

Memoria del I Simposio  
Internacional sobre

# ORGANIZACIÓN DEL CONOCIMIENTO: BIBLIOTECOLOGÍA Y TERMINOLOGÍA

Catalina Naumis Peña

Coordinadora

CONTROL VOCABULARIO

language structure

terminology

recuperación de información

lenguaje controlado

traduction littérale

indización temática

language communication and cognition

tesauros

LINGÜÍSTICA COMPUTACIONAL

natural language

traduction littérale

identity

contenido documental

language structure

informatique

transmission and diffusion

lenguaje controlado

LINGÜÍSTICA COMPUTACIONAL

recuperación de información

language in medicine

natural language

library science

transmission and

library systems

transmission and



La presente obra está bajo una licencia de:

[http://creativecommons.org/licenses/by-nc-sa/3.0/deed.es\\_MX](http://creativecommons.org/licenses/by-nc-sa/3.0/deed.es_MX)



Atribución-No Comercial-Licenciamiento Recíproco 3.0 Unported

Eres libre de:



copiar, distribuir y comunicar públicamente la obra



hacer obras derivadas

Bajo las condiciones siguientes:



**Atribución** — Debes reconocer la autoría de la obra en los términos especificados por el propio autor o licenciante.



**No comercial** — No puedes utilizar esta obra para fines comerciales.



**Licenciamiento Recíproco** — Si alteras, transformas o creas una obra a partir de esta obra, solo podrás distribuir la obra resultante bajo una licencia igual a ésta.

Esto es un resumen fácilmente legible del:  
[texto legal \(de la licencia completa\)](#)

En los casos que sea usada la presente obra, deben respetarse los términos especificados en esta licencia.



---

---

**Memoria del I Simposio Internacional sobre  
Organización del Conocimiento:  
Bibliotecología y Terminología**

**COLECCIÓN**  
**Sistematización de la Información Documental**  
**Centro Universitario de Investigaciones Bibliotecológicas**

---

---

**Memoria del I Simposio Internacional sobre  
Organización del Conocimiento:  
Bibliotecología y Terminología**

**Coordinadora**

**Catalina Naumis Peña**



**Universidad Nacional Autónoma de México  
2009**

**Z693.3**  
**T47S55**

Simposio Internacional sobre Organización del Conocimiento : bibliotecología y terminología ( 1 : 2007 : México, D.F.)

Memoria del I Simposio Internacional sobre Organización del Conocimiento : bibliotecología y terminología, 27 al 29 de agosto de 2007 / coordinadora, Catalina Naumis Peña.—México : UNAM, Centro Universitario de Investigaciones Bibliotecológicas, 2009 (Sistematización de la Información Documental) 627p.

ISBN: 978-607-2-00117-6

1. Organización del Conocimiento 2. Terminología 3. Bibliotecología

I. Naumis Peña, Catalina, coordinadora II.t.

***Diseño de portada:*** Mario Ocampo Chávez

Primera Edición 2009

DR © UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Ciudad Universitaria, 04510, México D.F.

Impreso y hecho en México

ISBN: 978-607-2-00117-6

# Contenido

<b>INTRODUCCIÓN</b> .....	xiii
---------------------------	------

## **SECCIÓN I**

### TERMINOLOGÍA EN DIFERENTES ÁMBITOS DE CONOCIMIENTO

LAS PALABRAS Y SU TELARAÑA. EL ORIGEN DE LA RED.....	3
Elizabeth Luna Traill	

DEFINING BIBLIOGRAPHIC 'WORKS:' NAÏVE CLASSIFICATION FOR TERMINOLOGY GENERATION.....	7
Richard P. Smiraglia	

LOS FILTROS TÉRMINO-LINGÜÍSTICOS PARA LA ACTIVIDAD DOCUMENTAL.....	19
Maria de Fátima Gonçalves Moreira Tálamo y Marilda Lopes Ginez de Lara	

LA TERMINOLOGÍA Y LOS CONCEPTOS EN LAS CIENCIAS HISTÓRICAS: UNA PERSPECTIVA INTERDISCIPLINARIA DESDE LA CIENCIA DE LA INFORMACIÓN Y LA DIDÁCTICA DE LA HISTORIA.....	37
Martha Sabelli	

## **SECCIÓN II**

### ANÁLISIS TERMINOLÓGICO

EL DESCRIPTOR Y EL TÉRMINO. LOS CONCEPTOS Y LA LINGÜÍSTICA.....	53
Ana María Cardero	

POSIBILIDADES Y LÍMITES DEL ANÁLISIS CUANTITATIVO DE CORPUS ESPECIALIZADOS.....	63
Elena Bogomilova Lozanova	

EL ROTULAJE NUTRICIONAL COMO FUENTE DE INFORMACIÓN PARA EL CONSUMIDOR: UN ESTUDIO EXPLORATORIO.....	79
Fernando Bittencourt Dos Santos	

LOS PROBLEMAS TERMINOLÓGICOS DE LA ORGANIZACIÓN Y DEL ACCESO A LA INFORMACIÓN.....	87
Nair Yumiko Kobashi	

### **SECCIÓN III**

#### **TRADUCCIÓN PARA TRANSFERIR INFORMACIÓN**

ESTRATEGIAS TERMINOLÓGICAS PARA LA RESOLUCIÓN DE TRADUCCIONES DE TÉRMINOS MÉDICOS.....	105
---	-----

Graciela Rosemblat

ANÁLISIS DE LOS TIPOS DE DEFINICIONES TERMINOLÓGICAS QUE APARECEN EN VARIAS ÁREAS DEL LENGUAJE ESPECIALIZADO.....	125
--	-----

Érika Ehnis Duhne

LA GESTIÓN DOCUMENTAL APLICADA A LA ELABORACIÓN DE UN BANCO DE CONOCIMIENTO ESPECIALIZADO.....	137
---	-----

Anna Aguilar-Amat, Bartolomé Mesa Lao  
y Pilar Sánchez-Gijón

EL TRATAMIENTO TERMINOLÓGICO EN LA LISTA DE ENCABEZAMIENTOS ARMARC PARA BIBLIOTECAS PEQUEÑAS.....	151
--	-----

Gloria Rojas León  
y Ageo García Barbabosa

PROCESOS EN LA TRADUCCIÓN DE “SUBJECT HEADINGS FOR SCHOOL AND PUBLIC LIBRARIES”.....	163
---	-----

Joanna F. Fountain

### **SECCIÓN IV**

#### **SEMÁNTICA EN LA INFORMACIÓN**

LA BIBLIOTECA SEMÁNTICA.....	183
------------------------------	-----

Juan Voutssás Márquez

EL TRABAJO TERMINOLÓGICO EN LA ESCUELA INTERAMERICANA DE BIBLIOTECOLOGÍA DE LA UNIVERSIDAD DE ANTIOQUIA.....	195
---	-----

María Teresa Múnera Torres y Carlos Arturo Montoya Correa

ONTOLOGÍA SOBRE EL DESPLAZAMIENTO FORZADO EN COLOMBIA.....	209
---	-----

Sandra Arenas Grisales

ONTOLOGÍAS PARA RELACIONAR CONTENIDOS EDUCATIVOS DIGITALES.....	227
--	-----

Clara López Guzmán  
y Francisco J. García Peñalvo



## **SECCIÓN V**

### **REPRESENTACIÓN Y RECUPERACIÓN DE INFORMACIÓN**

LA TERMINOLOGÍA COMO MÉTODO PARA REPRESENTAR Y ORGANIZAR DOMINIOS MULTIDIMENSIONALES.....	243
María José López Huertas	

LA ORGANIZACIÓN Y RECUPERACIÓN DE LA INFORMACIÓN DE ACUERDO CON EL CONTEXTO DEL USUARIO.....	263
Nydia Rojas Morales	

COMPORTAMIENTO DE BÚSQUEDA Y USO DE LA INFORMACIÓN AMBIENTAL POR INVESTIGADORES DEL ÁREA DE MEDIO AMBIENTE.....	275
Fernando Bittencourt dos Santos y Helen de Castro Silva	

PROCEDIMIENTOS TERMINOLÓGICOS Y DOCUMENTALES UTILIZADOS EN LA CONSTRUCCIÓN, IMPLEMENTACIÓN Y GESTIÓN DEL VOCABULARIO CONTROLADO USP-DESCRIPTORES EN PORTUGUÉS PARA LA REPRESENTACIÓN Y RECUPERACIÓN DE LA INFORMACIÓN.....	291
Vânia Mara Alves Lima, Nair Yumiko Kobashi, Vera Regina Casari Boccato, María Ángela de Toledo Leme y Mariza Leal de Meirelles Do Coutto	

ALGUNAS GRIETAS SEMÁNTICAS EN LA RECUPERACIÓN DE INFORMACIÓN: UNA PERSPECTIVA DECONSTRUCTIVA PARA UNA SOLUCIÓN PRAGMÁTICA.....	305
Rafael Ibarra Contreras	

CONOCIMIENTO DE LAS ARTES PLÁSTICAS Y LA ARQUITECTURA MEXICANAS.....	321
Elsa Barberena Blásquez	

## **SECCIÓN VI**

### **LA INFORMÁTICA EN LA RECUPERACIÓN DE INFORMACIÓN.**

PROCESAMIENTO DEL LENGUAJE NATURAL: ESTADO DE LA INVESTIGACIÓN.....	337
Alexander Gelbukh	

ACTUALIZACIÓN DEL CONCURSO SIMULTÁNEO EN EL USO DEL LENGUAJE LIBRE Y DEL CONTROLADO: FOLKSONOMÍAS Y TAXONOMÍAS.....	359
José Antonio Moreiro González, Jorge Morato Lara, Sonia Sánchez-Cuadrado y Vicente Palacios	
IMPORTANCIA DE LA TERMINOLOGÍA PARA LOCALIZAR Y RECUPERAR INFORMACIÓN CIENTÍFICA.....	387
Imelda Velázquez Montes	
LA PRIMERA BANDA DEL CONTÍNUM DE METADATOS, UN NIVEL POCO GESTIONADO POR LOS PORTALES BIBLIOTECOLÓGICOS Y LAS BIBLIOTECAS DIGITALES.....	399
Alejandro Uribe Tirado	
SOFTWARE LIBRE EN LA REPRESENTACIÓN, BÚSQUEDA, RECUPERACIÓN E INTERCAMBIO DE INFORMACIÓN.....	419
Dante Ortiz Ancona	
<b>SECCIÓN VII</b>	
LA TERMINOLOGÍA EN LA BIBLIOTECOLOGÍA	
ALGUNOS PROBLEMAS TERMINOLÓGICOS EN EL DOMINIO DE LA BIBLIOTECOLOGÍA Y DOCUMENTACIÓN. UNA BABEL TERMINOLÓGICA-CONCEPTUAL.....	435
José López Yepes	
LA ESTRUCTURA SINDÉTICA COMO MECANISMO DE RECUPERACIÓN DE LA INFORMACIÓN TEMÁTICA: LCSH YLEMB .....	467
Aurora Vega Martínez y Ageo García Barbabosa	
CATÁLOGO DE AUTORIDAD TERMINOLÓGICO PARA EL ANÁLISIS DE LA INFORMACIÓN DOCUMENTAL DE LA COMUNIDAD UNIVERSITARIA DE UNIDADES DE INFORMACIÓN ESPECIALIZADAS EN CIENCIAS SOCIALES (CUUICS). UNA EXPERIENCIA METODOLÓGICA COSTARRICENSE.....	479
Lucía Matamoros Segura	
LA RECUPERACIÓN DE LA INFORMACIÓN EN LA RED ESPECIALIZADA EN CIENCIAS SOCIALES “CUUIS” EN COSTA RICA Cristina Rocha Palma.....	497

LA NORMALIZACIÓN DE NOMBRES GEOGRÁFICOS CONTENIDOS EN ENCABEZAMIENTOS DE MATERIA.....	509
Noé Ángeles Escobar	

## **SECCIÓN VIII**

### **NORMAS Y USO DEL LENGUAJE EN UN AMBIENTE DIGITAL**

LAS TECNOLOGÍAS DEL LENGUAJE EN EL MANEJO DE INFORMACIÓN Y EN LA GENERACIÓN DEL CONOCIMIENTO.....	531
Gerardo Sierra Martínez	

ORGANIZACIÓN Y ESTRUCTURACIÓN DEL CONOCIMIENTO EN LA NORMALIZACIÓN INTERNACIONAL: EL CASO DE LAS NORMAS ISO 704 Y ANSI/NISO Z39.19-2005.....	547
María Pozzi	

INDIZACIÓN Y RECUPERACIÓN DE LA INFORMACIÓN EN INTERNET. EL USUARIO ANTE LA INTERPRETACIÓN Y TRADUCCIÓN DE LA DEMANDA.....	561
Marta Fuentes Agustí y José Hernández Serrano	

INDIZACIÓN Y RECUPERACIÓN DE LA INFORMACIÓN: TÉRMINOS QUE REPRESENTAN LAS INTERRELACIONES EN LA ZONA FRONTERIZA NORTE.....	571
Lil Bidart	

EL VOCABULARIO CONTROLADO DE LA BASE DE DATOS IRESIE: ESTRATEGIAS PARA SU CONSTRUCCIÓN.....	581
María Ángela Torres Verdugo y Rosa María Oriol Muñoz	

<b>DIRECTORIO DE AUTORES.....</b>	<b>599</b>
-----------------------------------	------------



## SECCIÓN I

### Terminología en diferentes ámbitos de conocimiento



# Las palabras y su telaraña. El origen de la red

ELIZABETH LUNA TRAILL  
*Instituto de Investigaciones Filológicas  
Universidad Nacional Autónoma de México*

**P**rimero que nada deseo agradecer el inmerecido honor que me han hecho el doctor Filiberto Felipe Martínez Arellano, director del Centro Universitario de Investigaciones Bibliotecológicas y la doctora Catalina Naumis Peña, organizadora de este Primer Simposio Internacional sobre Organización del Conocimiento: Bibliotecología y Terminología, de pronunciar unas palabras inaugurales. Inmerecido honor porque estoy muy lejos de ser una especialista en terminología; soy apenas una iniciada, gracias, por cierto, a la doctora Ana María Cardero.

En los albores del siglo XXI, este campo de conocimiento tiene, indudablemente, una vital importancia, al constituirse en una interdisciplina (o mejor llamarla transdisciplina) que parece ser una forma de investigación para la que difícilmente hay un camino de regreso. Bibliotecólogos, terminólogos, informáticos, lingüistas se reúnen hoy para intercambiar ideas y enriquecer el trabajo de comunicación del conocimiento que realizan. Y la felicitación para Catalina es doble porque no solamente ha logrado conjuntar a este selecto grupo de investigadores, sino porque lo ha hecho en nuestro medio académico, en el que, desafortunadamente, no se le ha dado a este tipo de investigaciones el lugar que merece.

Cometería yo una falta de respeto si en esta breve conferencia me atreviera a exponer ante ustedes, distinguidos especialistas aquí presentes, el estado del arte de la terminología y las preocupaciones que por hoy ocupan a quienes dedican sus afanes a estas investigaciones. Por ello me limitaré sencillamente a compartir con ustedes algunas breves reflexiones sobre la lengua.

Antes de adentrarme en el tema, quisiera enfatizar la singular relevancia que han adquirido los vocabularios especializados en la Sociedad de la información que hoy vivimos. Como bien señala Jacques Ellul, el hombre moderno no está subinformado sino superinformado, hasta el punto de que es incapaz de apreciar la exactitud, la importancia, la coherencia de las informaciones que lo aplastan. Sólo al seleccionarlas y organizarlas se les da un sentido a esas informaciones. Por otra parte y refiriéndonos específicamente a la Sociedad de la información (a la Sociedad del conocimiento, como también se le suele denominar) recordemos con Estela Morales “[...]que ya se habla de ella como una diferenciación histórica similar a la de la Revolución industrial o la sociedad posindustrial, lo cual denota que más que depender de las máquinas, dependemos de la información que obtenemos y sabemos utilizar, así como de los esfuerzos para convertirla en conocimiento”.

Como bien sabemos, los estudios sobre el lenguaje en Occidente nacen como parte de la Filosofía: Aristóteles es el primero en ocuparse sobre la división de las partes de la oración; Platón y su disquisición sobre el lenguaje en su famoso *Cratilo*, son dos ejemplos paradigmáticos. Hoy, los filósofos siguen ocupándose, permítaseme la expresión, de lo más humano que tiene el ser humano: el lenguaje. “Todo pensamiento, señala Cassirer, tiene que pasar por la prueba del lenguaje hablado o escrito, y también el sentimiento se prueba y corrobora al expresarse. Todos hemos pasado por la experiencia de sentimientos no pocas veces capaces de las cosas más asombrosas en este tipo de pensamiento “no formulado” característico de los sueños. Logramos resolver así, como jugando, los problemas más difíciles. Pero al despertar, todo se esfuma; la necesidad de cifrar en palabras lo conseguido nos lleva a darnos cuenta de que todo aquello no era más que una sombra vana. Por tanto, el lenguaje no es solamente, ni mucho menos,



algo que nos aleja de nosotros mismos; es, por el contrario, el camino que nos conduce a nosotros mismos y es eminentemente creador, por cuanto que sólo gracias a él se constituye nuestra conciencia del yo y nuestra autoconciencia. En otro momento, Cassirer señala: “De la razón investida en el lenguaje y que se expresa en sus conceptos se pasa a la razón científica. El lenguaje, con los medios de que dispone, no puede engendrar ni siquiera alcanzar, el conocimiento científico. Constituye, sin embargo, una etapa necesaria, en el camino que conduce a este tipo de conocimiento; es el medio sin el cual no puede nacer ni desarrollarse el saber en torno a las cosas. El acto de dar nombre a las cosas constituye la frase preliminar y la condición indispensable para llegar a determinarlas; es decir, para lo que constituye la función peculiar y específica de la ciencia. Así se comprende por qué la filología representa un aspecto necesario e integrante de la teoría del conocimiento”.

A mi entender las palabras de Cassirer expresan diáfana y claramente el quehacer que hoy nos reúne. En sus manos está utilizar una disciplina lingüística para encontrar los términos más pertinentes de cada campo del conocimiento, gracias al diálogo que mantienen entre sí y con los especialistas de aquellas áreas. Sus logros, vertidos en diccionarios, tesauros, terminologías, constituyen un elemento *sine qua non* para alcanzar la claridad que requiere el hecho de organizar y recuperar la información.

La lingüística norteamericana de las primeras décadas del siglo XX se caracterizó por su interés en la descripción de las lenguas amerindias. Este hecho permitió que sus lingüistas conocieran y describieran toda una variedad de lenguas muy diferentes de las que tradicionalmente había estudiado la filología europea. Indudablemente una de sus grandes figuras fue Edward Sapir, en quien se ha dado como en pocos lingüistas una síntesis perfecta de teórico y de empirista, como bien señala Sebastián Serrano en su historia de lingüística. Para Sapir el lenguaje organiza nuestro conocimiento del mundo, por ello a lenguas diferentes corresponden visiones del mundo diferentes.

Compromiso común de quienes se reúnen en este Simposio es la búsqueda de la difusión de los valores que cada sociedad posee porque, en efecto, cada lengua transmite un sistema de valores que le

es propio. Me llena de satisfacción saber que hay la preocupación de llevar a cabo investigaciones terminológicas en nuestra lengua para transmitir la información sin cargas culturales extrañas que la hagan menos comprensible. Si bien el mundo globalizado que nos ha tocado vivir exige definir los términos provenientes de otras lenguas, la recuperación y conservación de la identidad propia de cada lengua es fundamental.

Por lo anterior, me inquieta el contenido de dos artículos que tuve oportunidad de leer en estos días y en los que se habla sobre lo que en inglés se denomina *Plain language* y en español *Lenguaje ciudadano*: el lenguaje deja de ser un valor de cultura y se convierte en un objeto de comercio. Este movimiento, que ha sido ya transplantado a México se utiliza en la Secretaría de la Función Pública con el fin de formular mensajes claros y concretos para que el ciudadano reciba la información “que necesita” y que obviamente es la que se le quiere dar. Se empobrece así la lengua, desaparecen las variedades regionales y, por cierto, no se cuida de utilizar anglicismos y galicismos. Se pierde la identidad social; ¿se busca, acaso un esperanto del español? se pregunta uno de sus autores.

Ciertamente, como señala Olivier Reboul, en su *Lenguaje e ideología*: “La liberación del lenguaje no significa la liberación de los hombres. Sería demasiado simple, pero al menos la liberación del lenguaje es el paso obligado hacia la emancipación de los espíritus”.

# Defining Bibliographic ‘Works:’ Naïve classification for Terminology Generation

RICHARD P. SMIRAGLIA  
*Long Island University, E.U.A.*

**ABSTRACT:** Empirical methods are effective for the generation of descriptive terminology, especially concerning new or little-understood phenomena. Direct observation provides a critical base point from which terminology can be generated for further rational analysis. This procedure has been called “naïve classification” because it closely parallels the process of category generation in the laboratory. In the present paper, the evolution of terminology used to describe the phenomena concerning bibliographic “works” is demonstrated as a form of naïve classification.

## 1. COGNITION, CLASSIFICATION, NAÏVE CLASSIFICATION

**C**lassification is a near universal human phenomenon. When you say hello to a child and she says “Grandma,” it is because she has recognized that you are her grandmother, and therefore not any other person. She has created a classification with at least two categories—grandmother and not-grandmother—and she has assigned you as a member of one category and therefore not a member of the other. It is simple cognition at one level, but it is also classification. Therefore classification permeates human activity.

In a formal sense classification is acknowledged to have two uses for scholarship. First, scientists use classification to order their phenomena of study—often called *taxonomy*, this activity is essential for the advancement of knowledge. The second major use of classification is for the ordering of useful knowledge, and this can be seen as activity that crosses a broad spectrum of uses from social to scholarly to bibliographic. Dahlberg (2006) points out the common methods in use, which are the designation of objects of interest (knowledge elements), designation of the conceptual parameters for categories and relationships among them (knowledge units; i.e., this is the building of ontology), and the mapping of entities to the designated structure (knowledge systems). Your doctor checks a box on a diagnostic form, you find the tomatoes in the Italian foods section of your supermarket, you find mystery next to biography at your public library—all of these are examples of the use of classification for the useful ordering of knowledge.

In this paper I am concerned with the former definition of classification—that for use in the discovery process of science. I am particularly eager to demonstrate how this process has been used over the past two decades to investigate the nature of the bibliographic phenomenon called “works,” and the phenomenon of “instantiation” that is shared among diverse types of information objects. I bring this work to the attention of the present symposium as an example of the appropriate use of knowledge organization for the generation of terminology in the field of library and information science.

The specific research stream devoted to the definition of “works” and “instantiation” constitutes an example of what Beghtol (2003) has called “naïve classification.” The direct observation of phenomena provides a critical base point from which rationalism may be applied to provide conceptual connections and enhance understanding of research phenomena. More simply put, the researcher investigates empirically, records his observations, names his categories, and then adjusts the categories and the terms over time as new data are discovered. In this way naïve classification of works has yielded a rich typology of instantiation.

The empirical derivation of knowledge-elements, particularly in developing or evolving KO systems, provides a basis upon which conceptual systems can be built. This is exactly Beghtol's process of naïve classification for use in evolving scholarship. According to Beghtol (2003, 66), the process of naïve classification has many uses, including the discovery of gaps in knowledge, the reconstruction of historical evidence, and the revision or amplification of existing knowledge organization schema, among others. The process requires the scholar to articulate the purpose for his work so as to limit the empirical parameters, and then a variety of techniques may be employed, including paradigm-identification, and ordering (hierarchy, tree-structure, facing) techniques.

Beghtol refers to studies that report naïve classification of Chinese plates, paintings, religions, photographs, 13<sup>th</sup> century Spanish silks, and child-rearing practices, among others. Green and Fallgren (2007) use the techniques to analyze document structures for the revision of the *Dewey Decimal Classification*. Let us look at a very simple example. We begin by identifying the phenomena observed and creating simple groupings. Figure 1 shows a set of observations divided into two clusters.

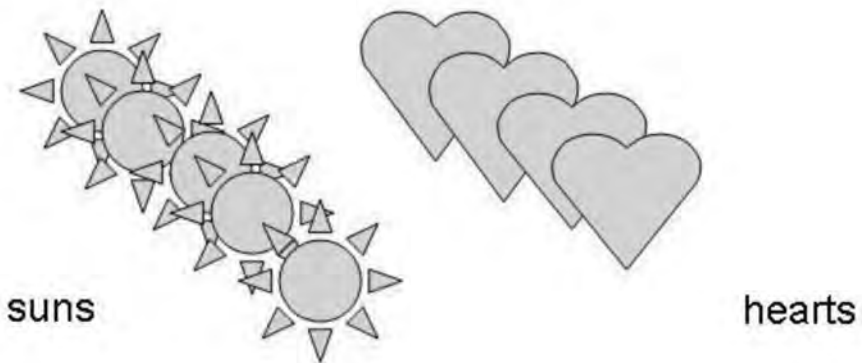


Figure 1. Suns and Hearts

In this illustration we have a naïve classification. There are nine objects in our laboratory, and they clearly can be divided by observable likeness into two categories, hearts and suns. There are five suns and four hearts. This classification is complete, because it identifies the knowledge elements (objects of study), knowledge units (hearts and suns), and a knowledge system “Suns and Hearts.” The classification is naïve because it represents merely the grouping of observations in this particular instance. We do not yet know why the suns are not hearts or why the hearts are not suns, we do not know what the suns and hearts have in common except that they both are found in this observation, we do not know why there are more suns than hearts or fewer hearts than suns, and we do not know why there are not other objects, such as moons or lightening bolts or stars. In research, the use of naïve classification is intended for just this process—to identify the paradigm and to create a matrix of its contents from which hypotheses can be generated to structure future research.

In the remainder of this paper we will define the concepts of “works” and “instantiation,” we will briefly summarize empirical methods that have been applied to their analysis and the naïve classifications that have resulted, and we will examine the typology of instantiation that exists at present. Finally we will look at this typology revealed by naïve classification over and against several rational schemes that have arisen elsewhere in literature. In this way I hope simply to demonstrate the usefulness of naïve classification for the advancement of knowledge through empirical research.

## 2. WORKS, AND INSTANTIATION

In library and information science we are much concerned with the artifacts of recorded knowledge, and in particular, we are motivated to organize these artifacts for retrieval, either in whole (as when a reader seeks a particular book) or in part (as when a scholar seeks a particular fact). We sometimes speak of these artifacts in specific terms: books, maps, journals, scores, videos, etc.—and we sometimes speak of them in generic terms: documents. In either event, we make a clear distinction between the objects themselves and their content. A

“work” is the intellectual content of a document. But a work is much more than that because research has shown that it is also the deliberate creation of intellectual content that is intended by its creator to be communicated to some audience. Thus a seashell may be described as a document because for some scholars it is considered informative, but it is not a work, per se. But a sculpture of a sea shell is a work, because its sculptor has intended it to be appreciated by its audience.

Most cataloging codes make a distinction between works and items, but until recently the distinction was not clear. Further, the structure of catalogs has failed to make organization and retrieval of works a primary objective. This becomes a problem when a single work exists in multiples—as when a novel becomes a screenplay, which becomes a motion picture, which becomes a video, whose music is issued on a CD, and so on. In these cases all of these diverse publications are likely to have the same name (e.g., *Harry Potter and the Order of the Phoenix*), therefore it becomes necessary to introduce order in the display to disambiguate the cluster. Another important aspect of the problem of works in the catalog is that of collocating all of the works of a particular creator. Again, ordering devices are required to disambiguate the collocated works under a particular heading (e.g. all Harry Potter books and their translations and all movies of Harry Potter books). These problems are treated fully in Smiraglia (2001).

The phenomenon of multiple iterations of a work over time is called “instantiation.” Instantiation occurs when some cultural catalyst acts like a market force and causes a work that has entered a canon to be translated, adapted, edited and reissued over time. The longer the period of cultural catalysm the greater the number of instantiations, and the more recent (say, the current decade, rather than the 19<sup>th</sup> century) the greater the complexity of the instantiation set. That is, we know from prior research that significant works have many editions and translations, and that works that have become significant recently will have many multi-media iterations. The problem for research has been to move beyond the simple concept of works present in cataloging codes and to understand the phenomenon more fully. The origins of this phenomenon are treated in Smiraglia (2005b).

## 2.1 EMPIRICAL METHODS FOR STUDYING WORKS AND INSTANTIATIONS

Research about works and instantiations has followed an empirical trajectory for more than two decades. Many studies (detailed in Smiraglia 2001, and treated as meta-analysis in Smiraglia 2002) have yielded interesting data about works and the extent of their instantiation. Roughly consistent statistical data have emerged showing that in most libraries the proportion of instantiating works follows roughly Lotka's law (thus around 30-40% instantiate), while most other works are unique singletons. These data emerge from empirical studies of library collections and bibliographic utilities in which a common methodology was employed. This methodology can be summarized briefly:

- Identify a sampling frame (e.g. an online catalog);
- Select a random sample of bibliographic records;
- Adjust the sample to make sure each "work" is unique in the sample, and to make certain each instantiated work in the sample is the progenitor in the sampling frame;
- Constitute the resulting list of works;
- Search all works in the bibliographic utilities and compile all sets of instantiations;
- Analyze the instantiation sets.

A taxonomy of instantiation was created by Smiraglia (1992), which included these categories:

1. Simultaneous derivations;
2. Successive derivations;
3. Translations;
4. Amplifications;
5. Extractions;
6. Adaptations; and,
7. Performances.



As research progressed new categories were discovered in the data. For instance, works that have large instantiation sets are often derived from earlier works, thus the category “predecessor work” was added. Also, many works are issued together with other works (a map, for instance, inside a book); these are called “accompanying works.” Vellucci (1997) found two categories of uniquely musical instantiation, which she called “notational transcription” and “musical presentation.” Smiraglia (2007) also found a category that occurs among best-selling works, which he called “persistent works.” In this way, the principle of naïve classification expanded the list of instantiation types for published works. As Beghtol suggests, naïve classification leads to new hypotheses, and in this case semiotic analysis of instantiation demonstrated that some of these categories represented new publications in which there had been little change in semantic content, while others represented tremendous change in semantic content. Thus, two meta-categories were developed—derivation, to describe instantiation with little or no change, and mutation, to describe instantiation with much change—that can be applied roughly hierarchically, thus:

Table 1. Naïve classification of instantiation types for publications

<b>Derivations</b>
simultaneous editions
successive editions
predecessors
amplificacions
extractions
accompanying materials
musical presentations
notational transcription
persistent works
<b>Mutations</b>
translations
adaptations
performances

## 2.2 CASE STUDY FOR STUDYING INSTANTIATION AMONG ARTIFACTS AND ARCHIVES

To extend this research, study of museum artifacts and archival documents was undertaken. Although the concept of “works” remains limited to deliberately created intellectual entities, the question was whether other informing objects would have instantiation networks that might require disambiguation for information retrieval. That is, while our naturally occurring sea shell is not a work, per se, if it is collected by a repository, there will be representations of it in the repository. Is it possible these would undergo the phenomenon of instantiation? The answer was resoundingly affirmative. Analysis of eight Etruscan artifacts from the University of Pennsylvania Museum of Archaeology and Anthropology was undertaken using case study method. The results are reported fully in Smiraglia (2005a). In this case, for each artifact the museum’s archives were searched thoroughly to locate all representations in-house. Subsequently, published literature was searched to locate additional representations. The results yielded a typology of both representations (photographs, models, etc.) and metadata sets (narrative descriptions), and both were discovered to exist in multiple instantiations both in-house and in publication. Further, additional research was conducted on the collected archival papers of the United States Merchant Marine Academy’s Class of 1942 Archives (reported in Smiraglia 2006). The papers were analyzed prior to digitization, and again instantiation among representations turned up. In this case we found correspondence that was present as typescript, carbon copy, and photocopy, we found photographs that reappeared on post-cards, which were in turn scanned and used as digital images on computer-generated documents, and so on.

The resulting naïve classification of works and instantiation can be summarized in its present form in the following table (Table 2).

Table 2. Naïve Classification of Comparative Instantiation Typologies

Bibliographic works	Artifacts - Metadata	Artifacts - Representations	Personal Papers
simultaneous editions	finding aids	field photos	Photocopies

*Terminología en diferentes ámbitos de conocimiento*

▶ successive editions	field notes	working images	Carbon copies
amplifications	letters	exhibition color images	Photos
extractions	conservation treatment notes	digitized exhibition images	postcard with photo
Musical presentations	register descriptions; object cards	conservation photos	digitized scan of postcard with photo
National Transcriptions	image order invoices	archived photographic negatives	reprint of photo
Persistent works	museum database records	archived photographic prints	digitized scan of photo
	catalog card records	archived photographic transparencies	
translations		object reproductions	
adaptations		drawings	
performances		3D models	
predecessors			
Accompanying materials			

From left to right are seen the typologies of instantiation that arose from the research. It is apparent that the bibliographic typology indicates types of publications, whereas the typologies from the museum and the archives indicate document or artifact types. There is no horizontal correspondence to be drawn from this table; rather the entries are simply arrayed by type for display. But the hierarchical grouping of derivations and mutations that we saw in table 1 persists here. The terms entered below the solid line represent instantiation types that could be considered to be mutations, because the content of the original appears in altered form. All of the instantiation types above the line are derivations—they represent re-iterations of the original content.

### 3. THE VALUE OF NAÏVE CLASSIFICATION FOR TERMINOLOGY GENERATION

The purpose of this paper was to demonstrate the use of empirical means to generate a KO structure, and in particular to highlight the evolution from a naïve classification of simple knowledge-elements

to a more useful model of a phenomenon, incorporating conceptual knowledge-units. Following Dahlberg (2006) we have identified a set of typologies of instantiating information objects as knowledge elements, we have grouped them into knowledge units, and we have arrayed them hierarchically into a knowledge system. Following Beghtol (2003) we have used the paradigm-generating approach of naïve classification by declaring our purpose (to study works and instantiations for the purpose of aiding disambiguation in information retrieval), we have identified gaps in knowledge of the phenomenon and generated new hypotheses sequentially, and as the research progressed we have allowed the classification to evolve.

In this instance what began as empirical research has yielded a base level naïve classification, from which additional future hypotheses can be generated. We have not created a sophisticated taxonomy of instantiation. Rather, we have generated a typology of derivations and mutations of informing objects. This naïve classification then, suggests the value of continued research into the phenomenon of instantiation. It demonstrates the potential breadth of terms that should be incorporated in any future taxonomy. And it suggests that scholars of knowledge organization in library and information science might well continue to benefit from the use of empirical methods for terminology generation.

## REFERENCES

- Beghtol, C. (2003). Classification for information retrieval and classification for knowledge discovery: Relationships between “professional” and “naïve” classifications. *Knowledge organization* 30: 64-73.
- Dahlberg, I. (2006). Knowledge organization: a new science? *Knowledge organization* 33: 11-19.
- Green, Rebecca and Fallgren, Nancy. 2007. Anticipating new media: a faceted classification of material types. In Ten-

### *Terminología en diferentes ámbitos de conocimiento*

- nis, J. ed. *North American Symposium on Knowledge Organization* <http://dlist.sir.arizon.edu/1911>.
- Smiraglia, Richard P. 1992. Authority control and the extent of derivative bibliographic relationships. PhD. Dissertation. University of Chicago.
- Smiraglia, Richard P. 2001. *The nature of a work: implications for the organization of knowledge*. Lanham, MD: Scarecrow.
- Smiraglia, Richard P. 2002. Further progress in theory in knowledge organization. *Canadian journal of information and library science* 26 n2/3: 30-49.
- Smiraglia, Richard P. 2005a. "Content metadata—an analysis of Etruscan artifacts in a museum of archeology." *Cataloging & classification quarterly* 40 n3/4:135-51.
- Smiraglia, Richard P. 2005b. Instantiation: Toward a theory. In Vaughan, L., ed., *Data, information, and knowledge in a networked world*; Annual conference of the Canadian Association for Information Science ... London, Ontario, June 2-4 2005. Available <http://www.caais-acsi.ca/2005proceedings.htm>.
- Smiraglia, R.P. (2006). Empiricism as the basis for metadata categorization: expanding the case for instantiation with archival documents. In Budin, G., Swertz, C. and Mitgutsch, K., eds., *Knowledge organization and the global learning society*; Proceedings of the 9th ISKO International Conference, Vienna, July 4-7 2006, pp. 383-88.
- Smiraglia, Richard P. 2007. The 'works' phenomenon and best selling books. *Cataloging & classification quarterly* 44n3/4: 179-95..
- Vellucci, Sherry L. 1997. *Bibliographic relationships in music catalogs*. Lanham, Md.: Scarecrow Press.



# Los filtros término-lingüísticos para la actividad documental

MARIA DE FÁTIMA GONÇALVES MOREIRA TÁLAMO  
*Programa de Posgrado en Ciências de la Información -  
PUC-Campinas, Brasil*

MARILDA LOPES GINEZ DE LARA  
*Escola de Comunicações y Artes/Universidade de São Paulo, Brasil  
Programa de Posgrado em Ciências de la Informação - ECA-USP*

## INTRODUCCIÓN

La interfaz Lingüística, Documentación y Terminología ha sido objeto de estudios importantes para el desarrollo de metodologías, tanto para el tratamiento y la recuperación de la información como para la construcción de lenguajes documentales. No obstante, el avance de estos estudios no se corresponde con la oferta de procedimientos pedagógicos que posibiliten el uso efectivo de estas metodologías y que perfeccionen la formación del documentalista.

Usualmente, los procedimientos de la Terminología son incorporados a las metodologías de enseñanza de la Documentación sin que haya una revisión de los mismos frente a los objetivos específicos de la formación del alumno. Además, no se discute apropiadamente la naturaleza discursiva de los propios vocabularios de especialidad, y se los reduce al aspecto meramente prescriptivo. La presente propuesta,

cuyo texto se encuentra dividido en cuatro partes, expone algunos fundamentos de la Lingüística Documental, lo que permite superar el abismo existente entre la teoría y su traducción, para fines de formación del alumno.

En la primera parte del trabajo nos referimos a la superioridad del lenguaje general o natural como lugar original del significado, mostrando que las diferentes metodologías, al privilegiar determinados aspectos del lenguaje, no corroboran la concepción general y única de ese sistema semiótico. El reconocimiento del lenguaje general como medio de acceso al conocimiento justificaría las metodologías de enseñanza y aprendizaje propuestos por la Lingüística Documental, las cuales se fundamentan en la concepción del lenguaje general y dentro de ella integran el funcionamiento del lenguaje de especialidad. La segunda parte del trabajo expone una propuesta para componer la interfaz entre Lingüística Documental y Terminología, cuyo foco en la Documentación permite elaborar cuadros de referencia comprometidos con la actividad documental. Inmediatamente después se presenta una experiencia didáctica desarrollada a partir del referencial lingüístico documental. Y finalmente, se ofrecen notas conclusivas que demuestran que la experiencia desarrollada contribuye no sólo para el avance de la formación del documentalista sino también para la discusión contemporánea de la Terminología acerca de su propio objeto.

## EL LENGUAJE GENERAL

Existe consenso acerca de la idea de que el lenguaje es el lugar original del significado. Del mismo modo se comparte la noción de que convivimos con varios lenguajes, lo que lleva necesariamente a admitir que existe una relación entre significado y contexto lingüístico. Ese aspecto social del lenguaje no sólo configura a éste como objeto de estudio, sino que también permite abordarlo bajo puntos de vista asociados a diferentes metodologías que evidencian aspectos, aparentemente diferentes de un evento único (Tálamo y Lara, 2006).

Bajo la perspectiva de la historia social del lenguaje, por ejemplo, la observación de los diferentes modos de interpretación a menudo condujo hacia la idea de la existencia de una jerarquía entre los



lenguajes utilizados por una misma comunidad. La importancia de tales estudios se asocia a la reflexión que induce a pensar sobre las relaciones entre lenguaje y poder las cuales, a su vez, necesitan ser analizadas en el ámbito de la propia distribución del lenguaje; es decir, de las políticas de enseñanza y de uso. Integran también los estudios de esa naturaleza, las lenguas escritas que ejercen algún poder especial, como es el caso del discurso religioso (Burke y Porter, 1993).

Los lenguajes de especialidad y el lenguaje de tratamiento de la información, aunque sean motivo de estudios más técnicos, forman parte de esa categoría y frecuentemente son considerados lenguajes artificiales, porque no son adecuadamente interpretados por la sociedad. Paralelamente a ese aspecto excluyente de la actividad lingüística es necesario también reconocer que la definición del campo de la Lingüística que organiza sistemas sígnicos y realidad cultural atribuyéndoles fundamentalmente la función de recrear —“crear de nuevo esa realidad” (Lopes, 1987:16)— indica de forma bastante objetiva las funciones creadoras y traductoras del lenguaje.

La comprensión efectiva de tales funciones se relaciona con la propiedad esencial del signo de comportarse tanto como signo objeto o como meta-signo. Semejante función fundamenta no sólo la noción de semiosis ilimitada sino también la jerarquización de los sistemas semióticos asociada a la idea de modelización. Las lenguas naturales, llamadas comunes, “ocupan la posición jerárquica predominante entre todos los sistemas semióticos porque ellas constituyen la *única realidad inmediata para el pensamiento de cada uno de nosotros*” (Lopes, 1987:20). Dada esa capacidad de traducir cualquier otro sistema sígnico, los lenguajes naturales reciben el nombre de sistema semiótico o modelizante primario.

En ese sentido, cualquier lenguaje de especialidad forma parte de la lengua general y natural. El vocabulario de especialidad comparte igualmente esa característica, de modo que el signo lingüístico puede realizarse ya sea como palabra, como término, o como descriptor en el caso del lenguaje de tratamiento de la información. Justamente ésa es la concepción del signo que fundamenta teóricamente a la Lingüística Documental y estructura la forma

por medio de la cual ella opera con la palabra, el término y el descriptor; es decir, el modo a través del cual ensambla los diferentes niveles de significación.

Coseriu (1969) se vale de una metáfora para explicar la relación existente entre los niveles de abstracción —por ejemplo, el concepto en el lenguaje de especialidad— y el nivel del fenómeno concreto del habla.

Y otro punto fundamental es que el botánico comprenda que los tipos ideales a los que ha llegado no tienen existencia autónoma en un mundo aparte, no existen fuera de las flores concretas e independientemente de ellas; que las formas abstractas resumen y generalizan lo concreto, pero no se le oponen. Queremos decir que es sumamente importante no considerar la abstracción como *otra realidad*, sino sólo como un aspecto formal y sistemático comprobado, por necesidades científicas, en los mismos fenómenos concretos, como una manera de encarar aquella realidad única e indivisible que es el lenguaje humano (Coseriu, 1969:17).

Con ese argumento, Coseriu evidencia que existe una dinámica lingüística que corrobora la función modelizante primaria del lenguaje natural (también denominado lenguaje general o común) que hace que el conocimiento de un tipo de lenguaje sostenga el conocimiento de otro tipo. Por lo tanto, refuerza la idea de que vivimos en una cultura de comunicación en la cual el contacto entre diferentes lenguajes fundamenta no sólo la ampliación de las formas interpretativas sino que también consolida flujos de comunicación en la sociedad.

Siguiendo una línea de razonamiento semejante, Sager (1993) enfatiza que, a la par de la importancia que tienen las terminologías en el mundo contemporáneo para promover la comunicación eficiente entre especialistas, para el gran público el vocabulario de especialidad “no es más que una jerga de especialistas para engañar, confundir o impresionar con conocimientos superiores...” (Sager, 1993:11). Dicha situación ambigua proviene de la aludida idea de la separación entre los lenguajes. Con ella sucumbe la posibilidad de que las terminologías promuevan una efectiva mediación práctica para acceder al conocimiento técnico-científico.

Sager considera que tal dificultad con la terminología está asociada al incipiente conocimiento que se tiene sobre los mecanismos de la información y de la comunicación, lo que se refleja, a su vez, en el estudio autónomo de los diferentes lenguajes y léxicos. Con ello se oculta la propiedad común compartida por los sistemas lingüísticos. El concepto de lenguaje general que muestra esa propiedad común, supone:

- a) el funcionamiento modelizante primario del lenguaje natural a través del cual se hace el aprendizaje de los lenguajes de especialidad;
- b) la existencia de diferentes niveles de conocimiento del lenguaje de especialidad; los traductores e intérpretes, por ejemplo, comprenden textos especializados sobre la producción de automóviles pero son incapaces de producirlos (Sager, 1993:14).

La enseñanza del lenguaje de especialidad, particularmente los recursos metodológicos utilizados para ello, depende del nivel de conocimiento exigido por la formación del documentalista. Según la Lingüística Documental, se parte del lenguaje general para entender el funcionamiento de los lenguajes de especialidad y de la relación de éstos con el lenguaje de tratamiento de la información. De modo específico le cabe a la Lingüística Documental armonizar las funciones descriptivas de la palabra, la prescriptiva del término y la normalizadora del descriptor, integradas al lenguaje general. Por lo tanto, se entiende que la mediación práctica que ejercerá el documentalista presupone el dominio de una mediación práctica que debe ser construida durante el proceso de enseñanza-aprendizaje.

## LINGÜÍSTICA DOCUMENTAL Y TERMINOLOGÍA

A lo largo de la integración propuesta por la concepción de lengua general y de la concepción de signo, adoptada por la Lingüística Documental, la enseñanza de la Terminología para la formación del documentalista, viene procesándose de un modo estacionario.

El campo teórico de la Terminología, tal como fue propuesto por Wüster predominó durante mucho tiempo entre los estudios termi-

nológicos a través de la Teoría General de la Terminología. Fundamentándose en las dicotomías término y palabra, y onomasiología y semasiología, tal teoría propone el vocabulario del área de especialidad —denominado terminología— como un sistema de conceptos —considerado universal—. Lo importante, entonces, es el establecimiento de una relación biunívoca entre concepto y término para la efectua-ción de comunicación eficiente entre los pares.

La Teoría General de la Terminología (TGT), al hacer viable la rela-ción unívoca entre el concepto y el término, presenta una solución se-ductora para los problemas lingüísticos enfrentados por la Documen-tación. Con ella, de hecho, se consolida por ejemplo, la idea del signo monosémico, importante para conferirle credibilidad al tratamiento de la información y confiabilidad a los sistemas de recuperación de la información. No obstante, semejante proposición compromete la com-prensión del contexto discursivo donde se da la selección de las unida-des documentales —descriptores y no descriptores— al mismo tiempo que se vale de la concepción autónoma del lenguaje de especialidad.

Sin embargo, es preciso ir más allá para entender la extensión de tal procedimiento. Aunque en la actualidad se considere que la ter-minología sea el conocimiento lingüístico sobre un tema, es preciso reconocer que para la TGT esa proposición no está suficientemente clara. Para Wüster, de hecho, terminólogo y especialista son términos sinónimos: los especialistas son “los únicos que poseían el conoci-miento suficiente sobre una materia para encontrar los términos más adecuados” (Cabré, 2005).

La primacía del concepto sobre el término —o del significado so-bre el significante— es factible porque le corresponde al especialista la elaboración del propio vocabulario. De hecho, el especialista ape-nas puede operar con el conocimiento sobre los conceptos de una ma-tería para hacer la selección de las denominaciones más apropiadas. Así, la concepción onomasiológica que rige el trabajo terminológico no puede ser universalizada como recurso práctico. Basta que sea so-metida a una reflexión más profunda para que su fragilidad salga a la superficie.

En semejante situación se encontró la Documentación cuando adoptó el tesoro como instrumento de tratamiento de la informa-

ción para fines de recuperación. La Documentación tomó esa idea de la obra de Peter Market Roget, titulada *Thesaurus of english words and pbrases, classified and arranged so as to facilitate the expression of ideas and assist in literary composition*, publicada en 1852. Dado que se trata de un diccionario de lengua general, Roget organiza conceptos y expresiones, valores y sus formas de expresión. Por eso los diccionarios de carácter onomasiológico son denominados también diccionarios terminológicos, ideológicos o analógicos.

La sólida relación existente entre la Terminología y la Documentación se estableció tradicionalmente a partir de la orientación onomasiológica. No obstante, esa tendencia es tributaria de la Lógica y no reconoce el papel del lenguaje a no ser para la denominación. Tal concepción tiende a considerar los procesos de conceptualización, como independientes y anteriores a su expresión en lengua. Pero cuando la Lingüística Documental reconoce la naturaleza lingüística de los procedimientos documentales, sugiere combinar los dos procedimientos, sea por la observación de la experiencia práctica de construcción de lenguajes documentales, sea por reconocer, tal como lo sugiere Béjoint (1989), que el terminólogo realiza, casi simultáneamente, la actividad de inventariar los sentidos de cada forma y buscar inscribirlos teniendo como referencia los conceptos formulados en el dominio explorado.

Se entiende por onomasiología a la actividad lingüística que parte de la idea o significado e identifica las posibles palabras para expresarla. Ese procedimiento permite describir las variadas formas por las cuales la idea encuentra su expresión en las palabras. Por lo tanto, los procesos de denominación —de la idea hacia el signo— se encuentran en los orígenes de los estudios de naturaleza onomasiológica.

La orientación onomasiológica es importante para la Terminología porque provee una base para prescribir la relación concepto-término. Ya para la Documentación, tal orientación operacionaliza la comparación que es la base de la normalización, por ejemplo, la relación de equivalencia en los tesauros. Sin embargo, es una orientación inocua para el desarrollo del aprendizaje. Es común oír que le corresponde a los especialistas desarrollar algunas actividades documentales. Ahora podemos reconocer una de las evidencias de tal afirmación: la orien-

tación onomasiológica que acompaña, al menos en teoría, la elaboración de instrumentos documentales.

Pottier (1992) entiende que el recorrido de la enunciación involucra posiciones relativas tanto para el enunciador (emisor) como para el enunciatario (receptor) que corresponden a los recorridos onomasiológicos y semasiológicos, respectivamente. En ese sentido, del concepto a lo dicho, se tiene a la onomasiología, y de lo dicho a su interpretación, a la semasiología. Consecuentemente, la orientación onomasiológica pertenece a la categoría del enunciador y la semasiológica a la de la recepción.

Por lo tanto, se entiende que la proposición de una terminología o de un tesoro se sostenga prioritariamente en una orientación onomasiológica. Se parte de la conceptualización de una idea, del mundo referencial, que es expresada inmediatamente en signos, según los medios ofrecidos por el sistema de la lengua general. En términos hjelmeslevianos: se parte de la sustancia del contenido hacia la forma del contenido, valiéndose de los procesos de designación.

Para fines del aprendizaje, tal orientación no es de las más adecuadas. Para que tenga éxito el proceso es necesario que se viabilice una orientación semántica compatible con la categoría de la recepción; es decir, del intérprete que, frente al discurso, lo conceptualiza. Se observa entonces que para Pottier, independientemente del lugar de partida, la comunicación se realiza a nivel conceptual. Es importante destacar que ambas orientaciones fundamentan la sólida relación entre el lenguaje y el pensamiento. La onomasiología responde por la transformación de las representaciones mentales en textos, enunciados y discursos —se inicia en el concepto— mientras que la semasiología orienta la interpretación de esos mismos objetos —llega hasta el concepto. Para ello, la orientación semasiológica dispone de etapas: la primera con respecto a los contextos en que aparece la unidad e, inmediatamente, su inserción en paradigmas.

Esa comprensión de los procesos de significación e interpretación —filtros término-lingüísticos— que hace la Lingüística Documental, permite compatibilizar mediaciones prácticas con los objetivos comunicacionales de las terminologías y de la recuperación de la información de los lenguajes documentales. Se reconoce, por esa vía, que los

procedimientos onomasiológicos y semasiológicos están asociados y son interdependientes.

## ELABORACIÓN DE VOCABULARIO CON ORIENTACIÓN SEMASIOLÓGICA

Con la intención de identificar las orientaciones de producción de la significación, desarrollamos una experiencia en una disciplina de la carrera de graduación de Biblioteconomía y Documentación<sup>1</sup> que explora la interfaz Terminología-Documentación, teniendo como foco la formación del documentalista.

Después de introducir el objetivo de la Terminología y destacar la posibilidad de promover la apropiación de conocimientos para la Documentación, se les sugiere a los alumnos que exploren un pequeño *corpus* que reúne, en orden creciente de especialización, textos sobre una temática actual (por ejemplo, gripe aviaria, dengue). Tales textos son leídos buscando identificar los términos que mejor auxilien en su comprensión, siguiéndole a esa actividad, una discusión colectiva para seleccionar aquéllos que constituyen el vocabulario conceptual y el vocabulario funcional,<sup>2</sup> siendo que sobre los primeros incidirán los procedimientos de descripción terminológica. El registro de los términos se hace en fichas de vaciado que, para los objetivos documentales, abarca el término elegido, sus variaciones (sinónimos) y los contextos de uso (transcripciones literales del término en el enunciado donde aparecen), además la fuente, el dominio propuesto, el nombre del responsable por el llenado y la fecha.

---

1 [http://infobservatorio.incubadora.fapesp.br/portal/int\\_terminol](http://infobservatorio.incubadora.fapesp.br/portal/int_terminol)

2 El vocabulario conceptual es relativo a los términos que, por su forma o significado, denominan las realidades específicas de la especialidad; ya el vocabulario funcional está constituido por las expresiones del lenguaje natural que forman parte del vocabulario de los especialistas (Dubuc, 1999).

**Fig. 1: Ejemplo de ficha de vaciado**

TERMO CANDIDATO	SINÔNIMO	CONTEXTO	FONTE	DOMÍNIO
Gripe aviária		“O vírus causador da gripe aviária provém das vias respiratórias e está presente também em fezes. Outros animais, como porcos, são infectáveis. Entre as aves, não só frangos podem albergar o vírus, pois diversos tipos também são atingidos. O vírus espalha-se por determinados ambientes e assim chega a aves, mas o contato direto é mais perigoso”.	Ref. 1	Saúde Pública
Gripe aviária	Gripe do frango	“(…) conhecida no Brasil como gripe do frango(…)”	Ref. 2	Saúde Pública
Influenza aviária		“A influenza aviária (IA) constitui enfermidade epizootica de aves, causada pelo vírus influenza A e seus diferentes subtipos, com distribuição mundial”.	Ref. 5	Saúde Pública
Influenza aviária		“O primeiro surto de influenza aviária A (H5N1) em humanos ocorreu em Hong Kong (1997). Na ocasião, 18 casos foram confirmados, com registro de seis óbitos. No caso da influenza H5N1, evidências apontam para a transmissão aves para homem (...) possivelmente meio ambiente para o homem e transmissão muito limitada, não mantida, inter-humana”.	Ref. 7	Saúde Pública
Influenza aviária		“A possibilidade de influenza aviária H5N1 deve ser considerada em todos os pacientes com insuficiência respiratória aguda grave em países ou territórios com influenza animal (H5N1), em particular naqueles expostos às aves. Alguns surtos em aves, porém, foram reconhecidos após a ocorrência de casos em humanos. Além disso, deve ser levantada a possibilidade de ocorrência de influenza A (H5N1) em pacientes com doença grave sem explicação, como diarreia e encefalopatia, em áreas com doença animal”.	Ref. 8	Saúde Pública

**Nota:** Las referencias de los textos constituyen un archivo aparte



Paralelamente a la actividad práctica, son gradualmente introducidos los conceptos teóricos, que abarcan desde los fundamentos sobre las tipologías de textos en el interior del lenguaje general —el cotidiano, de divulgación, especializados— hasta las características y funcionalidades de la palabra, del término y, posteriormente, del descriptor. Se pretende que el alumno comprenda los diferentes niveles de lenguaje y sus características.

Después de la elaboración de las fichas de vaciado cada grupo de alumnos selecciona alrededor de tres términos sobre los cuales desarrollará un trabajo más profundo. En esa fase, los contextos son analizados para identificar las características nocionales y, en seguida, organizados según atributos de semejanza, operaciones que serán registradas en fichas terminológicas de síntesis. Al comparar los atributos, los alumnos operan, también, con la idea del concepto como haz de rasgos semánticos. En ese momento, se observa que las condiciones de uso, o los contextos comunicacionales, constituyen el lugar en que se realizan la terminologización (actualización del término) y la actualización de la palabra. El propósito de esa etapa es evidenciar que la identificación nocional del término se asocia a su uso contextualizado.

El proceso experimentado permite que los alumnos no sólo interpreten los discursos sino que también asimilen los conceptos en juego y propongan una definición para el término. Se confirma, de ese modo, que a partir del procedimiento semasiológico los alumnos disponen de elementos para interpretar las informaciones reunidas, pudiendo proponer, en seguida, la expresión de la conceptualización (proceso onomasiológico).

Fig. 2: Ejemplo de ficha terminológica de síntesis

CONTEXTO	CARACT. 1	CARACT. 2	CARACT. 3	CARACT. 4
A influenza aviária (IA) constitui enfermidade epizootica de aves, causada pelo vírus influenza A e seus diferente subtipos, com distribuição mundial. A principal via de transmissão do vírus da	A influenza aviária (IA) constitui enfermidade epizootica de aves, causada pelo vírus influenza A e seus diferentes subtipos.(...)	A principal via de transmissão do vírus da IA é (...) a horizontal, representada, principalmente,		

**Primer Simposio Internacional sobre Organización...**

<p>▶ IA é, sem dúvida, a horizontal, representada, principalmente, por excreções e secreções de aves migratórias.(...)</p>		<p>por excreções e secreções de aves migratórias. (...)</p>		
<p>No caso da influenza H5N1, evidências apontam para a transmissão das aves para homem (...). O período de incubação é de um a três dias, mas dentro de um plantel pode chegar até a 14 dias. As aves aquáticas, principalmente patos, são os principais reservatórios naturais dos vírus influenza aviária. Em geral, os pacientes apresentam como sintomas iniciais febre elevada e manifestações de infecção em trato respiratório inferior. (...) A letalidade foi elevada entre os pacientes hospitalizados e o óbito ocorreu, em média, nove a 10 dias após o início da doença</p>	<p>(...) influenza H5N1</p>	<p>(...) evidências apontam para a transmissão das aves para homem (...). As aves aquáticas, principalmente patos, são os principais reservatórios naturais dos vírus influenza aviária.</p>	<p>O período de incubação [no homem] é de um a três dias, mas dentro de um plantel pode chegar até a 14 dias</p>	<p>os pacientes apresentam como sintomas iniciais febre elevada e manifestações de infecção em trato respiratório inferior. (...) A letalidade foi elevada entre os pacientes hospitalizados e o óbito ocorreu, em média, nove a 10 dias após o início da doença</p>
<p>O vírus causador da gripe aviária provém das vias respiratórias e está presente também em fezes. Outros animais, como porcos, são infectáveis. Entre as aves, não só frangos podem albergar o vírus, pois diversos tipos também são atingidos. O vírus espalha-se por determinados ambientes e assim chega a aves, mas o contato direto é mais perigoso. Pessoas entram excepcionalmente nesse processo. O contágio, muito fácil, dificulta a prevenção. O vírus é rotulado como H5N1 (...)</p>	<p>O vírus é rotulado como H5N1.</p>	<p>o vírus... provém das vias respiratórias e está presente em fezes. (...) Entre aves, não só frangos podem albergar o vírus (...)</p>	<p>Pessoas entram excepcionalmente nesse processo. O contágio, muito fácil, dificulta a prevenção.</p>	<p>O vírus espalha-se por determinados ambientes e assim chega a aves, mas o contato direto é mais perigoso.</p>
<p><b>GRIPE AVIÁRIA: PROPOSTA DE DEFINIÇÃO:</b></p>				
<p>A gripe aviária constitui enfermidade epizootica de aves, causada pelo vírus influenza A e seus diferentes subtipos. O vírus é rotulado como H5N1. A principal via de transmissão do vírus é horizontal, representada principalmente por excreções e secreções de aves migratórias (...)</p>				

En este momento los alumnos tienen condiciones para comprender teóricamente la organización de sistemas de conceptos, ya que los dominan abstractamente y pueden desvincularlos de los textos de origen. A continuación se les propone un ejercicio que consiste en crear un árbol relativo a la parte del dominio focalizado,<sup>3</sup> el cual es progresivamente reorganizado al moverse efectivamente las informaciones de los registros anteriores. Se verifica aquí la oportunidad de trabajar los conceptos de categorías y categorización, nociones éstas más desarrolladas en la Documentación que en la Terminología, de modo que vinculen los resultados del procedimiento terminológico con los objetivos de organización y representación de la información según los sistemas informacionales revisados. De las nociones intuitivas de conjunto y grados de generalidad, se pasa a las nociones lógico-lingüísticas de encaje, de asociación por contigüidad espacio-temporal y de sinonimia.

Fig. 3: Ejemplo de árbol de dominio



3 Árbol de dominio: diagrama o estrutura que organiza, de modo funcional, los conceptos de un área temática. Dicho árbol no representa una clasificación científica sino una manera funcional de agrupar los conceptos de acuerdo con su parentesco.



Por medio de las informaciones de las fichas terminológicas, la categorización y el árbol conceptual (el árbol de dominio) construido inicialmente, se someten a la continua reorganización, sustituyéndose el producto intuitivamente construido valiéndose ampliamente del lenguaje general y de la percepción del término (a nivel signifiante), por otra persona que observe los trazos de los términos de los conceptos y que tenga sus bases legitimadas en las informaciones de los contextos de uso efectivo de los términos (terminologización). Dicho de otro modo, los alumnos son conducidos a verificar que las decisiones sobre la categorización y la clasificación sólo pueden formularse a partir de la observación de los enunciados realizados en los dominios de especialidad. Se reemplaza, vía observación de los contextos de uso de los términos, la organización empírica, por otra fuertemente vinculada con la literatura de especialidad.

Fig. 6: Ejemplo de red relacional de términos en el tesoro

<b>Gripe aviária</b>	<i>NE Enfermidade epizoótica de aves, causada pelo vírus influenza A e seus diferentes subtipos. O vírus é rotulado como H5N1. A principal via de transmissão do vírus é horizontal, representada principalmente por excreções e secreções de aves migratórias.</i>
	UP Influenza aviária
	UP Gripe do frango
	TT Doenças
	TG Gripe
	TR Diarréia
	TR Dor de cabeça
	TR Encefalite
	TR Febre
	TR Pneumonia
	TR Vírose
<b>H5N1</b>	UP Vírus Influenza A
	TG Vírus influenza
	TR Hemaglutinina
	TR Neuramidase
<b>Vírus influenza</b>	TE H5N1

Vale recordar que las informaciones relevadas pueden, también, ser re-aprovechadas si se le adjunta al tesoro un glosario, a través del cual se les ofrecerán a los usuarios de los sistemas de información, elementos para interpretar adecuadamente los descriptores.

Gracias al recorrido semasiológico y onomasiológico, los alumnos tienen finalmente condiciones para proponer la red relacional del tesoro documental, realizando de forma consistente la distribución de los términos según los niveles de superordenación, subordinación, asociación y equivalencia.

## OBSERVACIONES FINALES

Los avances de los estudios teóricos de las investigaciones compartidas por las diferentes áreas de conocimiento son comprobados y ganan visibilidad cuando se los integra a las estrategias de enseñanza-aprendizaje. La Lingüística Documental, como campo de la Ciencia de la Información, busca la armonización de teorías para elaborar no sólo procedimientos metodológicos para la construcción de sistemas para la organización y recuperación del conocimiento, sino también estrategias de enseñanza.

La propuesta presentada surgió básicamente de una relectura de los procedimientos de la TGT a la luz de la práctica documental y de los avances de la Lingüística en lo que se refiere especialmente a los estudios de los procesos de significación involucrados en la constitución del léxico.

El concepto de lenguaje general inspirado en las nociones de modelización e interpretaciones con orientaciones semasiológicas, se muestra suficientemente explicativo, no sólo para fundamentar el modo según el cual el documentalista opera con diferentes lenguajes, sin lidiar con la idea de oposición entre ellas, sino también para afirmar la propia noción que fundamenta el campo de la Ciencia de la Información: la interdisciplinariedad.

Por todo lo anterior, la contribución de la Lingüística Documental es extremadamente importante, ya que integra los procesos de enseñanza, la interdisciplinariedad, las nociones de lenguaje general, del término y de la palabra como el lugar de la manifestación de trazos

que, si bien diferentes, sirven como medio de tránsito interpretativo. Se observa que la discusión actual de la Terminología consiste justamente en la dificultad de armonizar un objeto plural o poliédrico, como es el término, dentro de una concepción unitaria. Dicho de otro modo, cómo dar cuenta de la unidad del objeto a partir de diferentes vertientes (Cabre, 2005).

Esa situación, bastante común en las ciencias sociales aplicadas, encuentra en la propuesta hasta aquí presentada una posible salida asociada a la concepción del múltiplo, o de la diferencia, como manifestaciones típicas de un sistema complejo que permite la interpretación de cualquier otra persona que participe en la naturaleza semiótica.

## REFERENCIAS

- Béjoint, H., À propos de la monosémie en terminologie, en *Meta*, v.34, n.3, pp.405-411, 1989.
- Burke, P. & Porter, R., *Linguagem, indivíduo e sociedade*, São Paulo, Ed. da Universidade Estadual Paulista, 1993.
- Cabré, M. T., La Terminologia, uma disciplina em evolução: pasado, presente e algunos elementos de futuro, en *Revista Debate Terminológico*, n.1, 03/2005, [en línea] [http://www.riterm.net/revista/n\\_1/cabre.pdf](http://www.riterm.net/revista/n_1/cabre.pdf) [Consulta: 04 de junio de 2007].
- Coseriu, E., *Teoría del Lenguaje y Lingüística General*, Madrid, Ed. Gredos, 1969.
- Dubuc, R., *Manual práctico de terminología.*, 3.ed. corr. atualiz.; trad. de Ileana Cabrera, Santiago de Chile, Unión Latina, Ril Ed., 1999.

*Primer Simposio Internacional sobre Organización...*

Introdução à Terminologia aplicada à Documentação. Site didático *Infobservatorio*, [en línea] [http://infobservatorio.incubadora.fapesp.br/portal/int\\_terminol](http://infobservatorio.incubadora.fapesp.br/portal/int_terminol)

Lopes, E., *Fundamentos da lingüística contemporânea*, São Paulo, Cultrix, 1987.

Pottier, B., *Sémantique générale*, París, PUF, 1992.

Roget's Thesaurus of english words and phrases, classified and arranged so as to facilitate the expression of ideas and assist in literary composition, London, Longman, 1970.

Sager, J-C., La terminología, puente entre varios mundos. en Cabré, M.T. *La terminología: teoría, metodología, aplicaciones*, Barcelona, Editorial Antártida/Empúries, 1993.

Tálamo, M.F.G.M. y Lara, M.L.G., O campo da Lingüística Documentária, en *Transinformação*, v.18, n.3, 2006, pp. 203-211, [en línea] <http://revistas.puc-campinas.edu.br/transinfo/index.php> [Consulta: 02 de junio de 2007].



# La terminología y los conceptos en las ciencias históricas: una perspectiva interdisciplinaria desde la Ciencia de la Información y la Didáctica de la Historia

MARTHA SABELLI

*Escuela Universitaria de Bibliotecología y Ciencias Afines,  
Universidad de la República, Uruguay*

La ponencia enfoca, en este Primer Simposio Internacional sobre Organización del Conocimiento: Bibliotecología y Terminología, las relaciones interdisciplinarias entre la Ciencia de la Información y la Didáctica en el campo específico de la terminología utilizada en un dominio del conocimiento, la Historia y su aplicación en la enseñanza secundaria. Se centra en la importancia del rol que desempeña la terminología en la enseñanza de una materia, y su esencial contribución en el proceso de construcción y representación de las estructuras cognitivas.

La indagación se inserta en la problemática de los procesos de conceptualización abordados por la didáctica de una disciplina, y más precisamente en los procesos de internalización de los conceptos en la enseñanza de la Historia. La mirada del tema parte de las deficiencias y dificultades surgidas en la apropiación del vocabulario manejado por las Ciencias Históricas y las relacionadas a la asimilación e internalización de conceptos por parte de los educandos.

Paralelamente se destacan la falta de coordinación que existe por parte de los profesores, con respecto a la selección y uso de los conceptos y el vocabulario, y la falta de conciencia sobre el rol relevante que pueden y deben desempeñar éstos en el proceso de enseñanza -

aprendizaje de la materia a lo largo de los programas de estudio. Por ejemplo, en lo que se refiere a la coordinación entre docentes de la asignatura a nivel de los centros o institutos de enseñanza y, en especial, a su seguimiento a través de los diversos cursos para promover una continuidad del proceso de aprendizaje en espiral en el contexto de un plan de estudio.

Esta problemática se basa en la necesidad que existe en el área de enseñanza de la Historia, acerca del manejo por los alumnos de un “tejido base” del conocimiento histórico —un “tejido” de ideas y conceptos básicos— que incluya macro conceptos y conceptos básicos, conjuntamente con informaciones. El vocabulario histórico es un área estudiada desde hace décadas, pero actualmente se ha convertido en un centro de interés en la metodología de la asignatura. Y se da en un amplio marco pedagógico, donde el vocabulario utilizado por los alumnos de la enseñanza secundaria despierta enorme preocupación. Las carencias detectadas en su lenguaje es un dato frecuente en las conversaciones entre docentes. La pobreza del vocabulario común y el mal manejo de vocablos es una realidad fácilmente observable, como lo corroboran varias investigaciones sobre esta situación.

En este sentido, se reconoce tanto el rol desempeñado por los medios de comunicación de masas, como la pérdida de hábitos de lectura y el horario reducido de permanencia de los alumnos en el aula. Este último factor agrava los otros, al ser este espacio la única vía al alcance de un sector apreciable del alumnado para recibir mensajes bien expresados o “codificados”.

Debemos recordar que la terminología de las ciencias históricas, y la que por ende es empleada en su enseñanza, tiene un carácter interdisciplinario (economía, ciencia política, demografía, geografía, derecho, bellas artes, arquitectura...); y más bien, deberíamos preguntarnos qué temáticas de las ciencias sociales, humanidades y arte no están contempladas por los estudios históricos, como también sucede con amplias áreas de las ciencias en general.

En suma, la problemática es relevante porque al docente le resulta imposible avanzar en su estrategia didáctica si el alumno no ha aprendido el vocabulario histórico e interdisciplinario básico exigido por el nuevo aprendizaje. Asimismo debemos considerar que la no interna-

lización de un vocabulario significa el no aprendizaje de determinados vocablos-conceptos; es decir, el no cumplimiento de los objetivos propuestos. En relación con este aspecto observamos en las clases las enormes distancias que separan a unos alumnos de otros, pues algunos permanecen “aislados” por no dominar su propio idioma. La enseñanza de la Historia ratifica lo sostenido sobre las relaciones entre el lenguaje y la conceptualización.

Bajo la perspectiva de este análisis debemos destacar que estamos ante un dominio del conocimiento, la Historia y, más precisamente, en el proceso de enseñanza-aprendizaje de esta disciplina en la educación secundaria, escenario del desarrollo de la Didáctica. El dominio de conocimiento está delimitado aquí por los planes de estudio, los textos y el proceso de enseñanza-aprendizaje. Los actores que interactúan en el escenario de este dominio son los profesores y los alumnos, y también los textos o manuales, que aparecen como los mediadores documentales fundamentales en dicho proceso.

Los profesores de Historia que dictan sus cursos en el grado elegido conforman una comunidad que interactúa con otra comunidad mayor constituida por ellos y los alumnos, la que a su vez se inserta en instituciones educativas y medios sociales en ciertos casos semejantes y en otros muy diversos entre sí. El discurso de estas comunidades es un elemento esencial para analizar y mejorar la interacción entre sus actores, con el fin de obtener una coordinación posible y lograr los objetivos propuestos en los planes de estudio y en cada propuesta institucional y por área.

Se entiende a los profesores como una comunidad que tiene un referente común: un dominio específico del conocimiento: los contenidos y metodologías de la disciplina Historia, enfocada en la enseñanza a jóvenes entre 12 y 14 años. Tal dominio está determinado por ciertas variables vinculadas al proceso de enseñanza-aprendizaje del conocimiento en cuestión, y, como bien señala Hjørland, *el dominio refleja una construcción social*; en este caso en el ámbito educativo y en la disciplina Historia, y centrado en el campo de la didáctica de la Historia.

Por tanto, presentamos una problemática que puede ser investigada por dos disciplinas, la Ciencia de la Información y la Teoría de la Di-

dáctica, insertas en un escenario de “fertilización cruzada”, al decir de Foskett. Allí confluye un rico intercambio de ideas, metodologías de investigación y teorías basadas en los distintos ámbitos de la Ciencia de la Información, la Organización del Conocimiento y los lenguajes documentales, y la teoría de la Terminología, pero también las vinculadas con la propia conformación del campo disciplinar y de su objeto: la información.

La nueva visión de la información desarrollada en las últimas décadas por la Bibliotecología / Ciencia de la Información están muy relacionados con la problemática de la conceptualización. Nos adherimos a los enfoques socio-cognitivos, que han sabido aunar los aportes del cognitismo y constructivismo con las imprescindibles dimensiones sociales y culturales.

La nueva conceptualización de la información se ubica en las definiciones que la presentan como algo subjetivo, cognitivo, y considera las limitaciones del enfoque clásico centrado en una visión de la información como un objeto tangible. Aun el conocimiento registrado, reconocido por Kando (1994) como *información entidad objetiva* o por Hoffman (1993) *información como información documentada*, no está separado del ser humano, sino que es él quien le da sentido. O el construir el sentido (*sense making*) como lo denominó Brenda Derwin nos aporta un invaluable concepto para entender los paradigmas de los “estudios de usuario”. Esto ha llevado a considerar que el conocimiento registrado *no es directamente utilizable, pues se requiere un esfuerzo y la capacidad y destreza para seleccionar, interpretar y adecuar los datos a las necesidades y propósitos de aplicación* (Angulo Maciel, 1996).

Sin duda el constructivismo nos ha dado una mejor visión sobre el usuario y la generación de conocimiento. La información se da en la estructura cognitiva del individuo; la significación es el resultado de un proceso consciente e intencional y, en consecuencia, la calidad de la información dependerá mayormente de cómo se realiza dicho proceso en un contexto determinado donde los mensajes no pueden aislarse de los mediadores culturales.

De allí la importante contribución de las teorías socio-cognitivas y los enfoques holísticos y comprensivos, que conciben al sujeto con

sus pensamientos, emociones, actitudes, hábitos, costumbres y expectativas, y en su contexto familiar, laboral y educativo de su vida cotidiana.

Estos nuevos enfoques implican una nueva conceptualización del documento y su rol en la comunicación (teoría del documento) así como en las funciones que representan los sistemas de recuperación de información y del profesional de información (teoría del procesamiento de la información). Por su parte Vickery (1997) revalorizando el contexto de comunicación humana, inserta la información en los procesos de información humana. El ser humano al interactuar en el medio adquiere conocimientos de su entorno y le transmite información a otros.

#### LOS CONCEPTOS Y LA TEORÍA DEL ANÁLISIS DEL DOMINIO: UN POSIBLE ESCENARIO INTERDISCIPLINARIO

En este marco ubicamos el paradigma sociológico-epistemológico basado en la necesidad de hacer un “análisis de dominio” (*domain-analytic*), en el que se toman en cuenta las dimensiones socio-culturales; es decir, el ambiente de uso de la información. En este sentido el abordaje de Hjørland junto al de Albrechtsen, orientan la unidad de estudio desde el nivel individual hasta el social, el disciplinario y el de las comunidades de conocimiento, considerando que hay una relación dialéctica entre la comunidad y sus miembros. La perspectiva sociocognitiva que exponen se relaciona con la teoría de la actividad, la que provee un marco de estudio para el comportamiento humano y cognitivo.

El paradigma sobre el análisis del dominio difundido por Birger Hjørland, de la Royal School of Librarianship de Copenhague, nos ofrece un enfoque enriquecedor para la investigación al centrar sus estudios e investigaciones en una revisión de los paradigmas existentes en la Ciencia de la Información sobre los enfoques de los subjects (temas), y su relación con los documentos y los usuarios y para formular la teoría del análisis del dominio. Hjørland ha abordado sus principales ideas a través de cuatro documentos donde las expone y discute (Hjørland 1992, 1995, 2001 y 2002). La lectura y el análisis de estos

documentos nos aportaron una perspectiva crítica y un conjunto de enfoques sobre el papel de las comunidades y sobre el conocimiento en un marco epistemológico definido. De este modo fue posible conjugar sus presupuestos con una mejor comprensión y ahondar en las bases teóricas del estudio en cuestión.

De cara a esta profundización interesa destacar a continuación las principales ideas para apoyar esta presentación:

A. La teoría del dominio como enfoque epistemológico en el contexto de la Ciencia de la Información:

- necesidad de insertar la Ciencia de la Información en una visión social y en este caso, el análisis del dominio en un paradigma social;
- posicionamiento en un enfoque materialista / realista, rechazando visiones reduccionistas, en especial, el punto de vista empírico-positivista;
- visión socio-cognitiva del análisis del dominio;
- enfoque no sólo filosófico-realista sino también funcionalista que apunte a estudiar las funciones implícitas y explícitas de la información y la comunicación;
- relación de las corrientes socio-cognitivas de la Ciencia de la Información con teorías de las ciencias cognitivas y ciencia de la educación para permitir un enfoque transdisciplinario.

B. El subject (tema) y el dominio en el contexto de las comunidades:

- concepción del dominio del conocimiento como pensamiento o discurso de las comunidades inserto en la división del trabajo de la sociedad;
- definición del subject (tema) del documento y de sus potencialidades epistemológicas; los documentos tienen propiedades objetivas, con potencialidad cognitiva (o informativa);
- propiedades de los documentos los cuales emergen del uso de ellos, por ejemplo, leyendo el documento en relación con una actividad específica pueden surgir distintos significados del documento para distintos propósitos;

- concepción del lenguaje documental en un marco de la división lingüística que hay en la sociedad según la filosofía de “la división internacional del trabajo”.

Las ideas manifestadas por Hjørland en sus afirmaciones sobre la teoría del dominio como enfoque epistemológico en el contexto de la Ciencia de la Información, la visión social sobre todo su defensa de una perspectiva socio-cognitiva, permiten ver los objetivos y resultados del trabajo de selección de conceptos y vocabulario a la luz de estas ideas. La teoría crítica de la enseñanza, que descansa sobre principios de las corrientes de la acción participativa (Carr y Kemmis 1988), los enfoques sobre el currículo y su necesaria relación con el contexto (Gimeno Sacristán, 1989), y los aportes de Pichon Riviere (1984) como los relativos al campo más específico de la conceptualización (Novak, 1982; Novak y Gowin, 1988), manifiestan la misma intencionalidad de insertar la investigación dentro de la relación del proceso de enseñanza-aprendizaje con el medio social. Y más aún, la situación-problema que genera el estudio: las desigualdades existentes entre los alumnos en cuanto al acceso y uso de conceptos y terminología en general, están vinculadas a las distintas realidades sociales y culturales que éstos viven.

#### LOS RESULTADOS DE UNA INVESTIGACIÓN DESDE LA PERSPECTIVA DE LA TEORÍA DEL DOMINIO

El caso que presentamos se basa en una investigación que desarrollé durante un año sabático como profesora de Historia de enseñanza secundaria en Uruguay en 1990-1991. La formación y ejercicio profesional en el área de la enseñanza de la Historia (1972-1990) y didáctica de la asignatura en el Instituto de Profesores “Artigas” en los años 90, y también en el área de la Bibliotecología / Ciencia de la información, hizo posible una interdisciplinariedad y complementariedad en los abordajes y metodologías aplicadas.

El estudio se publicó con el título: *Los conceptos y el vocabulario en la enseñanza de la Historia* (Sabelli, 1991), y se centra en el plano conceptual y terminológico de un determinado “dominio”. Se conside-

ra que los conceptos y el vocabulario son componentes muy relevantes en el proceso de enseñanza-aprendizaje y constituyen instrumentos imprescindibles para coordinar la selección de centros de interés dentro de la asignatura, y con las afines. Como ya mencionamos, en los años 80 las nuevas tendencias metodológicas de la enseñanza en general, y especialmente del área de la enseñanza de la Historia, abrieron nuevas perspectivas en esta temática y la insertaron en un marco teórico que sirvió de referente para fundamentar la propuesta.

La investigación surge del interés por crear un nuevo instrumento que puedan utilizar los profesores de Historia del CBU (Ciclo Básico Único), basándose en los distintos niveles de alumnos y realidades sociales, y busca aportar un elemento que permita coordinar la selección de objetivos y contenidos. La fundamentación teórica conceptualiza el proceso de enseñanza-aprendizaje con ayuda de la teoría de la “constelación de ideas”, del aprendizaje de conceptos inclusores y la incidencia del lenguaje, y de la creatividad y la afectividad.

La situación-problema central de la investigación es la dificultad de coordinación que existe entre los docentes y los programas de estudio en el plano de la conceptualización de la disciplina y su contenido. La experiencia docente y el análisis del proceso de enseñanza-aprendizaje de la disciplina Historia en los primeros cursos (1° y 2° del Ciclo Básico) de enseñanza secundaria en los inicios de los años 90, indicaban una ausencia de coordinación en cuanto al manejo de los conceptos y del vocabulario por parte de los profesores y, lo que era más grave, importantes deficiencias y carencias en este plano por parte de los alumnos. En suma, una situación no resuelta ni abordada.

En este contexto la investigación sobre el proceso de aprendizaje de conceptos aparece como una de las áreas de estudio más relevantes sobre la enseñanza de cada asignatura del currículo.

La parte medular de la investigación consiste en elegir los macroconceptos o ideas fundamentales, los conceptos básicos y el vocabulario secundario para los dos primeros cursos de Historia del CBU. El listado de términos seleccionados se presenta en tres formas: 1. un listado por curso (primero y segundo curso de Historia del CBU) con términos jerarquizados a tres niveles: macroconceptos o ideas fundamentales (primera columna), conceptos básicos (segunda columna),



vocabulario secundario (tercera columna); 2.un listado alfabético de todos los términos con los códigos que permiten mostrar un seguimiento en los dos cursos; y, 3.un listado de macroconceptos o ideas fundamentales.

El trabajo incluye consideraciones sobre las estrategias de la instrucción en la enseñanza-aprendizaje de conceptos básicos y algunas conclusiones y perspectivas para continuar la investigación.

Con respecto a la metodología empleada, la primera etapa de la investigación consiste en la selección y estudio de bibliografía sobre las principales corrientes pedagógicas con el fin de ubicar el trabajo en un contexto teórico. Asimismo se consultaron obras sobre aspectos relacionados con el aprendizaje de conceptos y el lenguaje en general. En una segunda fase se analizaron los programas de Historia de los dos primeros cursos del CBU en 1990 y se identifican las principales unidades didácticas y macroconceptos. Luego se conformó una muestra de profesores de historia que se entrevistarían, quienes dictaban sus clases en los cursos mencionados y se elaboró un cuestionario. En la selección de profesores se buscó cubrir distintos niveles socioeconómicos y culturales del alumnado y una localización diversa de los cursos (Montevideo e interior del país). Gracias a las entrevistas se recopilaban datos sobre la opinión de los profesores sobre el uso del vocabulario, las metodologías empleadas para su estudio, los distintos niveles de su manejo por parte de los alumnos y las dificultades detectadas.

Paralelamente se desarrolló un arduo trabajo de “rastreo” de los términos en las fuentes consideradas más pertinentes: vocabularios históricos, otros vocabularios en ciencias sociales (sociología, ciencia política, demografía), manuales de historia uruguayos y extranjeros (españoles y franceses). Los textos de estudio seleccionados correspondían a los utilizados a fines de los años 80 en Uruguay, España y Francia. Totalizaron treinta libros, lo que implicó una difícil búsqueda, y para sistematizar la tarea se empleó un sistema de codificación.

Una primera versión del vocabulario se elaboró con base en los términos usados en nuestro trabajo docente y los extraídos de las fuentes de referencia. Luego se cotejó con los listados entregados por algunos profesores y se constató la unidad de criterios existentes entre ellos y la coherencia con manuales y fuentes nacionales y extranjeros.

El vocabulario incluye los términos jerarquizados por curso según unidades conceptuales, pero diferencia ideas fundamentales, conceptos básicos y el vocabulario secundario relacionado. También se desarrolló un “seguimiento” de los conceptos básicos y el vocabulario secundario del primero al segundo curso y se estableció una relación de continuidad del trabajo docente en ambos años.

Los conceptos seleccionados constituyen unidades de conocimiento, se relacionan entre sí con relaciones de dependencia, y se conforman grupos de conceptos. De esta manera se distribuyeron los conceptos agrupados en relaciones jerárquicas de tres niveles: macroconceptos, conceptos básicos y vocabulario secundario.

De este modo pudimos ver la estructura presentada —siguiendo a Dubuc (1999)— como un “árbol de domaine” (árbol de dominio). El diseño muestra relaciones de dependencia que pueden “leerse” en forma horizontal, de izquierda a derecha. Es una manera funcional de agrupar los conceptos de acuerdo con el parentesco encontrado, en niveles de conceptualización mayor. A la izquierda está el mayor nivel de abstracción (macroconceptos o ideas), en una segunda columna están los conceptos básicos, y la tercera columna pertenece al vocabulario más general relacionado con los conceptos básicos que contienen menor nivel de conceptualización pero que constituyen “puertas de entrada” para poder “aprehender” los conceptos.

El total de los términos es de 798, de los cuales 80 son macroconceptos, 287 conceptos básicos y 431 vocabulario general. Estas cifras ayudaron a los docentes a tomar conciencia de la dificultad por parte de los alumnos, a manejar un nivel amplio de términos, gran parte de los cuales no eran utilizados ni conocidos por ellos.

## CONCLUSIONES.

Con esta presentación se ha buscado describir esta investigación en el área de la terminología en ciencias sociales y, más específicamente, los conceptos y vocabularios que existen en el área de los conocimientos históricos a la luz de las principales ideas de la teoría del análisis del dominio de B. Hjørland. En función de lo expuesto en la última parte se puede concluir que las coincidencias en el enfoque

de la propuesta de investigación descrita con la teoría del análisis del dominio radican en los siguientes aspectos:

1. La investigación terminológica identifica y trabaja en una comunidad del conocimiento específica: docentes de Historia de los dos primeros cursos del ciclo básico de enseñanza secundaria, y encuentra coincidencias entre ellos en la selección y uso de conceptos y vocabulario relacionado con el contenido abarcado (Prehistoria, Época Antigua, Edad Media y Época Moderna). Asimismo, a lo largo de la investigación se comprueba la preocupación por coordinar en el ciclo básico el nivel de los conceptos, para facilitar un proceso de enseñanza-aprendizaje continuo por parte de los alumnos, más allá de la diversidad de los profesores y las particularidades de cada subcomunidad existente. Hjørland insiste en el rol que juegan las comunidades en los dominios y la necesidad de contar con un enfoque sobre el “colectivo” en las investigaciones.

2. El estudio busca un enfoque socio-cognitivo del rol de los conceptos y el vocabulario relacionado, basado en teorías constructivistas y de la investigación-acción. Las primeras inciden en la forma de definir y ubicar los conceptos en el proceso de construcción y estructura del conocimiento de los alumnos; el segundo, en la visión que tienen el docente y los alumnos en una comunidad que actúa, investiga y revisa sus resultados para reformularlos y seguir la acción del proceso de enseñanza-aprendizaje. Y como aspecto central de la investigación-acción, se busca la existencia de un mediador fundamental: su propio discurso, conformado por conceptos y demás términos relacionados.

3. Una tercera coincidencia es la definición, según Hjørland, del *subject* (tema) como las potencialidades epistemológicas del documento. La investigación “interroga” el corpus de textos (textos de enseñanza de Historia) para identificar sus conceptos básicos y secundarios. Las “respuestas” encontradas tras la revisión y fichaje realizado muestran grados muy altos de potencialidades semejantes, lo que confirma la hipótesis manejada: es posible que los conceptos puedan coordinar la didáctica de la asignatura. Los textos evidencian ser mediadores valiosos en este dominio de la disciplina y en el ejercicio profesional del área educativa.

4. Por último, defender un paradigma social fundamentado por Hjørland es parte fundamental de la base teórica de la investigación.

Para concluir, los conceptos y las relaciones conceptuales son un elemento básico en la recuperación de información. Actualmente existen corrientes revisionistas relacionadas con la estructura de los tesauros y demás lenguajes documentales tradicionales que buscan mejorar los sistemas automatizados. El trabajo aquí presentado surgió de la inquietud por encontrar representaciones de los dominios del conocimiento que mejor reflejen el saber de los autores y los usuarios, como sus intencionalidades e intereses.

## BIBLIOGRAFÍA

- Angulo Maciel, N. (1996), "Información: una nueva propuesta conceptual", en *Ciencias de la Información*, vol. 27, Núm.4, pp.190-194.
- Astolfi, Jean P. (2001), *Conceptos clave en la didáctica de las disciplinas*, Sevilla, DIADA.
- Carr, W. y S. Kemmis (1988), *Teoría crítica de la enseñanza: la investigación-acción en la formación del profesorado*, Barcelona, Martínez Roca.
- Díaz Barriga, Ángel (1985), *Didáctica y currículo: convergencia en los programas de estudio*, México, Nuievomar.
- Dubuc, R. (1999), *Manual de terminología*, Santiago, Chile, RIL.
- Gimeno Sacristán, José (1989), *Teoría de la enseñanza y desarrollo del currículo*, Salamanca, Anaya.
- Hjørland, B.(2002), "Epistemology and the socio-cognitive perspective in information science", en *Journal of the American Society for Information Science and Technology*, vol. 53 Núm. 4, pp. 257-270.

### *Terminología en diferentes ámbitos de conocimiento*

Hjørland, B. (2001), "Toward a theory of aboutness, subject, topicality, theme, domain, field, content ... and relevance", en *Journal of the American Society for Information Science and Technology*, vol. 52, Núm. 9, pp.774-778.

Hjørland, B. (1998), "Theory and metatheory of information science: a new interpretation", en *Journal of Documentation*, vol. 54, Núm. 5, pp.606-621, dec.

Hjørland, B. y H. Albrechtsen (1995), "Toward a new horizon in information science: domain-analysis", en *Journal of the American Society for Information Science*, vol. 46, Núm. 6, pp.400-425.

Hjørland, B. (1992), "The concept of "subject" in information science", en *Journal of Documentation*, vol. 48, Núm. 2, pp. 197-253.

Hoffmann, E. (1993), "Defining information: an analysis of the information content of document", en *Information Processing & Management*, vol.16, Núm. 291-340.

Ingwersen, P.; B. Larsen, y; E. Noyons. (2001), "Mapping national research profiles in social science disciplines", en *Journal of Documentation*, vol. 57, Núm. 6, pp.715-740.

Kando, N. (1994), "Information concepts reexamined", en *International Forum of Information and Documentation*, vol.19, Núm. 2, pp.20-24.

Litwin, Edith (1997), *Las configuraciones didácticas*, Buenos Aires, Paidós.

Novak, Joseph (1982), *Teoría y práctica de la educación*, Madrid, Alianza.

Novak, Joseph y D.Bob Gowin (1988), *Aprendiendo a aprender*, Barcelona, Martínez Roca.

Pichon Riviere, Enrique (1984), *El proceso grupal a la psicología social*, México, Nueva Visión.

Sabelli, M. (1991), *Los conceptos y el vocabulario en la enseñanza de la Historia : propuesta para los dos primeros cursos del C.B.U.*, Montevideo, Ediciones de la Banda Oriental.

Sabelli, M. (1989), *Lista de descriptores sobre urbanización y vivienda*, Montevideo, CIESU

Shapere (1984, "*Reason and the search for knowledge. Investigations in the philosophy of science*", Dordrecht, D.Reidel [citado por Hjørland, B.; Albrechtsen, H (1995), y Hjørland, B.(2001)].

Timpka,T. (1995), "Situating clinical cognition. Artificial intelligence in medicine", vol 7, pp.387-394, [citado por Hjørland, 2002].

Vergnaud, Gérard (1990), "La teoría de los campos conceptuales", en *Recherches en Didactique des mathématiques*, Vol.10, Núm. 2-3, pp133-170 ,(Traducción Juan D. Godino).

Vickery,B. (1997), "Metatheory and information science", en *Journal of Documentation*, vol. 53, Núm., 5, pp.457-484.

**SECCIÓN II**  
**Análisis terminológico**





# El descriptor y el término. Los conceptos y la lingüística

ANA MARÍA CARDERO  
*Facultad de Estudios Superiores Acatlán,  
UNAM, México*

## 1. INTRODUCCIÓN

a) Para desarrollar el tema que nos ocupa nos acercaremos al perfil de la Terminología, para posteriormente ubicar qué tipo de relación guarda con la Documentación, concientes de la complejidad de la tarea. Nuestro acercamiento al tema será desde la Lingüística aplicada y la Terminología.

La Terminología desde una perspectiva comunicativa es una disciplina completa en tanto que es teórica y aplicada; y los términos son unidades semántico-formales, y por tanto una de las áreas en las que se fundamentan es la Lingüística. Además, como los términos vehiculan conocimiento, otra área de relación son las ciencias cognitivas; por último, en tanto los términos sirven para la transferencia del conocimiento, se relacionan también con las Ciencias de la Comunicación.

Para Cabré (1997) la Terminología es una interdisciplinar, y él sostiene que: “Como una interdisciplina no define su campo de estudio como una adición lineal de los conceptos precedentes, sino que selecciona de las mismas un determinado número de conceptos y elemen-

tos, y posteriormente elabora a partir de esos conceptos un objeto y un campo propio”.

El término interdisciplina está etimológicamente constituido por una base, la palabra española “disciplina” que procede del mismo sustantivo latino *disciplina*. En la lengua española tal término es polisémico y una de sus acepciones hace referencia a ‘una rama del conocimiento’. Además la palabra que nos ocupa está formada también por el prefijo *inter* (preposición), que significa entre, en medio de, entre una cosa y otra. Esto, en general, implica que se trata de una relación entre dos elementos o más, por lo que interdisciplinar debe entenderse propiamente como “entre disciplinas”, considerando que se habla de relaciones que existen entre una disciplina y otra u otras, y no de relaciones internas dentro de una sola disciplina, en cuyo caso correspondería el uso de la preposición *intra*, que significa ‘dentro de, en el interior de’.

Pero al mismo tiempo la Terminología es una transdisciplina, donde la preposición *trans*, significa ‘más allá de’, ‘del otro lado de’, o bien, ‘a través de’, o ‘por encima de’. Como la preposición *trans* es polisémica, nosotros fijamos su significado como ‘a través de las disciplinas’, con base en varias razones: no hay ninguna materia especializada sin terminología; no se puede comunicar la especialidad sin terminología; no hay Terminología sin materia especializada.

Así, vemos que la Terminología mantiene con disciplinas diferentes, distintos tipos de relaciones a veces internas y otras externas. Las externas son de reciprocidad e intercambio, como ocurre con la Informática y la Documentación; en este punto coinciden la Terminología y la Documentación.

- b) Dentro del área de conocimiento de la Lingüística hemos ubicado a la Terminología dentro del subconjunto de la Lexicología; es decir, que la revisamos en relación con los parámetros e instrumentos que propone la Lingüística. Forman con ella un grupo la Lexicografía y la enseñanza de lenguas extranjeras según propone la lingüística aplicada.

De manera general, la Terminología es, en teoría y práctica, una disciplina que permite identificar el vocabulario de una especialidad;

es decir, los términos o unidades léxicas que adquieren valor como términos cuando se usan en un determinado contexto o situación especializada. Analizar el vocabulario y si es necesario crearlo entre el especialista y el terminólogo, además de normalizarlo en una situación concreta de funcionamiento con la finalidad de responder a las necesidades de expresión de sus usuarios.

La Terminología maneja unidades que son a la vez lenguaje y conocimiento, y su objetivo central es transferir ese conocimiento una vez organizado, el cual aparece como unidad de información en los textos o en el discurso, por lo que se concibe como un documento en el que la lengua, una vez organizada y transformada de acuerdo con las necesidades de la disciplina de la que se trate, se presenta a los usuarios.

La Documentación tiene como uno de sus objetivos la representación sintética del conocimiento que se transmite en un documento y éste se relaciona con la terminología en textos especializados. Mediante el proceso de *indización* los textos se reducen a marcas de significado que en general son términos. Entre la Terminología y la Documentación se consolida tanto una práctica interdisciplinar como transdisciplinar.

- c) El elemento que permite esta relación es la lengua. En este caso con un componente que marca la especialización y sus condicionamientos lingüísticos. Este material lo reutiliza la documentación, la que actualiza su práctica al establecer los procesos de gestión de información y agilizar las tareas documentales en estrecha colaboración con los recursos generados en la Terminología.

El estado de la lengua que presenta esta ciencia es un reto para la Documentación y hay que entender, en el caso mexicano, y creemos que en general también, que la Terminología es un reto para sí misma.

- d) En párrafos anteriores hemos señalado que los términos son unidades léxicas que adquieren valor como términos cuando se usan en un determinado contexto o situación especializada. El terminólogo trabaja con documentos que se producen en los

contextos de las especialidades, revisa los manuales, entrevista a los especialistas y selecciona, de acuerdo con estos principios, los términos adecuados para elaborar los glosarios terminológicos, es decir, que requiere de la Documentación para poder formar sus terminologías, como dice Cabré (1999), los requiere para 'existir'. Asimismo la Documentación se sirve de la Terminología, especialmente la normalizada, para describir o representar el contenido de los documentos.

## 2. CAUSAS POSIBLES DE UNA RELACIÓN COMPLEJA ENTRE LA TERMINOLOGÍA Y LA DOCUMENTACIÓN

Acercarse a los textos especializados no es tarea fácil, y llevan a cabo esto, entre otras ramas del conocimiento, la lingüística del texto especializado. Ésta tiene interés en aportar teóricamente algunos principios para poder caracterizar textos en una tipología del discurso especializado.

También a la ingeniería lingüística le conciernen los vocabularios especializados precisamente para desarrollar sistemas automáticos de reconocimiento de términos. Esta área de estudio está desarrollada tanto teórica como prácticamente, temas que finalmente tienen como meta la tipologización, como sucede en la tipificación del discurso especializado en la ingeniería computacional, y la efectividad para localizar automáticamente términos en los textos especializados, que finalmente tendrá que utilizar la documentación para describir o representar el contenido de los documentos.

La lingüística del texto (Ciapuscio 2002) nos presenta distintas propuestas de tipologización sobre los textos especializados y hemos seleccionado una sobre el tema de la *angiogénesis*.

Ciapuscio ilustra su propuesta para mostrar sus avances, con base en un artículo de ciencia e investigación, una entrevista y una noticia de divulgación científica con el tema del tipo de cáncer señalado. Propone cinco aspectos a los que se debe atender en cada uno de los textos que examina:

- Nivel I. Tipos de función: expresar/señalar/informar/dirigir.
- Nivel II. Tipos de situación: comunicación externa/comunica-

ción entre ámbitos y tipo de interlocutores. La relación puede ser externa (ciencia y medios), con un especialista o semi-lego, individual o colectiva y por medio de la comunicación escrita.

- Nivel III. Procedimientos temáticos y formas textuales primarias y derivadas. Para el especialista es obligatorio explicar y definir términos. En este caso la *angiogénesis*. En una entrevista el tema se puede desplegar con el equivalente del término. En este caso *sarcoma de Kaposi/angiogénesis* y en el artículo de divulgación será necesario explicar y definir los términos
- Nivel IV. Estructuración lingüística. Forma con la que puede aparecer o aparece el término: *sarcoma de Kaposi/angiogénesis*.
- Nivel V. Formulación. Tratamiento de la terminología y variación formal.

Ahora bien, cómo se involucra el documentalista en estas situaciones comunicativas. Su función es informar. Su situación comunicativa es externa. Debe facilitarle al usuario información sobre el tema. En la medida de sus posibilidades debe darle las equivalencias del término (sinonimia), y en el último nivel ubicar el tratamiento de la terminología y su variación formal (sinonimias y paráfrasis).

Como se puede observar, en este tema estamos involucrados tanto los lingüistas como los documentalistas en la interdisciplina y en la transdisciplina.

El documentalista emplea la lengua en el proceso de gestión e información. Con ella explora y explota la documentación que debe ofrecerle al usuario. Los documentos que contienen la información son lingüísticos. El documentalista como intermediario debe acceder al contenido de los documentos por medio de sus conocimientos léxicos, sintácticos, semánticos y morfológicos.

En nuestra responsabilidad como lingüistas consideramos que tenemos que hacer mucho trabajo en diferentes áreas; es decir, trabajar a profundidad para facilitar la comunicación.

Entendemos que en ocasiones el procesamiento de las terminologías no resulta transparente el comportamiento de los términos desde una perspectiva lingüística, para que los documentalistas operen con confianza en relación con los aspectos semántico-formales. A los pro-

blemas que en general pueden despistar al documentalista vamos a dedicar el siguiente inciso.

### 3. LAS LABORES DE LA LINGÜÍSTICA Y LA TERMINOLOGÍA

La Terminología no ha sido estudiada y revisada de manera especial por la Lingüística Hispánica.

Desde 1983 el Seminario de Lexicología, Lexicografía y Terminología, ha estudiado a esta última con un marco teórico propio de la Lingüística y básicamente desde una perspectiva neológica. Nuestro análisis parte siempre de la Documentación. Nuestro primer acercamiento a la materia que vamos a trabajar nos relaciona con los documentos propios de la materia de que se aborda. Nuestra labor incide en diferentes puntos de búsqueda. En dos de nuestros trabajos, para el trabajo del vocabulario de la cinematografía en México y en el trabajo relacionado con el control de satélites ha sido fundamental el bibliotecólogo que nos atendió. En el primer caso nos buscó bibliografía que resultó fundamental para nuestro trabajo, y en el segundo nos orientó por medio de un glosario y la forma en la que éste se había realizado: por medio de traducciones y equivalencias en otras lenguas. Pero como se dio cuenta que no era la realidad de la profesión sobre lo que buscábamos información, nos dirigió con los especialistas, y de ahí surgió un trabajo sobre la terminología de control de satélites en México.

En las primeras etapas de su trabajo el terminólogo depende de la documentación que nos muestre organizada la bibliografía sobre el tema y de los tesauros que encuentre relacionados con el tema.

En una segunda etapa busca a los especialistas para confirmar la vigencia de la temática que pretende trabajar y la forma de enriquecerlo con entrevistas con los usuarios. Luego hace un primer intento de organización del conocimiento en un árbol de dominio estructurado por áreas y subáreas temáticas.

La etapa de organización del conocimiento resulta especialmente delicada y la confrontamos cuando definimos los términos junto con los especialistas. Es ésta una labor interdisciplinaria en la que vamos reajustando el conocimiento en áreas y subáreas temáticas de las que nos estamos haciendo cargo.

Cuando estamos trabajando con los aspectos semánticos del material procedemos a identificar los procesos de sinonimia, polisemia y homonimia que nos salen al paso en la organización. Se evita la sinonimia, y se inclina la clasificación hacia la homonimia. Éste es un problema para el documentalista: organizar la documentación de la materia que se estudia.

¿Pero qué pasa con la polisemia? Ésta se mueve de la lengua general a la especializada cuando a una palabra del habla se le da un significado neológico que la especializa y la hace formar parte de una de las áreas de conocimiento de la terminología. En ocasiones observamos que el término también pertenece a otras áreas de conocimiento.

El asunto de la polisemia es un proceso muy frecuente y que no es deseable en la lengua especializada. En las estadísticas que manejamos el porcentaje mayor de términos es el que clasificamos de sentido lato, el cual puede provenir de la lengua estándar o de otras áreas de conocimiento. Al integrar los términos a un árbol de dominio de una especialidad ya les hemos otorgado un significado polisémico especializado que corresponderá a esa área o subárea temática. Esto ocasiona que distintas áreas temáticas de conocimiento compartan el mismo término, que al ser integrado a un árbol de dominio adquiere su valor de término especializado de x área de la materia que se esté trabajando en el momento.

El porcentaje de términos estrictamente pertenecientes a una disciplina es más bajo en algunas áreas. De ahí que conformemos nuestro léxico de términos en *sentido estricto*; es decir, con los términos que sólo pertenecen a esa disciplina, y en *sentido lato* los que comparten varias disciplinas o que provienen de la lengua general.

Tradicionalmente se ha considerado que los vocabularios especializados están constituidos en su mayoría por términos con formantes cultos grecolatinos. Los entramados lingüísticos de los vocabularios especializados son complejos porque están sometidos a procesos morfológicos o de formación que no corresponden a las expectativas, como es el hecho de que los términos generalmente aparezcan con formantes cultos grecolatinos. Desde una perspectiva morfológica lo que hacemos es revisar su constitución y los morfemas que la integran. De la revisión efectuada con bases de datos enormes, por ahora,

hemos concluido que los términos técnicos y científicos monoléxicos no están todos integrados por formantes cultos; es decir, que en su mayoría no lo están, sobre todo en la tecnología, como sin embargo ha asegurado la Terminología hasta ahora.

En relación con los aspectos sintácticos, la expansión de los sintagmas ha dado como resultado comportamientos muy interesantes, como es el de la expansión de los términos multiverbales alrededor de un mismo núcleo, para denominar procesos específicos o matizarlos, como en el ejemplo que señalamos a continuación en relación con el núcleo *subsistema* que tiene el significado de ‘elemento con una misión propia dentro de un sistema’, en la terminología de control de satélites:

- subsistema de rango (RSS)
- subsistema de propulsión (PSS)
- subsistema de plataforma
- subsistema de estructura
- subsistema de comunicaciones
- subsistema de comando (CSS)
- subsistema de carga útil
- subsistema de antena omni
- subsistema de energía eléctrica (EPS)
- subsistema de control térmico (TCS)
- subsistema de estado y control

A lo anterior hay que agregar un aspecto que nos resulta de gran complejidad que es el de los procesos de acortamiento de los términos, su abundancia. Éstos tradicionalmente son conocidos por la gramática como siglación y se integran en el grupo: acrónimos abreviaturas, siglas, inicialismos. Este tipo de clasificación ya no opera para la categorización de los términos, hemos tenido que trabajar en propuestas que nos permitan operar, Cardero (2006). Por un uso económico de la lengua los usamos y en ocasiones abusamos de ellos y no especificamos el tema que tratamos ni el área a la que pertenece. Una letra del alfabeto como la consonante *K* la pueden compartir hasta 35 áreas especializadas distintas.



Como vemos, con estos comportamientos las disciplinas guardan en sí mismas un grado de complejidad importante que colisiona con los objetivos de la Documentación. El documentalista dedicado a textos científicos y tecnológicos tiene que ser sensible a estos procesos para superarlos.

#### CONSIDERACIONES FINALES:

La Terminología es una asignatura pendiente de la Gramática, en la que percibimos algún interés en estos momentos porque se ha hecho consciente de que los vocabularios especializados son parte de la lengua en el subconjunto del léxico, y que ella no ofrece pautas adecuadas para la formación de términos.

Las líneas futuras de investigación deberán orientarse a concretar la elaboración de herramientas documentales a partir de la reutilización de los recursos lingüísticos, para consolidar así una práctica tanto interdisciplinar y como transdisciplinar existentes entre la Lingüística y la Documentación.

Otra solución, además de los problemas a los que nos debemos abocar tanto lingüistas como bibliotecólogos, puede ser la de tratar de sensibilizar a quienes elaboran textos científicos y tecnológicos, en el sentido de hacer más amigables los textos; la inclusión de glosarios nos facilitaría mucho la vida tanto a los documentalistas como a los terminólogos.

Y como tema urgente para ambas disciplinas: hay que crear índices temáticos, los cuales son de gran utilidad.

#### BIBLIOGRAFÍA:

- Arano, Silvia, *La ontología: una zona de interacción entre la Lingüística y la Documentación* [on line], Hipertext.net, núm. 2, 2003, <http://www.hipertext.net/web/pag220.htm#La%20lingüística%20y%20la%20documentación>

***Primer Simposio Internacional sobre Organización...***

Cabré Castellví María Teresa (1999), *La terminología. Representación y comunicación*, Barcelona, Institut Universitari de Lingüística Aplicada Universitat Pompeu Fabra.

Cardero García, Ana María (2003), *Terminología y procesamiento*, Universidad Nacional Autónoma de México, Escuela Nacional de Estudios Profesionales Acatlán, México.

Cardero García, Ana María (2006), “Abreviaturas, acrónimos, iniciales, siglas y símbolos en los vocabularios especializados. Una propuesta”, en: *Revista Debate Terminológico* no. 2 07/2006, ISSN 1813-1867.

Ciapuscio Guiomar, Otañi Isabel (2002), “La noción de esquema y la descripción del significado”, en: *Terminología, desenvolvimiento e identidad nacional. VI Simposio Ibero-americano de Terminología*, Lisboa, Edições Colibri, pp. 369-399.

## Posibilidades y límites del análisis cuantitativo de corpus especializados

ELENA BOGOMILOVA LOZANOVA  
*Centro de Estudios Lingüísticos y Literarios,  
El Colegio de México*

**L**as aplicaciones de los corpus textuales en terminología son múltiples y permiten conocer cómo se produce, transmite y enseña el conocimiento especializado, y también identificar unidades simples y complejas de carácter terminológico y su uso en textos especializados, detectar procesos de creación neológica y su grado de productividad, para mencionar sólo algunas. Los datos aquí obtenidos son la base para elaborar y actualizar bancos de datos terminológicos y otras herramientas de gran utilidad para los especialistas del área y los profesionales de la lengua, la informática, la información y la documentación.

Este trabajo enfoca el análisis cuantitativo de los corpus especializados y presenta una valoración de las posibilidades y los límites de este tipo de análisis. Con el fin de permitir la verificación de hipótesis sobre procesos y fenómenos lingüísticos, el estudio cuantitativo debe basarse en un corpus textual representativo que debe tener lo siguiente:

- una gran diversidad de textos que asegure la aparición del mayor número de términos del ámbito de conocimiento especializado en estudio, es decir, que permita la entrada de temas muy diferentes

- una adecuada estratificación de los textos que permita obtener buenos resultados en el campo de la dispersión y en el uso estadístico
- una longitud suficiente de los textos, que permita la recuperación del significado global en que aparezcan los términos.

La aplicación de estos criterios a textos del derecho ambiental mexicano le permitió a la autora conformar un corpus especializado con las características requeridas que comprende 83 documentos de esta área jurídica correspondientes a 576,689 palabras. El propósito de la constitución de ese corpus fue identificar los procedimientos de formación de términos en el ámbito jurídico-ambiental. Los resultados obtenidos en el estudio sistemático de los rasgos cuantitativos del corpus del derecho ambiental constituyen una base objetiva para el análisis cualitativo de índole morfosintáctica y semántica de los candidatos a términos de este ámbito especializado. Hay que resaltar que el acercamiento a esta parte del léxico es sólo el primer paso del análisis. Los datos obtenidos son señales de la probabilidad de que un vocablo adquiera valor especializado y se convierta en término; estos indicios deben de ser validados mediante el estudio cualitativo.

Lo que se deseaba obtener del análisis cuantitativo del corpus de datos era:

- a) Un número elevado de unidades léxicas probables candidatas a términos.
- b) Una base objetiva de selección de aquellas unidades léxicas posibles unidades terminológicas de carácter sintagmático.

Con vista en estos objetivos, se elaboraron dos hipótesis que fundamentan los criterios cuantitativos de identificación terminológica y que guían el análisis:

1. Un vocablo determinado tiene más probabilidades de ser un término técnico cuantas más veces aparezca en un texto reconocido como técnico y, en comparación con un corpus de lengua usual, tenga una mala dispersión en la población léxica,

sesgada precisamente por el carácter técnico del texto en que se encuentre,

2. Un sintagma determinado tiene más probabilidades de ser una expresión terminológica en cuanto:

- a) Todas las palabras que parezcan constituir la expresión terminológica tengan una frecuencia de aparición notablemente alta (sin que se puede decir, *a priori*, qué tan alta) en un texto técnico,
- b) La frecuencia de aparición de toda la expresión sea más alta de lo que se podría encontrar en un corpus de la lengua general.

## ANÁLISIS

Para investigar estas hipótesis se procedió a estudiar las fuentes primarias del derecho ambiental, —leyes, reglamentos y normas oficiales mexicanas—. cuantificando con el programa KWIC la totalidad de los tipos de vocablos<sup>1</sup> que aparecen en éstas. Como resultado se obtuvieron 23,535 tipos de vocablos. De acuerdo con la frecuencia total de estos tipos, el conjunto se dividió en dos grupos; el primero contiene los tipos que aparecen por lo menos dos veces en el corpus jurídico-ambiental y abarca 14,138 tipos, el segundo grupo está conformado por 9,397 tipos de vocablos de una sola aparición.

---

1 La unidad de lengua que me interesa en el análisis es el vocablo al que corresponde, como sintetiza Lara, (*Investigaciones lingüísticas en lexicografía*, México: El Colegio de México, 1979, p.12) la ocurrencia en el habla y el tipo en la  $\Sigma$  hablas. El vocablo es, según Lara (Teoría del diccionario monolingüe, p.119), “una forma léxica abstracta, de naturaleza social y elaborada a lo largo de la historia de la comunidad lingüística”; es la representación de un conjunto de formas léxicas (los tipos de vocablos) que ocurren en el habla como palabras. Así, *aprovechamiento* y *aprovechamientos* son dos tipos que corresponden al vocablo **aprovechamiento**. En la elaboración del vocablo inciden la simplicidad de la forma propuesta, su brevedad en términos del número de letras que la constituyen, su frecuencia absoluta y su capacidad para construir formas en su paradigma o de derivados.

Con vista en el objetivo del presente trabajo los estudios a continuación se restringen al primer grupo mencionado debido a que el mayor uso de un tipo de vocablo en los textos de un ámbito determinado es una señal de que este tipo es un elemento significativo en la construcción de ese discurso.

Cuando un tipo está documentado una sola vez en el corpus, las conclusiones que se podrían hacer sobre su comportamiento léxico, en general, tendrán un muy bajo grado de confiabilidad, puesto que se puede tratar de un uso casual del vocablo,<sup>2</sup> una preferencia individual del autor del texto o una característica específica del texto que se usó como muestra que no refleja el uso en la sociedad. Pero también tenemos que pensar en la posibilidad de errores de captura o verdaderos errores ortográficos, llamados *hapax*, que se introducen de manera aleatoria en el corpus.

La reducción del umbral de validez a 2,<sup>3</sup> me permite no perder posibles candidatos a términos y captar un mayor número de términos diferentes que forman la riqueza del léxico jurídico-ambiental. Considerar los vocablos con una frecuencia igual o mayor que 2 me garanti-

---

2 Aún hay que añadir que la fijación de un umbral en 2 y no en un número mayor de ocurrencias obedece a factores pragmáticos; estudiamos una terminología que está surgiendo en las últimas décadas y que presenta una variación considerable. Por ello es de suponer que habrá un número significativo de vocablos que a pesar de su baja frecuencia sean términos del derecho ambiental.

3 Contrario a la opinión de muchos especialistas de considerar una frecuencia mínima igual o mayor que 4 para que un vocablo tenga valor en el estudio de la estructura léxica de una lengua. Es muy revelador para esta investigación el comentario crítico que hace Lara (1990: 91-92) respecto a estos valores de la frecuencia absoluta: "(...) si bien la documentación de cuatro o cinco ocurrencias de un vocablo parece suficiente para un estudio lexicológico, cuando se trata de los materiales que requiere la lexicografía, tal cantidad de ocurrencias de vocablo resulta, muchas veces, insuficiente si desea uno atenerse estrictamente a esos datos; pues salvo los vocablos que encuentran una dispersión homogénea en todo el corpus y que, por esa razón, suelen corresponder a un uso generalizado y no marcado social o terminológicamente, todos los demás no alcanzan a definir claramente sus usos y se convierten en accidentes aleatorios más o menos cercanos al verdadero núcleo léxico que uno pretende reconocer, y en síntomas solamente de la extensión real de un vocablo en el uso mexicano".

za poder identificar los términos representativos del derecho ambiental mexicano.

Aparte del importante valor operativo de la frecuencia absoluta de un vocablo en un *corpus* determinado, conviene tomar en cuenta también la dispersión de este vocablo entre los distintos géneros que constituyen el *corpus*; si comparamos dos vocablos del cual el primero ocurre dos veces en la misma fuente y el segundo se manifiesta con igual frecuencia, pero en dos fuentes distintas, podemos concluir que la mejor distribución del segundo vocablo tiene su origen en el mayor uso dentro de la comunidad lingüística respectiva y que el primer vocablo se usa por un grupo menor de hablantes en registros específicos de la lengua.

En una segunda fase del análisis cuantitativo de los textos jurídico-ambientales, encaminada a comprobar la primera hipótesis postulada al inicio del trabajo, comparé el corpus jurídico con el *Corpus del español mexicano contemporáneo* (CEMC) que es un corpus de lengua general.

Un vocablo determinado tiene más probabilidades de ser un término técnico cuantas más veces aparezca en un texto reconocido como técnico y, en comparación con los usos comunes de los demás vocablos de la lengua, tenga una mala dispersión en la población léxica, sesgada precisamente por el carácter técnico del texto en que se encuentre.

## ESTUDIO COMPARATIVO ENTRE CADAM Y CEMC

El CEMC está constituido por 996 textos, de aproximadamente 2000 palabras gráficas cada uno, seleccionados de obras escritas y de grabaciones magnetofónicas originarias de toda la República Mexicana. El equipo de investigadores del proyecto del *Diccionario del Español de México* (DEM) de El Colegio de México, que realizó esta recopilación agrupó estos textos en catorce géneros que corresponden a la lengua culta, la lengua estándar coloquial, denominada también lengua sub-

culta, y la lengua no-estándar que incluye las variedades regionales del español de México.<sup>4</sup>

Debido a que este trabajo se centra en el comportamiento de vocablos con usos específicos en un área científica que tienen una función predominantemente referencial, supongo que éstos se usan, ante todo, en el registro sociolingüístico de la lengua culta tal y como la definen los autores del CEMC. Son los géneros de este nivel de lengua los que me interesan más y, en especial, los textos que provienen de fuentes periodísticas, científicas y técnicas. Las entidades que protegen el derecho ambiental son al mismo tiempo también el objeto de estudio de distintas disciplinas científicas y técnicas. Pienso, en particular, en las ciencias naturales como geología, botánica, zoología, las ciencias fisicoquímicas, y los ámbitos científico-técnicos como son las distintas ramas de la ingeniería o las técnicas como las agropecuarias, la caza o el transporte. Los especialistas en estas áreas estudian los seres vivos, las transformaciones de sustancias, las propiedades de la materia y de la energía. El aspecto que el derecho norma y regula son las conductas humanas que pueden influir de una manera relevante en los procesos de interacción entre los sistemas de los organismos vivos y sus sistemas de ambiente e incidir en su protección, conservación, preservación y utilización. El fin que persigue el legislador es garantizar uno de los derechos fundamentales del ser humano consagrado en el artículo cuarto de la Constitución Política de los Estados Unidos Mexicanos: *Toda persona tiene derecho a un medio ambiente adecuado para su desarrollo y bienestar*. Estas reflexiones permiten suponer un número importante de vocablos de origen científico-técnico en los textos del derecho ambiental objeto de estudio.

Pero también el ámbito periodístico se considera como una fuente importante de vocablos especializados, puesto que aquí se usan por primera vez términos que hasta el momento estaban restringidos sólo a un ámbito científico-técnico. Asimismo, en los textos periodísticos se emplean denominaciones originadas en otras lenguas para reflejar problemas actuales de la sociedad que aún no han encontrado reflejo en las ciencias o técnicas del país.

---

<sup>4</sup> Información detallada sobre la composición del CEMC puede ser consultada en Lara (1979) y Lara (1984).



De acuerdo con los objetivos de estudio de la presente investigación, se tomaron en cuenta en el análisis del CEMC los siguientes índices estadísticos.<sup>5</sup> a) la frecuencia absoluta, b) la frecuencia relativa y c) el índice normalizado de dispersión: C.

Mientras que la frecuencia absoluta cuenta el número de ocurrencias de cada vocablo dentro del *corpus*, las frecuencias relativas miden el porcentaje de ocurrencias de cada vocablo respecto al total dentro de cada género y entre los géneros. La frecuencia absoluta refleja la intensidad con la cual se emplea el vocablo analizado. Para los propósitos de la elaboración del DEM, mientras más alta fuera la frecuencia de aparición absoluta en el *corpus*, mayor sería la importancia del vocablo y más merecería ser tomado en consideración. Sin embargo, si se trata de identificar los vocablos con significados especializados y con un uso limitado a determinados ámbitos del conocimiento científico-técnico, sucede lo contrario: interesan mucho más los vocablos ubicados en los niveles bajos de la escala de los valores de frecuencia absoluta. La relación entre la frecuencia absoluta y la probabilidad de ser término científico-técnico es inversamente proporcional: a una frecuencia absoluta alta corresponde una probabilidad baja.

Varios aspectos que cuestionan la adecuación de la frecuencia absoluta como ordenadora exclusiva de los vocablos son la influencia de los tipos de género en el uso de los vocablos y del tamaño del CEMC en cada género.<sup>6</sup> Así, los vocablos especializados pueden aparecer en cualquiera de los catorce géneros,<sup>7</sup> pero su frecuencia será siempre

---

5 El cálculo de los índices estadísticos se realizó con el Analizador gramatical automático desarrollado para el proyecto del DEM y que es un parser morfosintáctico con un componente estadístico y un productor de concordancias (García Hidalgo en Lara/Ham 1979: 85-155).

6 Lara, L.F. et al. (1979). *Investigaciones lingüísticas en lexicografía*, Jornadas 89. México: El Colegio de México. pp.51-52.

7 **Lengua culta**<sup>7</sup> :

1. Literatura (g<sub>1</sub>): obras de literatura, cuentos y ensayos aparecidos en revistas y suplementos culturales
2. Periodismo (g<sub>2</sub>): reportajes de autores mexicanos, editoriales, reseñas polí-

más elevada en el género especializado de tal vocablo. Si se considerara únicamente la frecuencia absoluta se le estaría atribuyendo a una palabra la misma importancia que a otra que tiene la misma frecuencia absoluta aunque su aparición no fuera exclusiva de un género y se repartiera entre todos. En el segundo caso, el mayor tamaño de un género dentro del corpus también le da mayor oportunidad a la aparición de sus vocablos propios y especializados. Las observaciones anteriores han llevado al equipo lexicográfico del DEM a incluir el índice normalizado de dispersión C propuesto por Roberto Ham Chande que mide la importancia y el orden de los vocablos tomando en cuenta su frecuencia de aparición en cada género y el tamaño relativo de cada género. Este índice queda establecido en un rango que varía entre 0 y 1 donde 0 indica la distribución más desigual y 1 la más uniforme. Con vista en que C me dirá cuándo un vocablo está concentrado en uno o varios géneros, cuándo es propio del vocabulario general al encontrarse bien distribuido entre todos los géneros, o cuándo hay una situación intermedia y en qué grado, consideré indispensable incluirlo junto con la frecuencia absoluta y la frecuencia relativa de un vocablo como parámetro en la comparación cuantitativa entre CEMC y CADAM.

---

ticas, reseñas sociales, reseñas culturales, reseñas deportivas, reseñas policíacas, reseñas taurinas

3. Ciencias (g<sub>3</sub>): bibliotecología, filosofía, historia, culturas indígenas, educación y pedagogía, psicología, antropología, arqueología, derecho, economía, geografía, política, sociología, astronomía, matemáticas, electrónica y electricidad, física, geofísica, computación, biología, química, administración, contabilidad, comercio, medicina y veterinaria, medicina humana, arquitectura, artes coreográficas, artes plásticas, artes gráficas, arte dramático, música, cine y fotografía.
4. Técnicas (g<sub>4</sub>): correos y filatelia, periodismo, publicidad, radio y televisión, transporte, mercadotecnia, ingeniería civil, ingeniería industrial, ingeniería química, ingeniería automotriz, ingeniería aérea, ingeniería de ferrocarriles, ingeniería naval, ingeniería de minas, carpintería, electricidad, mecánica, dibujo técnico, enfermería, corte y confección, albañilería, plomería, herrería, agropecuarias, caza y pesca, ejército, charrería, textos del hogar.

El hecho de que la muestra del CEMC se haya obtenido de documentos del español usado en México entre 1921 a 1974, sin que se reflejen las décadas de los ochenta y noventa, y de que no se cuenta con muestras más recientes de tal extensión y posibilidades de aplicación, limita la efectividad de la comparación.

El análisis comparativo comprende varios pasos realizados, en parte, con la ayuda de distintas herramientas computacionales, como Informix, Keywords in context (KWIC), Word y Excel:

- a) confrontación de los tipos del corpus especializado (CADAM) con el corpus de lengua general (CEMC),
- b) agrupación de los tipos de CADAM según su aparición o no aparición en CEMC,
- c) agrupación de los tipos de CADAM que aparecen en CEMC con un índice de dispersión menor que 0.6 y
- d) agrupación de los tipos recogidos bajo c) según la clasificación genérica de la muestra de textos de CEMC.

La fijación del índice de dispersión en los valores menores de 0.6 se fundamenta en la hipótesis de que en cuanto más restringida es la distribución de un vocablo dentro del CEMC tanto mayor es la probabili-

- 
5. Discursos políticos (g5): discurso político.
  6. Religión (g6): religión.
  7. Habla culta (g7): habla culta de la Ciudad de México.

**Lengua sub-culta** :

8. Literatura popular (g8): novela rosa, telenovela, fotonovela, historieta, novela popular.
9. Habla media (g9): habla media de la Ciudad de México.
10. Lírica popular (g10): habla media, habla regional.

**Lengua no-estandar** :

11. Textos dialectales (g11): textos dialectales.
12. Documentos antropológicos (g12): documentos antropológicos.
13. Textos jergales (g13): textos jergales.
14. Textos del hampa y conversación popular (g14): textos del hampa, conversaciones populares.

dad de que se trate de un candidato a término. Los tipos de vocablos que tengan un mayor índice de dispersión que 0.6; adolecen de una restricción genérica y tienen muy poca probabilidad de convertirse en términos técnicos, mientras que los tipos con un bajo índice de dispersión están limitados a ciertos géneros del *corpus*, por lo que tienen una probabilidad más alta de adquirir usos específicos.

De la primera cuantificación del CADAM resultó una lista de 23,535 tipos de vocablos que muestra las 100 palabras más frecuentes en el *corpus*. Depuré de esta lista todas las palabras gramaticales, números y combinaciones de letras que corresponden a partes de distintos vocablos y que por limitaciones del programa computacional kwic no fueron identificados como tales (kwic interpreta el acento como marca delimitativa de una palabra gráfica) y obtuve las 100 palabras plenas con mayor número de ocurrencias en el *corpus* jurídico.

Realicé este recuento con el fin de obtener los vocablos de mayor importancia en los textos del derecho ambiental; es decir, las palabras clave que describen el contenido de los textos en estudio. Identifiqué seis grandes grupos temáticos relacionados con:

- a) las instituciones involucradas en la elaboración y el cumplimiento de la legislación ambiental mexicana (p.ej. *secretaría(s), estados*)
- b) los distintos tipos de disposiciones jurídicas en la materia y su estructura interna (p.ej. *artículo (abreviatura: art), norma(s), reglamento, ley, diario, fracción, acuerdo, disposiciones*)
- c) el bien jurídico protegido (p.ej. *ambiente, naturales, pesca, recursos, especies, agua(s), ambiental, forestal(es), medio, área(s), mar, zona, equilibrio, silvestre, etcétera*)
- d) las actividades humanas que tienen consecuencias negativas para el ambiente (p.ej. *aprovechamiento, uso, desarrollo, residuos, peligrosos*)
- e) las actividades humanas que tienen consecuencias positivas para el ambiente (p.ej. *protección, manejo, control, conservación, almacenamiento*)
- f) el léxico general del discurso jurídico (p.ej. *objeto, materia, plazo, términos, condiciones, fecha, autorización, especifi-*

*caciones, medidas, cumplimiento, procedimiento, debe(n), establece, realizar, refiere)*

Estos temas sintetizan la estructura conceptual básica del derecho ambiental y me facilitan, como no especialista en la materia, una primera visión panorámica de los contenidos del mismo. Además, estas altas frecuencias de uso pueden estar basadas en el carácter terminológico de los vocablos; por el momento, me sirven sólo como indicios que habría que confirmar o invalidar en los siguientes pasos del análisis cuantitativo y cualitativo.

La próxima fase de la investigación centrada en los 14,138 tipos de vocablos que aparecen, por lo menos dos veces en el CADAM, arrojó que 7,296, es decir, el 51.60%, aparecen también en el *Corpus del español mexicano contemporáneo*. No se reflejan en la lengua ordinaria 6,842, esto es, el 48.39%, de los tipos del CADAM. Estos datos, sistemáticamente recogidos, permiten suponer que la mayoría del léxico empleado en los textos jurídico-ambientales es parte del español mexicano contemporáneo. Al mismo tiempo, hay que subrayar que más de la mitad de los tipos de vocablos que comparten ambos corpus muestran un índice de dispersión más bajo que 0.6 en las fuentes textuales del corpus general, lo que indica que su distribución es muy desigual; el conjunto restante de 3,290 tipos de vocablos está definido por un índice de dispersión mayor que 0.6. El primer grupo está conformado por 4,006 tipos que reducidos a 3,907 tipos al eliminar los tipos con un error de ortografía, corresponden finalmente a 2,987 vocablos. Como se verá a continuación, el uso de estos vocablos es muy restringido considerando el amplio abanico de los géneros de la lengua usual.

## ESTUDIO DE LOS VOCABLOS POR GÉNERO

Con el objetivo de comprobar la hipótesis planteada acerca del uso restringido de vocablos con significado especializado, estudio a continuación la distribución genérica dentro del CEMC de aquellos tipos<sup>8</sup> del

---

<sup>8</sup> A continuación utilizaré la forma corta tipo(s) para referirme a tipo(s) de vocablo(s).

corpus jurídico que muestran un índice de dispersión menor que 0.6 en el corpus de lengua general.

Los resultados finales de este estudio confirman los supuestos anunciados al inicio del capítulo acerca de los géneros de CEMC más importantes para la presente investigación.

Identifiqué una concentración de las ocurrencias de los tipos analizados en los textos de las fuentes periodísticas, científicas y técnicas; 3,578 del total de 3,907 tipos; eso es, 91.58%, ocurren en uno o varios de estos géneros. Al lematizar estos 3,578 tipos obtuve 2,729 lemas.

Con vista en estos resultados decidí enfocar el análisis en los tipos de los tres géneros arriba mencionados, suponiendo que es más probable que de estos géneros provengan los candidatos a términos del derecho ambiental.

La comparación de la distribución de los usos entre estos tres géneros me indica, por un lado, que la mayoría de los tipos analizados aparecen en uno o dos géneros, mientras que sólo alrededor de 18% de los tipos se usan en los tres géneros y, por el otro, que los textos provenientes de las ciencias y técnicas parecen ser una fuente más abundante que los textos periodísticos para vocablos de uso limitado en el español mexicano contemporáneo y que posiblemente sean términos del derecho ambiental mexicano.

Respecto a su pertenencia a una categoría gramatical, el análisis me indica que el mayor número de vocablos es una forma nominal. Este dato confirma mi hipótesis inicial con respecto a la naturaleza nominal de la gran mayoría de unidades léxicas posibles candidatas a términos jurídico-ambientales. Estos 2,729 vocablos constituyen junto con los 168 vocablos limitadas al *corpus* jurídico la base de los estudios de carácter cualitativo realizados con el propósito de identificar los términos propios del derecho ambiental.

#### CONFRONTACIÓN DE LOS TIPOS DE VOCABLOS EXCLUSIVOS DE CADAM CON UN DICCIONARIO DE LENGUA GENERAL —DRAE—

La confrontación con la 22ª edición del DRAE tiene el objetivo de recabar aún más datos objetivos e imparciales sobre si los tipos de vocablos del *corpus* del derecho ambiental que no pudieron ser identifica-

dos en el *Corpus del español mexicano contemporáneo*, son parte de la lengua común o se usan con una acepción técnica; el hecho de que un vocablo aparezca sólo en el corpus jurídico y no en el *Corpus del español mexicano contemporáneo* y además no se relaciona con una entrada en el DRAE, eleva la probabilidad de que este vocablo sea término del derecho ambiental.

Una primera revisión del listado de los tipos exclusivos del corpus jurídico sugirió confrontarlos con un diccionario de lengua general de reciente edición suponiendo que aquellos tipos que no aparecen en este diccionario, tengan mayor probabilidad de ser términos del área estudiada. Elegí el DRAE y hallé del conjunto inicial de 6,842 tipos de vocablos restringidos al *corpus* jurídico-ambiental, 2,618 no figuran en el DRAE y 4,224 tipos corresponden a una entrada de este diccionario. El número inicial de tipos 2,618 se elevó a 2,699, puesto que algunos tipos corresponderían a distintas categorías gramaticales; este número de 2,699 tipos al aplicar los criterios de exclusión se redujo a 198 tipos de vocablos, que a su vez corresponden a 168 lemas; es decir, se eliminaron 2,501 tipos que pertenecen a uno de los conjuntos definidos con base en los criterios de exclusión<sup>9</sup>. Entre los lemas identificados se encuentran *abulonero* (adj.; frq.: 3), *acreditamien-*

---

9 La primera revisión de los datos llevó a establecer criterios que permitan excluir además aquellas expresiones que por razones obvias no pueden ser parte del vocabulario propiamente jurídico-ambiental. Las expresiones finalmente excluidas pertenecen a uno de los grupos abajo mencionados:

- a) Denominaciones comunes y del latín de taxonomías científicas, ante todo, de biología y química; denominaciones de especies biológicas y de elementos, sustancias y procesos químicos, como por ejemplo, *guayacán*, *guerrerensis*, *germanio*, *grafito*, *sulfonación*, etc.
- b) Expresiones que aparecen únicamente en uno de los dos apartados *fuentes bibliográficas* o *grado de concordancia con normas y recomendaciones internacionales* de los textos del corpus del derecho ambiental.
- c) Denominaciones de los numerales, fórmulas matemáticas, símbolos matemáticos y científicos, en general: *r = radio de la centrifuga = spindle to the center of the bracker*, etc.
- d) Denominaciones de nombres propios (personas, lugares, instituciones, organizaciones, empresas, lugares): *Guadalajara*, *Galindo*, *Guzmán*.
- e) Denominaciones de publicaciones periódicas, e.g.: *Gaceta Ecológica*, *Diario Oficial de la Federación*.

to (sust.; frq.: 11), *acuacultur* (sust.; frq.: 2), *acuacultura* (sust.; frq.: 111), *acuitardo* (sust.; frq.: 3), *agroforestería* (sust.; frq.: 8), *antropogénico* (adj.; frq.: 2), *autocero* (sust.; frq.: 3), *bioacumulables* (adj.; frq.: 2), *corraleo* (sust.; frq.: 4), *desforestación* (sust.; frq.: 3), *ecotono* (sust.; frq.: 2), *entrecara* (sust.; frq.: 2), *redoblamiento* (sust.; frq.: 10), *traslocación* (sust.; frq.: 5), *zoosanitaria* (adj.; frq.: 4).

## CONCLUSIONES DEL ANÁLISIS CUANTITATIVO

Serán objeto del análisis cualitativo sólo aquellos vocablos que aparecen tanto en el corpus jurídico-ambiental como en el *Corpus del español mexicano contemporáneo* que muestran en este último corpus un índice de dispersión menor que 0.6 y que fueron identificados en textos de los ámbitos periodísticos, científicos y técnicos. Asimismo forman parte del conjunto de datos identificados con base en el estudio cuantitativo los vocablos que figuran sólo en el corpus especializado y que no corresponden a una entrada de la 22ª edición del *Diccionario de la Real Academia Española* (DRAE).

En relación con el grado de objetividad del inventario de vocablos candidatos a términos conviene señalar la diferencia entre juzgar los resultados a partir de una consideración intuitiva de la “realidad” léxica con lo que la evaluación se torna imposible al quedar sujeta a la experiencia de cada especialista en materia jurídico-ambiental, juzgarlos en comparación con trabajos realizados bajo muy diferentes enfoques (por ejemplo, con diccionarios especializados elaborados con objetivos distintos a los presentes en esta tesis) y evaluarlos tras un análisis

---

f) Denominaciones de las distintas disposiciones jurídicas, como son reglamento, Norma Oficial Mexicana en materia de ecología, ley, etc.

g) Nombres de medidas y equipos e instrumentos de medición: *centímetros*,

Se excluyen estas expresiones, debido a que por razones obvias no pueden ser parte del vocabulario propiamente jurídico-ambiental. La razón fundamental para ello es el hecho de que con estos vocablos se denominan los objetos de la legislación jurídico-ambiental en sí y no nos dicen nada sobre el cómo de la actividad legislativa. Este trabajo debe permitir a detectar los términos sobre que se ha legislado y que se restringen al ámbito medioambiental.



cuantitativo basado en criterios objetivos e imparciales. El hecho de que el corpus es representativo para el ámbito especializado en estudio y no era necesario limitarlo a un muestra de un determinado número de párrafos de las fuentes y la forma en que se realizó el estudio cuantitativo, me permite esperar un gran acercamiento a las realidades del léxico propio a esta área del derecho.

Los resultados obtenidos me permiten constatar que el estudio cuantitativo ha sido de gran utilidad debido a los siguientes datos:

1. Pude limitar con métodos objetivos y confiables el conjunto de vocablos candidatos al derecho ambiental.
2. La comparación de un *corpus* especializado con un *corpus* de lengua general con el fin de obtener información confiable sobre el primero, es un método de análisis aún poco usual en el ámbito de la terminología. Es a partir de la primera década del siglo XXI cuando aparecen los primeros estudios terminológicos basados en una comparación con textos de la lengua común.<sup>10</sup>
3. El estudio confirmó las hipótesis planteadas al inicio del capítulo.

Llegué a la conclusión de que las hipótesis planteadas sirvieron para filtrar de una considerable muestra de textos del derecho ambiental aquellos vocablos que con mayor probabilidad pueden ser términos o parte de una unidad terminológica compleja.

Al mismo tiempo debo anotar que no existe un índice numérico que indique con certeza si se trata de una expresión terminológica o no. Es decir, no es posible decir, por ejemplo, si la formación terminológica en derecho ambiental comienza a partir de un índice de dis-

---

10 En el año 2003 Patrick Drouin esboza en *Terminology*, la revista más renombrada en esta área, por primera vez, una nueva técnica para la extracción de términos de corpus especializados basada en la comparación con un corpus no especializado; su objetivo es “to reduce the amount of noise in the list of candidate terms (CTs) by restricting the lexical items that can appear incide candidate terms”.

persión de 0.3 o 0.4, o que una expresión que aparezca por lo menos 3 o 4 veces en el corpus jurídico, es candidata a término. Junto con esta limitante del análisis cuantitativo hay que aludir a las restricciones temporales del CEMC y a la falta de un corpus de fecha reciente.

El carácter terminológico de una expresión simple o compleja se puede corroborar sólo a través del estudio semántico de los contextos de uso de la expresión; no basta con la cuantificación de los significantes, sino que hay que estudiar también el plano del significado de cualquier signo lingüístico. Es decir, tenemos que abandonar el ámbito de la estadística para adentrarnos en la sustancia del contenido.

## BIBLIOGRAFÍA

Drouin, Patrick, "Term extraction using non-technical corpora as a point of leverage", en: *Terminology* 9:1, 2003, 99-115.

Lara, Luis Fernando *et al.* (1979), *Investigaciones lingüísticas en lexicografía*, Jornadas 89, México, El Colegio de México, pp.51-52.

Lozanova Bogomilova, Elena, *Estudio de los procedimientos de formación de términos en el derecho ambiental mexicano*, Tesis de doctorado, El Colegio de México, (en proceso).

Real Academia Española, *Diccionario de la lengua española*, 22ª edición.

# El rotulaje nutricional como fuente de información para el consumidor: Un estudio exploratorio

FERNANDO BITTENCOURT DOS SANTOS  
*Universidade Estadual Paulista "Júlio de Mesquita Filho"*  
*Departamento de Ciência da Informação, Brasil*

## 1 INTRODUCCIÓN

**E**ste trabajo ha buscado identificar la frecuencia de lectura de la información nutricional contenida en los rótulos de los alimentos por los consumidores de la ciudad de Campinas, con el objetivo de obtener un mayor entendimiento (comprensión) de esta lectura, y también en sentido más estricto se propuso identificar la comprensión de las informaciones contenidas en el rótulo de los alimentos por los consumidores; determinar la influencia de la información nutricional en la compra de un determinado producto por el consumidor, y diagnosticar si este consumidor utiliza otras fuentes de información y cuáles son éstas.

Ante el problema de que no siempre la información presente en los rótulos de los alimentos es de fácil comprensión para el consumidor y la supuesta falta de agudeza y/o veracidad de las informaciones contenidas en los rótulos de los alimentos más la inexistencia histórica de una política eficaz para educar (orientar) y concientizar a la población en cuanto a la educación nutricional, además del hecho de que una de las formas del consumidor para evaluar el producto que adquiere son sólo las informaciones contenidas en el rótulo de los alimentos que él

mismo compra, este trabajo se propone contestar algunas cuestiones relativas a la información nutricional contenida en los rótulos de los alimentos ante el contexto de las actividades de marketing de la industria de alimentos.

Para eso ha sido necesario basar este trabajo en un referencial teórico-conceptual que sirva como baliza para los análisis que serán presentados más adelante. Como principio, hemos caracterizado el rótulo y la información nutricional y hemos descrito en forma minuciosa el problema, la justificación y el objetivo del trabajo y también la metodología que hemos utilizado para fundamentar nuestros análisis.

## 2 EL RÓTULO Y LA INFORMACIÓN NUTRICIONAL

El rótulo, según Macedo (1999), es toda inscripción, leyenda e imagen o, toda materia descriptiva o gráfica que esté escrita, impresa, estampada, destacada o pegada sobre el embalaje del producto.

La información que delimita el contenido de un producto alimenticio o que destaca sus propiedades o sus contenidos físicos/químicos, se denomina información nutricional, Rego (1989, p. 07).

Entre las informaciones que el rótulo debe contener, destacamos:

- nombre del producto
- lista de ingredientes
- contenido líquido
- identificación de origen
- identificación del lote
- fecha de duración mínima
- instrucciones para uso, cuando sean necesarias.
- valor calórico
- carbohidratos
- proteínas
- grasas totales
- grasas saturadas y transaturadas
- fibra alimentaria

- calcio
- hierro y sodio

“No contiene colesterol; enriquecido con hierro; contiene gluten; bebida de bajas calorías”, éstos son algunos ejemplos de las diversas informaciones disponibles en el día a día del consumidor y que, aunque no preponderantemente, pueden influir en la toma de decisión de los individuos sobre su alimentación, Salay (2003). Si estos alegatos son adecuados o suficientes o están en exceso son cuestiones que se pueden contestar, partiendo de la abstracción del mercado (industria de los alimentos) en donde las fuerzas de demanda y oferta interactúan; y se ofrece un cierto nivel de información nutricional sobre el producto.

Por varios motivos, sin embargo, ese mercado, que opera por sus propias fuerzas, no satisface necesariamente ni el volumen ni la adecuación de información nutricional necesaria para orientar al consumidor, Belik (2003). Primeramente, porque el consumidor no siempre está preocupado con la alimentación y la nutrición, lo que genera una demanda débil de este atributo. Con poca demanda, entonces, no siempre las empresas tendrán incentivos para innovar y ofrecer informaciones nutricionales.

Otra razón es que el proveedor tendría costos mucho más altos si dispusiera voluntariamente datos nutricionales poco atractivos de su producto, como el alto contenido de sodio o el bajo contenido de vitaminas. La tendencia es enfatizar las propiedades positivas de los alimentos ofrecidos, como por ejemplo, el alto contenido de fibras o de minerales.

La mayoría de las informaciones científicas nutricionales sin embargo, son de carácter público, es decir que cuando están disponibles en el mercado pueden producir beneficios para toda la sociedad. En este sentido esa información debe ser generada y actualizada frecuentemente, y estar disponible para los ciudadanos.

El rótulo nutricional se constituye en una importante fuente de información para el consumidor, pues se refiere a su salud y viene a influir de manera significativa en el alcance de una alimentación de calidad.

### 3 METODOLOGÍA

Esta investigación ha utilizado una metodología exploratoria, que obtiene los datos por cuestionario, el cual presenta preguntas abiertas y cerradas. Los sujetos, foco de la investigación fueron cerca de 100 consumidores elegidos en la ciudad de Campinas.

Los cuestionarios se aplicaron en el periodo comprendido entre junio a septiembre de 2005 por los propios miembros de esta investigación y algunos voluntarios.

Los sujetos de esta investigación fueron en su mayoría amas de casa, universitarios y personas que suelen frecuentar el supermercado. Como el cuestionario no fue tan extenso y es de fácil comprensión, abordamos a las personas en los supermercados de Campinas y les pedimos para que ellas mismas respondieran el cuestionario.

### 4 RESULTADOS

#### 4.1 Lectura de las informaciones

- 60% frecuentemente leen las informaciones contenidas en el rótulo;
- 30% ocasionalmente leen las informaciones;
- 10% nunca leen las informaciones.

Gran parte de los consumidores lee las informaciones contenidas en el rótulo de los alimentos.

#### 4.2 Comprensión de la información nutricional

- 10% leen, pero no la entienden;
- 75% entienden parcialmente;
- 5% encuentran fácil la lectura;
- 10% no leen el rótulo.

Entre los consumidores que leen las informaciones contenidas en el rótulo de los alimentos una gran parte entiende parcialmente lo que

está escrito; algunos leen pero no entienden nada lo que está escrito, pocos encuentran facilidad en la lectura, y el resto de los consumidores no lee el rótulo nutricional.

#### 4.3 Influencia de la información en la compra de un producto

- 92% Sí
- 8% No

Se nota que, en el momento de comprar un producto, la mayoría de los consumidores es o ha sido influenciado por la información contenida en el rótulo del alimento.

#### 4.4 Utilización de otras fuentes de información

- 84% utilizan otras fuentes de información;
- 16% no utilizan otras fuentes de información.

La mayoría de los consumidores utiliza otras fuentes de información. Entre las fuentes utilizadas por los consumidores, destacamos:

- revistas,
- sitios de salud & nutrición,
- bases de datos,
- libros y
- nutriólogos.

### 5 CONSIDERACIONES FINALES

La necesidad de rotulaje nutricional como fuente de información sirve para que las personas puedan discernir qué alimentos son más saludables y puedan así elegir sus productos de acuerdo con su voluntad.

La salud es indispensable para el ser humano, así como la información es el instrumento central para que el individuo adquiera calidad de vida y busque una alimentación saludable.

La Ciencia de la Información es una ciencia interdisciplinar y el campo de la salud y nutrición puede ser un nuevo espacio de actuación para el profesional de la información. Este sector necesita organizar, clasificar e indexar los datos/informaciones nutricionales para los propios profesionales de nutrición, y para ser utilizados y orientar al consumidor en su búsqueda por una alimentación de calidad.

Según Chambolle (1995, p. 12), la información de los consumidores sobre los alimentos y la alimentación no les interesa sólo a ellos. Para los productores de bienes y servicios (agricultores, industrias, distribuidores), ésta es también una herramienta de orientación para un consumo más rentable de los productos y una contribución a la imagen de las empresas. Los profesionales de la información utilizan ese saber para desarrollar relaciones económicas y culturales con el público. Otros grupos de interés (médicos, nutrólogos, investigadores) se interesan por el desarrollo de una información de calidad, porque asegura la legitimidad y el prestigio de sus actividades.

El trabajo realizado ha sido muy satisfactorio y representa una utilidad pública si se tiene en cuenta que la salud, nutrición e información son temas correlacionados y que forman parte del cotidiano de las personas y de los profesionales que diseminan esta información para la población.

## BIBLIOGRAFÍA:

Belik, Walter (2003), *Segurança Alimentar: a contribuição das universidades*, São Paulo, Instituto Ethos, 2003, 88p.

Chambolle, M (1995), *L`information sur les aspects nutritionnels et technologiques des aliments*, París, Económica, 1995, 77p.

Maestro, V. ; Salay, E (2006), "Vantagens e dificuldades do oferecimento de informação nutricional em restaurantes do município de Campinas-SP", en XIX



CONGRESSO BRASILEIRO DE NUTRIÇÃO, 2006, São Paulo, CONBRAN 2006, São Paulo, Associação Brasileira de Nutrição ASBRAN e Associação Paulista de Nutrição APAN, p. 71-71.

Rego, Raul Amaral (1989), *A informação nutricional em rotulagem de produtos alimentícios: um estudo de caso*, São Paulo, 1989, 197p. (dissertação de mestrado).

Salay, Elisabete, *Informação nutricional e mercado: um desafio para o governo*, Campinas, 2003. 2p.



# Los problemas terminológicos de la organización y del acceso a la información

NAIR YUMIKO KOBASHI

*Escola de Comunicações e Artes - Universidade de São Paulo, Brasil*

*Departamento de Biblioteconomia e Documentação*

## 1 INTRODUCCIÓN

La información es un concepto analizable desde distintas perspectivas. En este texto nos acercamos a tal concepto en el contexto de los archivos; es decir, de “la memoria institucionalizada y funcionalizada como información” (Abril, 2004, p.9). Así, remitimos el término al dominio de la Ciencia de la Información, el campo que estudia la información y los dispositivos de almacenamiento y recuperación de información para garantizar flujos.

García Gutiérrez (2002, p. 9) nombra esos dispositivos como “exomemorias”; esto es, memorias exteriores en las que se encuentran inscritos “los pensamientos, experiencias, ilusiones, emociones”.

La interacción con las exomemorias, como también su producción, es afectada por la intervención mediática y tecnológica (Orozco Gómez, 2006). Las tecnologías de la información determinan, en gran medida, las formas de almacenar y hacer disponible la información.

En principio facilitadoras de los procesos informacionales, las tecnologías ubican nuevos problemas: se necesitan mediaciones cada vez más complejas para reconocer, encontrar y comprender la información. Las mediaciones permiten que las inscripciones del pasado sean

reinterpretadas para la construcción/reconstrucción “de la cultura, la conciencia y las identidades del presente” (García Gutiérrez, 2002, p. 13). La producción de informaciones, su almacenamiento en las exomemorias y el acceso no son, por tanto, procesos triviales.

## 2. PRODUCCIÓN DE EXOMEMORIAS

La producción y uso de las exomemorias pueden analizarse en el cuadro de las reflexiones acerca de la comunicación y del lenguaje. Para aclarar el acercamiento se definen preliminarmente los términos directamente involucrados en el campo de la producción y el uso de las exomemorias:

- a) *Información* es representación de contenidos inscritos en artefactos de información. En este sentido, información es una forma simbólica que representa los documentos recoletados y preservados en instituciones de memoria (bibliotecas, archivos y museos, físicos o electrónicos).
- b) Documento e información no son términos equivalentes. El documento es un soporte de mensaje; esto es, un objeto que tiene una dimensión material. La *información*, a su vez, es un objeto simbólico al que se accede por el lenguaje.
- c) *Información documental* es una información formateada para las exomemorias (Kobashi, 1994, 2006). Aunque la información documental es una forma simbólica, se trata de un tipo particular de representación que, en sus distintas formas de expresión (conjunto de descriptores, resúmenes, símbolos de un sistema de clasificación) busca responder a las funcionalidades de los sistemas de recuperación de información.

En la información, obtenida por un lado, por la neutralización del poder expresivo del texto y, por el otro, moldeada de acuerdo con reglas previamente determinadas, prevalece la idea de algo que, aunque formalmente distinta del original, y por lo tanto “representación”, es

equivalente al original desde el punto de vista del contenido. En este sentido la información documental es una representación construida a partir de un objeto efectivamente presente, el documento. Es, así, información segunda, de naturaleza referencial, producida por operaciones metonímicas (Kobashi, 2006).

Svenonius afirma que la organización de la información en sistemas presenta tres dimensiones: la social, la teórico-metodológica y la operacional. La dimensión social es del orden de la institucionalización de las actividades informacionales; la dimensión teórico-metodológica, del orden del conocimiento crítico acumulado que ofrece soluciones para los problemas de la producción y el flujo de la información, y la dimensión operacional es de naturaleza técnica y tecnológica, y se asocia a las técnicas e instrumentos utilizados en estos procesos (SVENONIUS, 2000).

Hay por tanto tres aspectos indisolubles en la producción de las exomemorias: la *esfera teórica*, en la cual se determina el objeto, las funciones, los instrumentos y los procedimientos metodológicos; la *esfera de la producción*, donde se engendran las reglas de generación de las diversas modalidades de información documental y la *esfera pragmática*, en la cual se verifican las condiciones de adherencia entre los usuarios y los sistemas de información (Kobashi, 1994, 2006).

## 2.1 Organización y recuperación de información en exomemorias

Organizar información en exomemorias es, en una primera aproximación, una operación de clasificación que reúne por semejanza y diferencia. Este criterio, en su aparente simplicidad es fuente de muchos problemas, uno de los cuales es la necesidad de contar con un sistema de referencia para representar las informaciones. Cada clase del sistema de referencia debe ser auto-excluyente, y además debe ser lo suficientemente discriminante como para abrigar un conjunto manipulable de objetos que pueda responder a las necesidades de información de distintos usuarios. En otras palabras, clases muy genéricas no son aptas para discriminar información; por otro lado, clases muy específicas pulverizan en lugar de reunir. Los extremos no son útiles para organizar la información.

La información puede ser recuperada de distintas maneras. Inspeccionar cada documento de una colección es una forma elemental de hacer búsquedas. Recurrir a los dispositivos que reúnen y sistematizan información en clases, como las bases de datos bibliográficos, es, en principio, mucho más eficaz (Wellisch, 1987).

Clasificar es también estandarizar. Por medio de la estandarización se obtienen conjuntos de documentos con cierta homogeneidad. Esta operación le otorga previsibilidad a las búsquedas, lo que concurre simultáneamente a la estabilidad y a la economía de los sistemas.

La estandarización se aplica a las cuatro funciones prioritarias de los sistemas de información: a) en la descripción de la forma del documento; b) en la descripción de contenido; c) en la transcripción de los datos para un formato estandarizado —los metadatos— y, finalmente, d) en la búsqueda (Wellisch, 1987).

Un sistema de información no opera, por tanto, sin la mediación de un lenguaje de codificación y de decodificación. Esto implica afirmar que el lenguaje de un sistema de información, para cumplir sus funciones, debe ser moldeado de acuerdo con la comunidad discursiva que va a utilizarlo. Un sistema simbólico compartido es, por tanto, capital en las exomemorias.

El sistema de referencia de los sistemas de información es lenguaje, y, como parte constitutiva de la realidad, sus conceptos presentan dinamicidad; o sea, están en conexión dinámica con la realidad y los sentidos.

## 2.2 Lenguajes documentales

La construcción de lenguajes de codificación y búsqueda de información es una de las tareas más complejas de los sistemas de información. Las categorías aristotélicas definen el principio de estructuración de gran parte de los lenguajes documentales. Las categorías promueven la estabilidad de un universo de conocimiento y permiten segmentar y jerarquizar los conceptos. Los lenguajes documentales (LD), son, además, un conjunto de términos relacionados semánticamente. En los tesauros documentales y ontologías, por ejemplo, la jerarquización es apenas uno de los modos posibles de relacionar

conceptos. Las relaciones son, por tanto, multidimensionales: los términos se asocian por sinonimia o por otros tipos de relaciones espacio-temporales (causa y efecto, proceso y producto, agente y paciente, por ejemplo).

Para que los LD sean funcionales, deben integrar los términos de los lenguajes de especialidad. Así, los LD son afectados, necesariamente, por las asimetrías que se desprenden de la dinámica de la producción de conceptos.

Los lenguajes documentales no están constituidos únicamente de unidades terminológicas; su eficacia se vincula a la asociación de diversos paradigmas de designación: el lenguaje de los dominios técnico-científicos, el lenguaje común o natural y el lenguaje de los documentos.

En ese conjunto, la terminología es el referente fundamental, el sistema a partir del cual se establecen las equivalencias entre los diferentes lenguajes en juego. Aunque el referente del descriptor sea el término, el descriptor no presenta el rigor del término (Cintra, 2003). Su valor es instituido igual por operaciones orientadas por la economía de los sistemas de información.

En efecto, es común en los lenguajes documentales, establecer equivalencias entre términos de mayor intensión y términos de mayor extensión. Éste es un recurso para limitar la cantidad de unidades del léxico documental y dotarlo de mayor eficacia.

Por otro lado, la producción de nuevos términos crea tensiones permanentes en los sistemas de información. Los LD deben por tanto, actualizar permanentemente su base léxica para que no se cristalice y pierda su potencialidad de codificar y transmitir información.

El lenguaje documental, en esa medida, solicita apertura permanente para incorporar nuevas unidades y nuevas relaciones, sin perder, con todo, el cierre de un sistema. No es difícil comprender semejante dinámica al establecerse una analogía entre lenguaje natural y lenguaje documental. El lenguaje natural incorpora nuevas nociones de forma permanente sin perder su condición de sistema. Este atributo es también constitutivo de los lenguajes documentales.

### 3. ESTUDIO EMPÍRICO DE LA CODIFICACIÓN DE INFORMACIÓN EN BASES DE DATOS

La elaboración de repositorios de alta calidad es condición *sine qua non* para preservar y hacer recuperable la memoria colectiva. El estudio de dos bases de datos de tesis brasileñas, presentado a continuación, demuestra algunos de los problemas de indización encontrados.

Los resultados más amplios y los aspectos teóricos, metodológicos y operacionales de estos estudios se detallan en otros trabajos de los investigadores del Grupo Scientia<sup>1</sup> (Eliel, 2006; Igami, Nozaki, Kobashi, 2007; Kobashi, Santos, 2005, 2006a, 2006b, 2006c; Santos, Kobashi, Igami, Bressiani, 2006; Igami, Nozaki, Kobashi, 2007).

#### 3.1 Análisis temático de las tesis de la Ciencia da la Información

Entre 1978 y 2001, han sido defendidas **833 tesis** en el área de Ciencia de la Información. En la indización de las tesis le han sido atribuidos **1929 descriptores** al conjunto. Para obtener un mapa temático del área se han seleccionado 33 descriptores cuya frecuencia de uso está por arriba de 10. Los 33 descriptores (2% del total) han sido atribuidos a 78% del total de documentos del *corpus*. La Fig. 1, abajo, presenta la distribución de los descriptores.

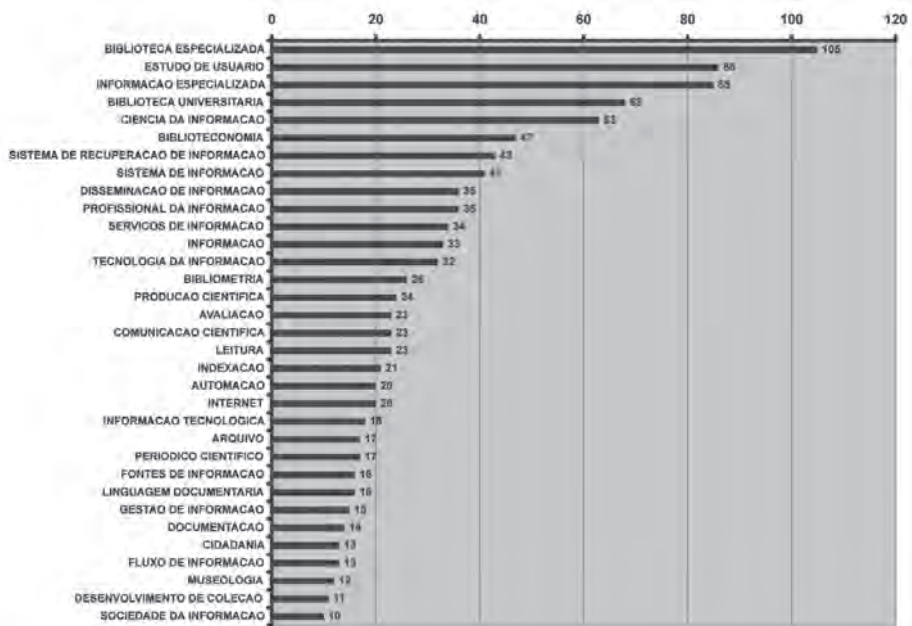
Se han detectado, en el análisis, los problemas siguientes: a) descriptores muy genéricos o, al contrario, muy específicos; b) ausencia de estandarización de conceptos equivalentes y c) uso excesivo o insuficiente de descriptores en cada documento.

---

1 El Grupo Scientia es constituido de investigadores de la Universidade de São Paulo, Universidade Federal de Santa Catarina, Instituto de Pesquisas Energéticas y Nucleares de São Paulo, estudiantes de grado y de posgrado.



Figura 1: Temas de las tesis de Ciencia de la Información (1971-2001)



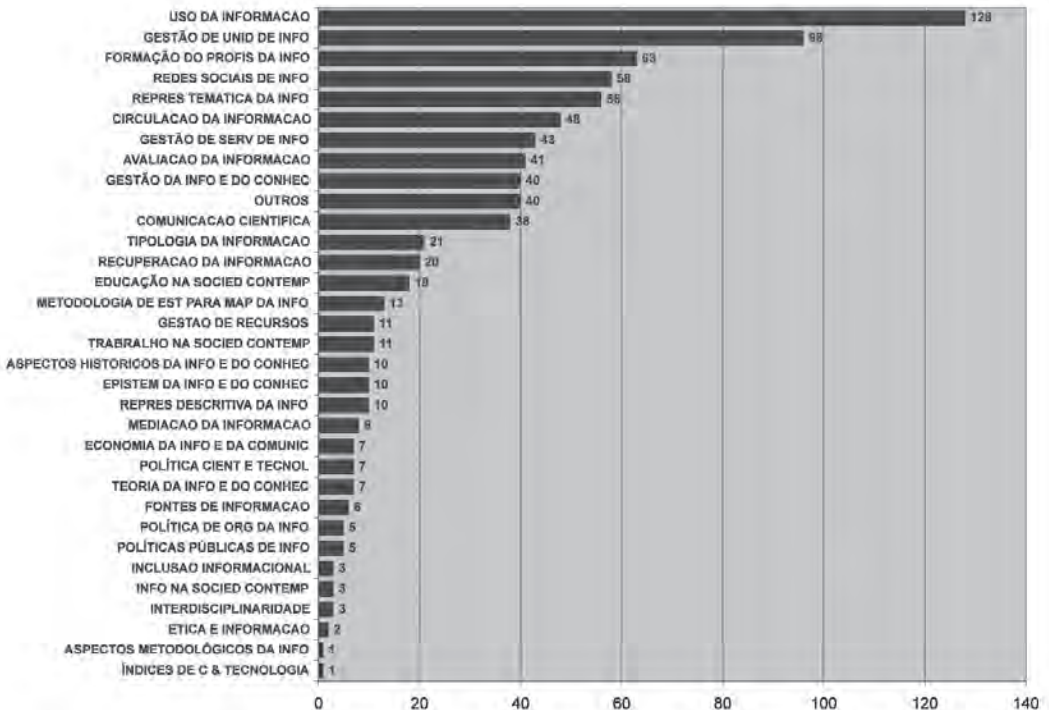
En la indización se han utilizado términos como “Producción científica” y “Comunicación científica” que, en ciertos contextos, representan un mismo proceso. “Información especializada” e “Información tecnológica” pueden ser también términos equivalentes. La ausencia de estandarización sintáctica de los descriptores (singular y plural, por ejemplo) aliada a la falta de uso de un vocabulario controlado de referencia para la indización, como se observa en la figura de arriba, indica que hay disparidades que comprometen la confiabilidad de la recuperación de información.

Para obtener una visión más exacta de los temas y simultáneamente promover diferentes efectos (*precision y recall*), se creó una base de datos *ad hoc* estructurada en: *a) campo de descriptores*: atribuidos por el autor de la tesis; *b) campo de macro-descriptores*: atribuidos *ad hoc*.

*Primer Simposio Internacional sobre Organización...*

Para obtener una visión más global de los datos, se adoptaron las denominaciones de los Grupos de Trabajo de la Associação Nacional de Pesquisa em Ciência da Informação e Biblioteconomia (Ancib) igual como criterio de agrupamiento temático de las tesis (Eliel, 2006). Los nombres de los Grupos de Trabajo, que agrupan las ponencias de los encuentros nacionales, se presentan como categorías temáticas reconocidas y legitimadas por investigadores de Ciência de la Información de nuestro país. Operan, por tanto, como marcos de referencia adecuados para clasificar las tesis en grupos homogéneos.

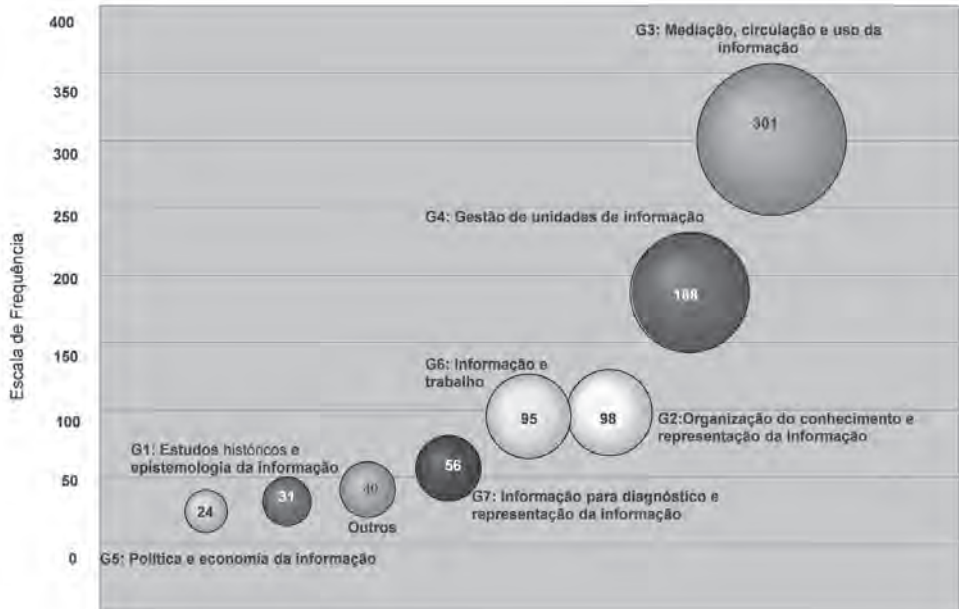
Figura 2 – Distribución de las tesis según términos descriptivos de los Grupos de Trabajo de ANCIB.



La Fig. 2 muestra la distribución de las tesis en un primer nivel de la descripción de los GT. El sistema de referencia utilizado permite ob-

tener una visión más exacta del área, y evitar la dispersión provocada por la atribución de descriptores por parte de los autores de las tesis.

Figura 3 - Tesis de Ciencia de la Información (1978-2001), según los Grupos de Trabajo de la ANCIB.



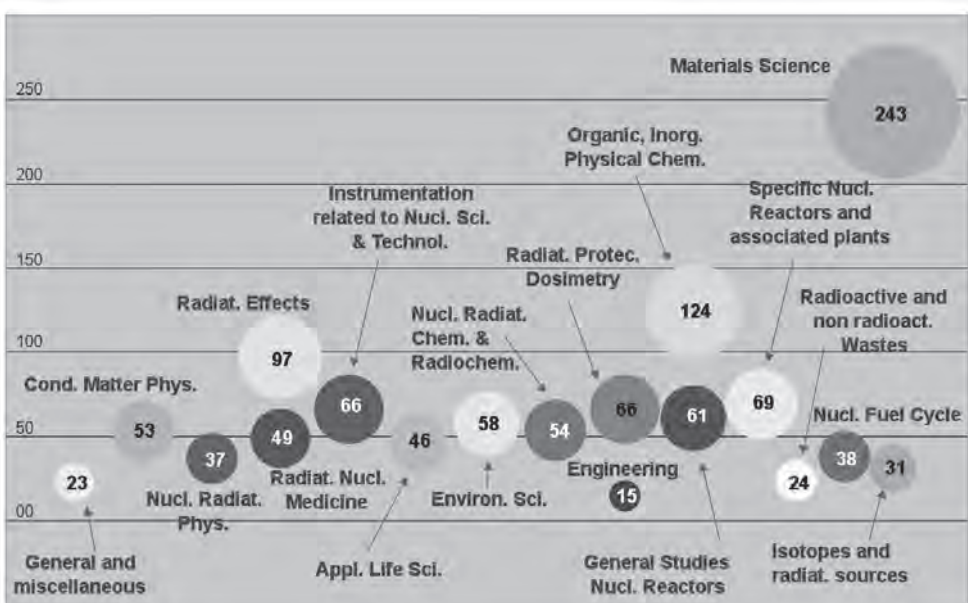
La Fig. 3 presenta un grado aún mayor de generalización de los temas del *corpus* considerado. En orden decreciente vemos: 1) Mediación, Circulación y Uso de la información; 2) Gestión de Unidades de Información, Organización del Conocimiento e Información y Trabajo. Las categorías con menor cantidad de ítems son Información para Diagnóstico, Estudios Históricos y Epistemológicos y Política y Ética de la Información. En la categoría Otros, están las tesis que no se identifican con ninguna de las 7 categorías.

### 3.2 Análisis de las tesis de Energía Nuclear

El *corpus* del estudio del área de Energía Nuclear está constituido por las tesis producidas en los últimos 30 años (1978-2005) en el Programa de Posgrado del Instituto de Pesquisas Energéticas e Nucleares (Santos, Kobashi, Igami, Bressiani, 2006). Los descriptores originales se han integrado a categorías más amplias para obtener conjuntos homogéneos y significativos. Se redondea, así, la dispersión provocada por descriptores muy específicos.

La Fig. 4 presenta la distribución temática de las tesis. La terminología adoptada representa el lenguaje del *Subject categories and scope descriptions* de INIS (2002, 2004).

Figura 4 – Distribución temática de tesis del Programa de Pos-rado de IPEN (1978-2005).



Es visible, en la Figura 4, la prevalencia de tres áreas: *a) Materials Science*, *b) Organic, Inorganic and Physical Chemistry* y *c) Radiation Effects*. Los resultados han sido considerados útiles para el planeamiento de la política de investigación del Programa Institucional de posgrado.

#### 4. CONSIDERACIONES FINALES

Clasificar es esencialmente un proceso descriptivo, que consiste primeramente en elegir las dimensiones de los datos empíricos que serán luego divididos en elementos discretos y enumerables (Granger, 1974, 1989). Para los fines específicos de esta investigación, fue necesario neutralizar la dispersión y constituir nuevos conjuntos de informaciones.

Las clases son, en definitiva, aproximaciones que expresan “semejanzas de familia” (Wittgenstein, 1996), una operación lógico-semántica que permite reunir objetos que presentan trazos comunes. Son, con todo, agrupamientos con alguna inexactitud y fluidez, a semejanza de los descriptores de los lenguajes documentales. La abstracción conceptual es, en ese sentido, la operación que permite caracterizar los datos del *corpus*, y luego compararlos y resignificarlos.

El análisis ha demostrado que las bases de datos de tesis de Ciencia de la Información no presentan la misma calidad y consistencia que la base de tesis de IPEN. Algunos factores explican las diferencias encontradas. En el caso de la Ciencia de la Información, los metadatos han sido producidos descentralizadamente. Cada curso de posgrado envía sus datos a un repositorio central, pero no se hace un trabajo preliminar de darle consistencia a los datos antes de ser incluidos en la base de datos.

Para reorganizar los datos de la Ciencia de la Información, medida necesaria para obtener resultados significativos, fue necesario hacer un trabajo intenso de estandarización. Por esta razón los datos producidos sin recurrir a un sistema de referencia compartido provocaron dispersión de información.

En el caso de IPEN, la consistencia de la base de datos es resultado de la política de documentación de la biblioteca, unidad responsable de la catalogación e indización de la producción científica local.

Quedó evidenciada en este estudio la necesidad de establecer políticas de estandarización y de consistencia de los repositorios producidos cooperativamente. Es necesario destacar que la ausencia de directrices puede comprometer la calidad y la confiabilidad de los sistemas cooperativos. El proceso conocido por auto-archivamiento, si no es controlado por directrices, puede, a largo plazo, comprometer la interoperabilidad de los sistemas.

Los dispositivos de información son sistemas dinámicos continuamente modificados por la introducción de nuevas informaciones. En definitiva, la utilidad de los sistemas de información se establece en la tensión entre la manutención de la estructura y la modificación de sus constituyentes. Los constituyentes son representaciones documentales de discursos de la ciencia, cuya característica fundamental es la creación conceptual, expresada en términos. Por tanto, un sistema de información que no incorpora la dinámica de la creación terminológica, está condenada al fracaso. Se evidencia, de ese modo la importancia de incorporar los métodos de la Terminología en la construcción de los lenguajes de representación de la información y en las operaciones de exploración de las memorias colectivas.

En el caso específico de la Ciencia de la Información, es evidente la necesidad de directrices y de un instrumento compartido de control terminológico para estandarizar la base de datos nacional de tesis. Un Tesoro de Ciencia de la Información actualizado es imprescindible para establecer los parámetros de indización.

Un tesoro documental, en la conceptualización que adoptamos, tiene como referente fundamental los términos producidos en la actividad discursiva de los dominios científicos. Esto es señal de la necesidad del diálogo permanente que debe existir entre la Ciencia de la Información y la Terminología para producir instrumentos de control terminológico adherentes a la dinámica de producción de conceptos.

La Ciencia de la Información reflexiona sobre la producción de las exomemorias, y presenta estructuras lógicas y significantes adecuadas para integrar flujos. El reconocimiento de que la información participa de diferentes estructuras significantes ha sido factor de motivación para renovar los métodos de elaboración de lenguajes documentales, ontologías y taxonomías. Sin esos instrumentos, se estrangula el flujo.

La Ciencia de la Información debe, por tanto, reflexionar sobre las prácticas informacionales y proponer acciones relativas a la producción de la información, su distribución y formas de acceso, en contextos mediáticos y tecnológicos. Por eso es urgente abandonar las visiones simplistas acerca del lenguaje, la representación de información y la producción de sentido.

**Agradecimientos:** expreso mi gratitud a Raimundo N. M. Santos, Priscila Nozaki y Mery P. Z. Igami, compañeros del Grupo Scientia de Investigación; al CNPq y a la Pró-Reitoria de Pesquisa-USP por las becas y el apoyo a la investigación. Un agradecimiento especial a Mery P. Z. Igami por su revisión del texto.

## BIBLIOGRAFÍA

- Abril, G. Prólogo, en GARCÍA GUTIÉRREZ, A., *Otra memoria es posible: estrategias descolonizadoras del archivo mundial*, Sevilla, Un. de Sevilla; Buenos Aires, La Crujia, 2004, pp. 9-13.
- Cintra, A.M.M. *et al.*, Do termo ao descritor: estudo exploratório, en *Revista Comunicações e Artes*, São Paulo, 1993.
- Cintra, A.M.M. *et al.*, *Para entender as linguagens documentárias*, São Paulo, Polis, 2005.
- Ducrot, O., Todorov, T., *Dictionnaire encyclopédique des sciences du langage*. París, Seuil, 1972.
- Eliel, R.A., *Institucionalização da Ciência da Informação no Brasil: estudo da convergência entre a produção científica e os marcos regulatórios da área*, Campinas, PUC-CAMPINAS, 2007, 115f, (dissertação).

**Primer Simposio Internacional sobre Organización...**

- García Gutiérrez, A., *La memoria subrogada: mediación, cultura y conciencia en la red digital*, Granada, Universidad de Granada, 2002.
- García Gutiérrez, A., *Otra memoria es posible: estrategias descolonizadoras del archivo mundial*, Sevilla, Un. de Sevilla; Buenos Aires, La Crujia, 2004.
- Granger, G., *Pensamento formal e ciências do homem*, Lisboa, Presença, 1974, (v. 1 e 2).
- Granger, G., *Por un conhecimento filosófico*, São Paulo, Contexto, 1989.
- Hutchins, W.J., *Languages of indexing and classification: a linguistic study of structures and functions*, Herts, Peter Peregrinus, 1975.
- Igami, Mery P. Z. ; Nozaki, Priscila, y Kobashi, Nair Y., A gestão de qualidade em bases de dados da produção científica, en Congresso brasileiro de biblioteconomia, documentação e ciencia da informação, 12, 8-11 julho, 2007, Brasília, DF. *Anais*, (1cd).
- International Nuclear Information System - Inis, *Joint thesaurus*, Vienna, International Atomic Energy Agency, 2004, ( ETDE/INIS Joint reference series, nº 1, rev. 1).
- International Nuclear Information System - Inis, *Subject categories and scope descriptions*, Vienna, International Atomic Energy Agency, 2002, (ETDE/INIS Joint reference series nº 2).
- Kobashi, N. Y., *A elaboração de informações documentárias: em busca de uma metodologia*, São Paulo, ECA-USP, 1994 (Tesi de doutorado).
- Kobashi, N.Y., *Exomemórias*, São Paulo, ECA-USP, 2006, (Tese de livre-docência) .



- Kobashi, N.Y.; Santos, R.N.M., Information science institutionalization in Brazil: a bibliometric study, en Guerreiro-Bote, Vicente P. (Ed.), Current research in information sciences and technologies. Multidisciplinary approaches to global information Systems, International conference on multidisciplinary information sciences and technologies, 1<sup>st</sup>, october 25-28<sup>th</sup>, 2006, Merida Spain, en *Proceedings...* Badajoz, España, Open Institute of Knowledge, 2006c. v.1 pp. 610-614.
- Kobashi, N.Y.; Santos, R.N.M., Institucionalização da pesquisa científica no Brasil: cartografia temática e de redes sociais por meio de técnicas bibliométricas., en *Transformação*, v.18, p.27-26, 2006b.
- Kobashi, Nair Y.; Santos, Raimundo N. M., dos. Arqueologia do trabalho imaterial: uma aplicação bibliométrica à análise de dissertações e teses, en Encontro Nacional de Pesquisa em Ciência da Informação - 7.,2006, Marília, *Anais...* Marília: FFC/UNESP, 2006b, 1 CD ROM, (VII ENANCIB).
- Lara, M.L.G., *A representação documentária: em jogo a significação*, São Paulo: ECA-USP, 1993, (Dissertação de Mestrado).
- Lima, V. M. A., *Da classificação do conhecimento científico aos sistemas de recuperação de informação: enunciação de codificação e enunciação de decodificação da informação documentária*, São Paulo: USP, 2004.,(Tese de Doutorado).
- Orozco Gómez, G., Aprendiendo con video juegos. III Congreso Educarede, São Paulo, maio/2006, disponível em: [www.educarede.gov.br](http://www.educarede.gov.br)
- Santos, R.N.M.; Kobashi, N.Y.; IGAMI, M.P.Z.; Bressiani, J.C. Brazil's insertion into nuclear energy research: bibliometric analysis of dissertations developed at the Instituto de

***Primer Simposio Internacional sobre Organización...***

Pesquisas Nucleares- São Paulo, en Guerreiro-Bote, V.P. (Ed.), Current research in information sciences and technologies. Multidisciplinary approaches to global information Systems, International conference on multidisciplinary information sciences and technologies, 1<sup>st</sup> October 25-28<sup>th</sup>, 2006, Merida Spain, en *Proceedings...* Badajoz, España: Open Institute of Knowledge, 2006. v.1, pp.553-556.

Svenonius, E., *The intellectual foundation of information organization*, Cambridge, Mass, MIT, 2000.

Wellisch, H.H., *Uma teoria dos sistemas de recuperação da informação*, Brasília, IBICT, 1987.

Wittgenstein, L., *Investigações filosóficas*, Petrópolis, Vozes, 1996.

SECCIÓN III  
Traducción para transferir información



# Estrategias terminológicas para la resolución de traducciones de términos médicos

GRACIELA ROSEMBLAT

*Biblioteca Nacional de Medicina,  
Institutos Nacionales de la Salud, Bethesda,  
MD, EE.UU.*

## INTRODUCCIÓN

La lengua inglesa ha venido a ocupar un lugar de privilegio en la globalización lingüística, y se ha convertido en una especie de *lingua franca* para la comunicación y transmisión de información científica. Es el idioma elegido para las conferencias científicas internacionales y las publicaciones técnicas de alto nivel. Académicos e investigadores de países no angloparlantes deben enfrentar una gran presión institucional y profesional para presentar sus escritos en publicaciones internacionales prestigiosas (en su gran mayoría en inglés y con un alto factor de impacto), y de esta manera diseminar sus teorías y dar a conocer sus trabajos, para obtener un mayor reconocimiento. Esta situación, que data de varias décadas y ha sido descrita en numerosos trabajos [1, 2, 3, entre otros], se observa en todos los países no angloparlantes y en casi todos los campos del conocimiento, incluso en temas de medicina o relacionados con la salud y se refleja en la Web. La economía de los países de habla inglesa permite destinar enormes presupuestos para lograr avances técnicos y científicos. De ahí que esos países lideren las investigaciones en esos campos y publiquen sus resultados, a pesar del sostenido aumento participativo de los demás países en la producción científica mundial.

En el campo de la salud en particular, y especialmente en los Estados Unidos, el creciente número de internautas de habla hispana en los últimos años ha dado lugar a la creación de sitios Web y recursos de información bilingües y en español. El objetivo es brindar apoyo a los hispanohablantes facilitando el acceso a la información en línea en su propio idioma. Esta bienvenida apertura trae aparejada otra serie de problemas. En muchos casos aunque los contenidos se presentan en español, la falta de una redacción pulida y coherente, el uso de frases forzadas o terminología sin traducir (y sin ninguna explicación al respecto) plantean serias dudas sobre la procedencia de la información y, en consecuencia, la fiabilidad de esos sitios Web. La terminología exacta en temas de salud es fundamental para las búsquedas y recuperación de la información e indización de documentos dentro del marco de la comunicación científica y especializada, en todos los idiomas. Con el advenimiento de nuevos tratamientos, compuestos farmacológicos y procedimientos terapéuticos, en su mayoría descubiertos o creados en los Estados Unidos y, por tanto, originados lingüística y conceptualmente en inglés, el rezago terminológico en otros idiomas es inevitable. En consecuencia, muchos términos médicos sólo están fácilmente disponibles en inglés, a pesar de los excelentes diccionarios y recursos multilingües que o no brindan suficiente cobertura en campos específicos, o no están actualizados. Esta ponencia presenta varios métodos para recuperar terminología en español utilizada por científicos y profesionales involucrados en el proceso de la comunicación científica y en los criterios para su validación.

## MARCO CONTEXTUAL

Las estrategias que se describirán a continuación surgieron a partir de un proyecto de investigación de la Biblioteca Nacional de Medicina (parte de los Institutos Nacionales de la Salud de los Estados Unidos) para proporcionar información en español acerca de la investigación clínica, y facilitarle a los usuarios de habla hispana el acceso a información en línea sobre los ensayos clínicos asentados en los registros de dicha institución. Para ello se desarrolló un sistema de recuperación de información y visualización en español: las consultas en espa-

ñol eran traducidas al inglés para recuperar los ensayos clínicos pertinentes de esa base de datos, originalmente escritos en ese idioma. Como resultado de la consulta, los documentos o ensayos clínicos recuperados se presentaban en un formato abreviado, en el cual los títulos eran traducidos manualmente al español, al igual que el lenguaje controlado correspondiente a los tratamientos y trastornos médicos objeto de estudio en cada ensayo [4].

En ese proyecto de investigación, además de la traducción de terminología especializada referente a técnicas o terapias inéditas, la interpretación de las siglas o acrónimos representó un gran desafío, dado que los proveedores de datos no tienen la obligación de incluir las expresiones completas que corresponden a dichas abreviaturas. Es común que los acrónimos den lugar a ambigüedades al referirse a diferentes conceptos. En algunos casos se utilizan como una manera ingeniosa de acortar el título de un ensayo clínico, formando una palabra relacionada con el texto, simbolismo prácticamente irreproducible en la traducción a otros idiomas.

Un recurso incomparable que incluye tanto traducciones al español como expansiones de siglas es el Unified Medical Language System® (UMLS® [5]). La fuente terminológica comprendida en el Metathesaurus, parte del UMLS®, constituye el recurso léxico más exhaustivo para el campo médico. En esa base de datos, una sigla y su expansión en la expresión completa son representados como sinónimos o variantes de un mismo concepto, junto con las correspondientes entradas multilingües o términos traducidos. Los términos MeSH [6] y SNOMED CT son las principales fuentes de datos para las entradas en español del UMLS®. MeSH, o su equivalente en español, los Descriptores en el Campo de la Salud (DeCS), es considerado como un lenguaje controlado en el sentido que se utiliza para la indización de textos, lo cual facilita luego su recuperación. Además, es una forma de estandarizar el vocabulario técnico. No obstante, pese a su gran cobertura, muchas entradas en el UMLS® carecen de traducciones al español: en la versión 2007AB, de un total de 1.436.182 conceptos, sólo 342.123 (24%) tienen al menos un sinónimo en español (1.454.317 términos en español frente a 3.547.627 en inglés). Esta falta de cobertura es genera-

lizada: a pesar de los grandes avances informáticos, las bases de datos bilingües con glosarios especializados y las aplicaciones de software para la recuperación automática de textos multidisciplinares multi-lingües no son suficientemente exhaustivas para el ámbito biomédico.

Dentro del marco contextual de este proyecto de investigación, la metodología desarrollada para el tratamiento terminológico está basada en técnicas de búsqueda en Internet eficaces y principios lingüísticos para obtener traducciones correctas, no fácilmente disponibles, al español, de terminología médica especializada en inglés. La estrategia combinada que se describirá se presta a la validación y convalidación terminológica, y lleva a la recuperación de las mejores opciones de traducción o posibles traducciones. La estrategia está dirigida tanto a traductores especializados, lexicógrafos y bibliotecarios como a científicos y representantes de las ciencias de la comunicación que se dedican al desarrollo de contenidos de páginas Web, especialmente en el ámbito de la medicina y temas relacionados con la salud. La metodología puede aplicarse además en otras disciplinas y en otros idiomas ajustando los principios lingüísticos a éstos.

## METODOLOGÍA

Se describirán tres métodos diseñados para su utilización de un modo iterativo y complementario, que pueden combinarse en una sola estrategia por etapas para obtener traducciones al español a partir de las expresiones en inglés. Estos métodos consisten en:

- 1) el aprovechamiento de textos ya alineados o *corpus* paralelos inglés-español, empleados con frecuencia en la traducción automática;
- 2) la utilización de señales semánticas y del contexto conceptual, semántico y sintáctico para cubrir los vacíos terminológicos en español;
- 3) la aplicación de conocimientos lingüísticos específicos para el español, ateniéndose a las características morfosintácticas propias de la lengua.



Antes de poner en práctica estos métodos, un paso preliminar de gran utilidad consiste en realizar búsquedas manuales de la expresión original en inglés en las fuentes primarias; es decir, en hacer búsquedas monolingües en textos en inglés, lo cual permite verificar:

- a) la inexistencia de posibles errores ortográficos, tipográficos o de uso: la ortografía correcta puede corresponder a una expresión para la cual ya existe una traducción al español y no procede continuar con la búsqueda.
- b) si la expresión en cuestión es una marca registrada y no se ha incluido el símbolo correspondiente. Un nombre propio o una marca registrada debe emplearse en la traducción tal cual aparece en el idioma original, con el agregado del símbolo correspondiente, y no procede continuar con la búsqueda.
- c) el contexto léxico-semántico de la expresión que da la pauta de su significado y uso correcto, además de la visualización de otros términos con los cuales se utiliza generalmente la expresión y que también constituyen un recurso en búsquedas puntuales para refinar los patrones lingüísticos en los que se utiliza, derivando contextos más limitados dentro del ámbito especializado;
- d) el concepto exacto denotado por la expresión o sinónimos de la misma, dado que el reconocimiento del significado es clave para una indización correcta [7]. Además de analizar el contexto semántico y el uso de una expresión, se puede también obtener el significado del concepto ingresando “define:xxx” en la casilla de búsqueda (buscador Google®), reemplazando “xxx” por la expresión deseada.
- e) el campo del conocimiento científico, que puede presentar variaciones en el uso de una misma expresión. Para determinar el campo apropiado se pueden limitar las búsquedas a ciertos encabezamientos de páginas Web, o recurrir a sistemas que permitan la detección automática o semiautomática de términos o candidatos a términos<sup>1</sup> [9].

---

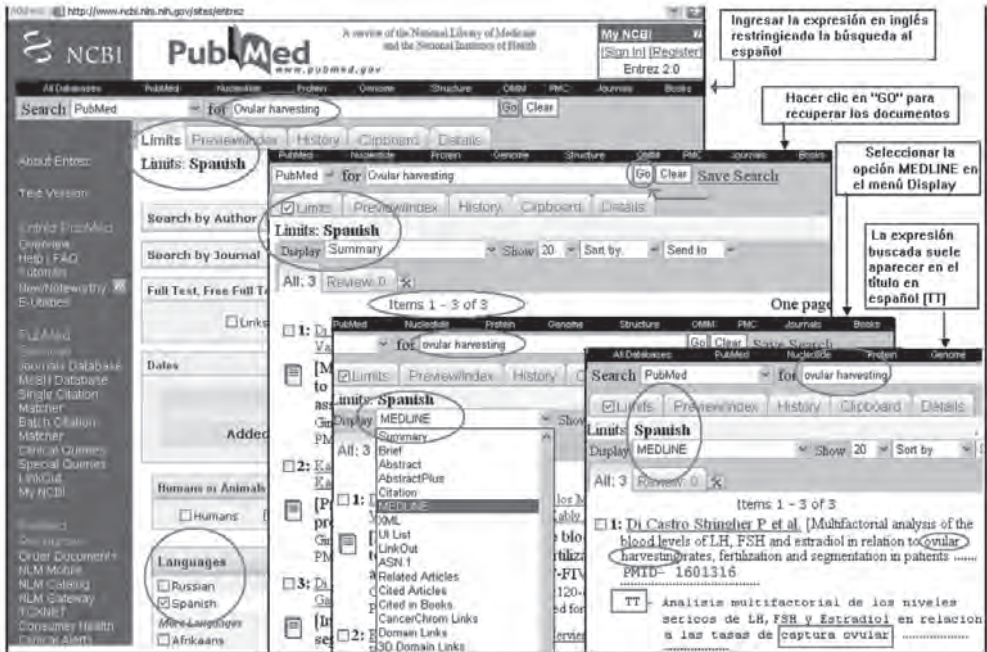
<sup>1</sup> Véase, por ejemplo, la herramienta YATE [8] que combina estrategias probabilísticas y lingüísticas.

Según los resultados de este paso preliminar, se procede o no a las técnicas de búsqueda, que se explicarán a continuación.

1) *Aprovechamiento de textos bilingües paralelos* existentes en línea, escritos en el idioma original (que puede ser tanto inglés como español) por especialistas en la materia, con su traducción correspondiente al otro idioma. Estos textos constituyen un inestimable recurso terminológico, aun cuando solamente una parte del documento presente traducción, que es generalmente la parte que corresponde al resumen o al título. El texto no necesariamente está alineado oración por oración o párrafo por párrafo (texto paralelo), método muy utilizado en la traducción automática, sino que puede presentar una alineación más local y conceptual, o temática [10], también llamada texto o *corpus* comparable [11].

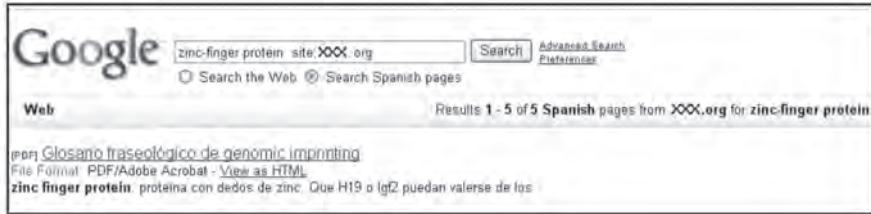
La búsqueda de estos corpus paralelos puede realizarse en sitios Web especializados u otros generales. En el primer caso, si existiera la posibilidad de que la traducción de una expresión en inglés pueda aparecer en publicaciones en español, se puede hacer una búsqueda en PubMed [12] para ver qué términos utilizan los investigadores de habla hispana en las fuentes primarias para referirse a la expresión original. La búsqueda en PubMed comienza ingresando la expresión en inglés en la casilla de búsqueda. En la pestaña o ficha denominada Limits (Límites), se puede restringir la búsqueda a los documentos originalmente escritos en español seleccionando “Spanish” en el menú de idiomas (Languages) (Fig. 1). Si hubiera textos recuperados como resultado de la búsqueda, al seleccionar la opción ‘MEDLINE’ en el menú de visualización (Display), se obtendrá la referencia completa de los artículos, con los títulos originales en español en la etiqueta [T], que a menudo incluyen la traducción de la expresión en inglés consultada, u otros términos relacionados con la misma (sinónimos o contexto léxico-semántico).

Figura 1. Búsqueda en PubMed de una expresión en inglés restringida al español



Además de las búsquedas avanzadas en PubMed restringidas a publicaciones en español, se puede aplicar el mismo principio para restringir una búsqueda avanzada en textos en español (*Search Spanish pages*) a algunos de los sitios Web especializados y de alta fiabilidad que ofrecen resúmenes o glosarios en ambos idiomas (inglés y español). Para ello, utilizando el buscador Google®, se indica el sitio Web deseado en la casilla de búsqueda a continuación de la expresión buscada, como en el siguiente ejemplo:

Figura 2. Búsqueda avanzada de una expresión en inglés en textos en español restringida a un sitio determinado, utilizando el buscador Google®



La expresión en inglés, ya sea completa o en partes, también puede buscarse en fuentes no específicas (más generales) disponibles en español, que proporcionan resúmenes o textos en ambos idiomas. Aun cuando no estén alineados, los resúmenes pueden dar lugar a la recuperación de posibles traducciones en español para su posterior validación en los textos monolingües. Las posibles traducciones pueden identificarse por la semejanza morfológica en ambos idiomas, por la correspondencia semántica, o por una comprensión clara del concepto denotado, junto con los correspondientes microcontextos léxico-semánticos (colocaciones típicas) y sintagmáticos en ambos textos.

## *2) Señales semánticas y contextuales en casos de vacíos terminológicos*

Cuando las búsquedas para obtener traducciones o posibles traducciones no son exitosas y no se recuperan documentos en español, o no se puede determinar fehacientemente la existencia de una traducción estándar o aceptada por la comunidad científica, resulta útil recurrir a un análisis del contexto léxico-semántico de la expresión en inglés. Este nivel de fiabilidad no se determina por un número mínimo aceptable de apariciones en Google® establecido a priori, sino por una conjunción de factores: origen de las fuentes primarias en cuanto a país (hispanohablante) e instituciones confiables, número de apariciones en textos especializados (PubMed, Google®) y el sentido común y conocimientos del traductor o terminólogo. A partir de este análisis del contexto léxico-semántico pueden surgir varias opciones

de traducción al español. Una estrategia consiste en hacer búsquedas en las fuentes primarias en español de los términos que aparecen regularmente en el contexto de la expresión en inglés cuya traducción nos interesa. Aunque esta búsqueda recupera documentos en lugar de expresiones, la terminología utilizada para la traducción de la expresión en inglés puede identificarse si se tiene una comprensión del concepto incorporado en la semántica de la expresión original, de la mano de un minucioso análisis léxico y semántico del microcontexto, el cual difiere en cada caso y es determinado por el terminólogo-traductor, o las colocaciones típicas en que aparece la expresión, que incluye los términos vecinos, los modificadores directos o indirectos y los complementos, además de las relaciones semánticas y conceptuales entre todos estos. El microcontexto puede ir desde el sintagma nominal hasta el párrafo según la granularidad descriptiva del texto y la introducción de conceptos afines.

Por ejemplo, una búsqueda en PubMed de los títulos en español ([TT]) que contengan una traducción de la expresión en inglés ‘*second-look (surgery)*’, definida como “cirugía realizada después del tratamiento primario...”,<sup>2</sup> no obtuvo resultados. En cambio, una búsqueda avanzada en Google® obtuvo los siguientes resultados: (*cirugía secundaria*) (100 citas), (*cirugía de segunda exploración*) (15), (*de segunda inspección*) (6), (*de segunda intención*) (3), (*de segunda observación*) (1). A pesar de que el término correspondiente a ‘look’ no aparece en ninguna de las posibles traducciones para la expresión ‘second-look’ (surgery), los contextos de las expresiones recuperadas muestran una clara correspondencia semántica con el concepto en inglés. En este caso, las opciones de traducción incluyen:

- Utilizar una de las expresiones en español que más textos recuperan e incluir a continuación la expresión original en inglés entre paréntesis: *cirugía secundaria* (second-look surgery);
- Utilizar dos de las expresiones halladas en los textos recuperados en español, la segunda entre paréntesis a modo de aclaración

---

2 Dictionary of Cancer Terms, National Cancer Institute, NIH

de la primera: *cirugía secundaria (cirugía de segunda exploración)*;

- Utilizar sólo la expresión en español que más textos recupera, por ejemplo, *cirugía secundaria*, siempre que la fiabilidad de la fuente primaria pueda comprobarse mediante la institución y el país de procedencia, la calidad del resto del texto o la consonancia con expresiones similares que sean más conocidas, entre otros criterios.

Las posibles traducciones pueden validarse examinando su uso en textos monolingües en español. La recuperación de un número alto de textos de fuentes primarias fidedignas en español brinda buena validación para una traducción, especialmente cuando se utiliza como consulta de sitios Web principalmente originados en países latinoamericanos o en España. Para acotar la búsqueda en Google® a documentos de ciertos países, se debe incluir en la casilla de búsqueda el país de origen deseado, a continuación de la expresión que se necesita consultar, de modo similar a la restricción de la búsqueda a fuentes específicas, como en la Figura 2. En el siguiente ejemplo, la expresión “vía de señalización de Notch” tal como está escrita será buscada en sitios exclusivamente de España (.es):

Figura 3. Búsqueda avanzada de una expresión en inglés en textos en español restringida a sitios de España, utilizando el buscador Google®



Para los vacíos terminológicos de expresiones en inglés que no tienen traducción correspondiente en español, una explicación corta, puntual o una paráfrasis en español de la expresión buscada puede

utilizarse a modo de traducción con el fin de proporcionar el significado, dado que no es admisible incluir una expresión en inglés —que puede ser desconocida— en medio de una narrativa en español, sin antes introducir esta expresión mediante una descripción explicativa. Por ejemplo, actualmente no existe en español una traducción para el término ‘mindfulness’,<sup>3</sup> comúnmente utilizado en psicología en la expresión ‘mindfulness-based therapy’. Muchos textos/documentos, aun los provenientes de países hispanohablantes, utilizan el término en inglés. Sería difícil que hubiera un cierto consenso entre diferentes profesionales acerca de cómo explicar el concepto en español, ya que puede hacerse con mayor o menor detalle. Algunas opciones son:<sup>4</sup>

- ‘Técnica terapéutica utilizada en psicología (llamada “mindfulness” en inglés)’
- ‘Técnica terapéutica que incluye observación activa, meditación y aceptación (conocida como “mindfulness” en inglés)’
- ‘Procedimiento terapéutico que consiste en tener plena conciencia de sí mismo de una forma deliberada y neutral, con observación activa, meditación y aceptación (conocido como “mindfulness” en inglés)’

### *3) Aplicación de conocimientos lingüísticos específicos para el español*

De más está decir que los conocimientos lingüísticos y las características morfosintácticas propias de una lengua son una valiosa herramienta que ayuda a inferir posibles traducciones, especialmente en casos de variantes terminológicas que denotan un mismo concepto. Esta herramienta cumple un rol fundamental ante la falta de una traducción estándar (o estandarizada) que pueda ser validada mediante búsquedas en textos monolingües en fuentes primarias en español. Algunos campos lingüísticos en los que pueden aplicarse estos tipos

---

3 The practice whereby a person is intentionally aware of his/her thoughts and actions in the present moment applied to bodily actions and the mind's own thoughts and feelings. [Definición de ‘mindfulness’: [www.wikipedia.org](http://www.wikipedia.org)]

4 Opciones de traducción, G.R.

de conocimiento son: (3a) los procesos morfológicos que tienen lugar en la derivación terminológica, (3b) los sintagmas nominales en inglés compuestos de dos sustantivos y (3c) las cuestiones de uso de la lengua.

3a) Los sufijos utilizados en la morfología derivativa para la formación de las palabras y las categorías gramaticales a menudo le agregan al núcleo semántico de la raíz la información sintáctica, específica para cada idioma. Suele suceder que las correspondencias en diferentes idiomas de muchos términos médicos [13] y comunes sólo difieran en su terminación [14]. Por ejemplo, el sufijo adjetival inglés ‘-al’ puede corresponder a diversos sufijos adjetivales en español, uno de los cuales es también ‘-al’ (Cuadro 1). Esto último puede dar lugar a generalizaciones erróneas en las traducciones del inglés al español, a pesar de que en ciertos casos, con el tiempo y por imposición del uso, el anglicismo termina siendo aceptado como sinónimo de la expresión que corresponde en español, como es el caso del término inglés “viral” y las traducciones ‘vírico’ o ‘viral’ en español, o “aneurysmal” y ‘aneurismático’ o ‘aneurismal’ en español.

Cuadro 1. Terminaciones adjetivales en inglés y en español

INGLÉS	ESPAÑOL
<b>-al</b>	<b>-al</b>
distal, spinal, vaginal	<i>distal, espinal, vaginal</i>
mitral, transforaminal	<i>mitral, transforaminal</i>
monoclonal, hepatorenal	<i>monoclonal, hepatorenal</i>
	<b>-ico</b>
paroxysmal, surgical	<i>paroxístico, quirúrgico</i>
liposomal, menopausal	<i>liposómico, menopáusico</i>
	<b>-ivo</b>
educational, informational	<i>educativo, informativo</i>
	<b>-oso</b>
parenchymal, variceal	<i>parenquimatoso, varicoso</i>
mesenchymal	<i>mesenquimatoso</i>
	<b>-ario</b>
embryonal, diaphyseal	<i>embrionario, diafisario</i>
	<b>-ano</b>
carpal, tarsal	<i>carpiano, tarsiano</i>
retinal, cranial	<i>retiniano, craneano</i>
	<b>-dor/-tor</b>
explorational, reductional	<i>explorador, reductor</i>



3b) El proceso de formación de sintagmas nominales compuestos de dos sustantivos en inglés es inespecífico con respecto a la relación entre los sustantivos que lo componen, mientras que el español lexicaliza la relación entre ambos. Las traducciones de estos compuestos a menudo se expresan revirtiendo el orden de los sustantivos e insertando una preposición entre ellos. El traductor o lexicógrafo debe seleccionar la preposición que indica y especifica la relación implícita en el sintagma en inglés, y es idiomática en español (Cuadro 2). En algunos casos existe la opción entre dos preposiciones y ambas pueden ser igualmente válidas, como por ejemplo: *trastorno de ansiedad / trastorno por ansiedad* ('anxiety disorder');

Cuadro 2. Sintagmas nominales de dos sustantivos en inglés y en español

INGLÉS	ESPAÑOL
macrolide antibiotic	<i>antibioticoterapia con macrólidos</i>
interferon treatment	<i>tratamiento con interferón</i>
anodal stimulation	<i>estimulación por corriente anódica</i>
mercury poisoning	<i>intoxicación por mercurio</i>
magnesium oxide	<i>óxido de magnesio</i>
transmission condition	<i>condición para la transmisión</i>

Otros sintagmas nominales compuestos de dos sustantivos en inglés se traducen al español mediante un complemento adjetival pos-modificador del sustantivo núcleo:

Cuadro 3. Sintagmas nominales de dos sustantivos en inglés, sustantivo + sintagma adjetival en español

INGLÉS	ESPAÑOL
blood disorders	<i>trastornos hematológicos</i>
telephone intervention	<i>intervención telefónica</i>
telemedicine intervention	<i>intervención basada en telemedicina</i>

3c) Los idiomas difieren en el potencial léxico de las palabras cognadas, y estas diferencias afectan la traducción. Por ejemplo, el inglés

tiene pares de palabras relacionadas que se refieren a los procesos y sus resultados, como es el caso de ‘neoplasia’-‘neoplasm’. El español tiene las mismas palabras cognadas (*neoplasia* - *neoplasma*), pero la primera se usa generalmente tanto para el proceso como para el resultado. Por consiguiente, en español el término ‘*neoplasia*’ se usa como traducción del vocablo inglés ‘neoplasm’. Otros ejemplos son ‘sonogram’ - *ecografía*; ‘angiogram’ - *angiografía*; ‘arteriogram’ - *arteriografía*.

En el caso del español, las variaciones dialectales entre los diferentes países cobran mayor relevancia por cuanto atañen también a conceptos básicos en cuestiones de medicina. El ejemplo más común está dado por los diferentes vocablos para denotar las partes del cuerpo. Estas variaciones dialectales agregan otro factor de dificultad a las conocidas variantes terminológicas existentes en cada idioma (sinónimos, paráfrasis, polisemia, entre otras). El consenso terminológico está dado por el uso y la adopción. Cuando no existe consenso terminológico, dos o más variantes terminológicas pueden utilizarse para un mismo concepto, según el país o área geográfica de procedencia. Por ejemplo, ‘pelvic floor’ se traduce tanto como *suelo pelviano* (España) o como *piso pélvico* (América Latina) sin que ello implique una diferencia en cuanto a su significado.

## ESTRATEGIA COMBINADA

La estrategia general que se propone consta de la aplicación de los tres métodos descriptos, que son complementarios y pueden utilizarse para la validación y convalidación terminológica. La secuencia para derivar la traducción más estándar existente en español se indica en los siguientes pasos, en el orden en el cual se presentan, a partir de la expresión en inglés:

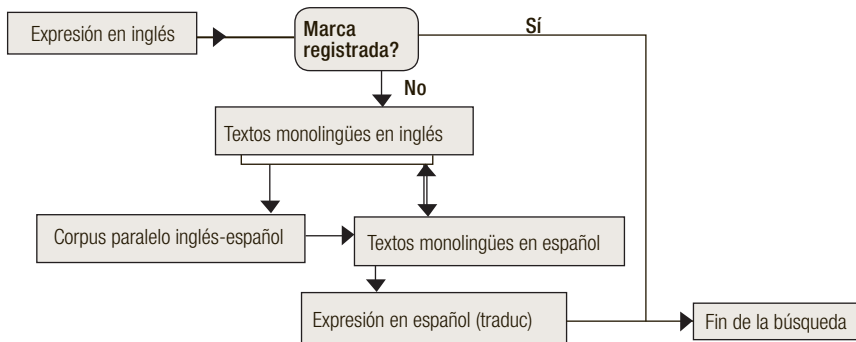
- Paso 1. Búsqueda en fuentes primarias en inglés (monolingües) para confirmar la ortografía, el significado y el contexto de uso de la expresión en inglés. Si la expresión es una marca registrada, la búsqueda concluye aquí. Este es el caso, por ejemplo, de ‘Lupron therapy’ [*Tratamiento con Lupron*®].
- Paso 2. Búsqueda de la expresión en inglés en fuentes en español, para identificar textos paralelos en los resúmenes o los títulos.

los, o corpus paralelos o equivalentes (comparables). Por ejemplo, 'maxillary sinus lift' [*elevación del seno maxilar*]<sup>5</sup>. Las expresiones recuperadas en español siempre deben validarse en textos monolingües en español (Paso 3).

- Paso 3. Por último, búsqueda de las posibles traducciones (expresiones recuperadas) en fuentes primarias en español (monolingües), para confirmar la ortografía, el significado y el contexto del uso en este idioma. Además del número de fuentes que utilizan una u otra posible traducción (frecuencia de uso), las fuentes confiables ya conocidas por el profesional y la procedencia de un país hispanohablante confieren una innegable señal de peso que sirve de guía para la decisión final del lexicógrafo o traductor.

Si el paso 2 no arroja ninguna traducción o posible traducción debido a la falta de corpus paralelos que contengan la expresión en cuestión, o por vacíos terminológicos en español, tendrán lugar varias reiteraciones de los pasos 1 y 3 hasta que los criterios lingüísticos sean satisfechos. Puede resultar útil recurrir también a la traducción automática para obtener algunas pautas de búsqueda, por paralelismo con otras traducciones en las cuales se emplea la expresión o expresiones objeto de búsqueda. El paso 3 siempre será el último, dado que en todos los casos las traducciones deben validarse en textos monolingües en español, como se muestra a continuación:

Figura 4. Secuencia de los pasos. La flecha doble representa una o más reiteraciones.



5 <http://www.scielo.isciii.es/pdf/medicor/v10n4/09.pdf> [02/01/2007]

## EVALUACIÓN Y PRUEBA DE CONCEPTO

Para determinar la exactitud y eficacia de la estrategia descrita, se realizaron dos tipos de evaluaciones de las traducciones en el trabajo de investigación anteriormente mencionado, obtenidas mediante la aplicación de la metodología que depende de diferentes técnicas manuales utilizadas de manera complementaria, con validación y convalidación. Una evaluación preliminar de los resultados de esta metodología fue realizada por un traductor médico experto, inglés-español, sobre la base de 200 títulos de los ensayos clínicos, de un promedio de 25 palabras cada uno, y 300 expresiones de lenguaje controlado utilizadas en diferentes ensayos. Los resultados demostraron inequívocamente la validez de la estrategia, con la calificación más alta (traducción impecable) en el 95.4% de los casos. Otra evaluación de diferente carácter fue llevada a cabo por los participantes en un estudio formal de utilización [15] sobre el proyecto citado, quienes a través de las observaciones positivas sobre los títulos y expresiones traducidos al español brindaron validación externa de la eficacia de la metodología presentada, desde el punto de vista de los usuarios respecto de la información aportada.

## CONCLUSIONES

Si bien el conocimiento del dominio es indudablemente un gran recurso, también se requiere competencia lingüística y del contexto de uso para obtener o derivar la mejor traducción posible. Dejando de lado la controversia acerca de si verdaderamente existe una traducción perfecta, los recursos metodológicos y lingüísticos presentados en este trabajo no constituyen una lista exhaustiva, sino que se han incluido para ilustrar un sistema que puede asimismo aplicarse en otros idiomas u otros campos del conocimiento para obtener los resultados deseados.

Cabe mencionar aquí las conocidas advertencias acerca de la falta de autoría oficial y revisión en Internet, para todos los idiomas. No toda la información en español es confiable y cada usuario debe determinar la fiabilidad de las fuentes. Cuando las búsquedas se realizan

de manera sistemática en un ámbito especializado, se llega a un buen grado de conocimiento de cuáles sitios Web son confiables y valederos, lo cual facilita tanto las búsquedas como la obtención de resultados. A pesar de que los métodos mencionados son manuales, para uso de traductores o terminólogos en el transcurso de su trabajo, sería posible automatizarlos hasta cierto punto, por ejemplo, creando barras de búsqueda para Internet Explorer que permitan realizar búsquedas simultáneas en 3 tipos de páginas Web (monolingües en inglés o en español, PubMed, páginas bilingües), con la opción de restringir las búsquedas a páginas Web procedentes de países hispanohablantes o de ciertas instituciones de reconocida fiabilidad por la experiencia del profesional.

Una característica importante de los recursos citados es la posibilidad de emplearlos como estrategia interdisciplinaria y multisectorial fuera del marco de la traducción, aplicable a la mejora de la indización y la recuperación de documentos, al suministro de un marco apropiado para la evaluación terminológica, a la normalización lingüística y estilística en medios multilingües, a campos como la bibliotecología, la elaboración de tesauros, la prensa hablada y escrita, la revisión editorial, la investigación científica y la docencia, así como a la creación de sitios Web idóneos y fidedignos, por citar sólo algunos. Estas aplicaciones prácticas permitirán atender la demanda de información multilingüe y en español, en particular.

Para finalizar, vale recordar que generalmente las traducciones van acompañadas de alguna limitación. La falta de especificación del contexto puede dar lugar a una interpretación equivocada (polisemia). Pueden existir varias traducciones estándar para una expresión dada (sinonimia), o ninguna (vacío terminológico). Sin embargo, esto se aplica a todos los idiomas y para todos los campos del conocimiento, y no menoscaba la validez de los métodos presentados.

## AGRADECIMIENTOS

Deseo expresar mi agradecimiento a Martha Daza, Julia Aymerich y Olivier Bodenreider por la inestimable retroalimentación brindada. Este proyecto ha sido patrocinado por el Programa de Investigaciones

Intramurales de los Institutos Nacionales de la Salud de los Estados Unidos, Biblioteca Nacional de Medicina, Centro Nacional Lister Hill para las Comunicaciones Biomédicas (LHNCBC). Quedo también muy agradecida al Centro Universitario de Investigaciones Bibliotecológicas de la UNAM por la gentil invitación a esta conferencia.

## REFERENCIAS BIBLIOGRÁFICAS

- [1] Carli A, Calaresu E., “Le lingue della comunicazione scientifica. La produzione e la diffusione del sapere specialistico in Italia”, en *Ecologia linguistica. XXXVI Congresso Internazionale di Studi della Società di Linguistica Italiana*, Roma, 27-74, 2003.
- [2] Curry MJ, Lillis T., “The global imperative to publish in English: Brokering multilingual scholars into English academic publishing”, en *NYS TESOL Idiom*, 34(2), 8-9, 2004.
- [3] Arunachalam S, Singh UN., “Where do Israeli scientists publish?” en *Journal of Information Science*, 14(4), 221-231, 1988.
- [4] Roseblat G, Tse T, Gemoets D, Gillen JE, Ide NC., Supporting access to consumer health information across languages, ICML 9/CRICS 7, Brazil, 2005.
- [5] *UMLS Knowledge Sources*, US National Library of Medicine, Documentation, 2007 AA, <http://www.nlm.nih.gov/research/umls/documentation.html>
- [6] Nelson SJ, Schopen M, Schulman JL, Arluk N., An Interlingual Database of MeSH Translations.
- [7] Fugmann R., Subject analysis and indexing: theoretical foundation and practical Advice, Frankfurt/Main, INDEKS Verlag, 1993, en <http://www.nlm.nih.gov/mesh/intlmesh.html>

- [8] Vivaldi J., Extracción de candidatos a término mediante combinación de estrategias heterogéneas, Barcelona, Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, 2004, Serie Tesis 9. 2001.
- [9] Feliu, J., Relacions conceptuals i terminologia: anàlisi i proposta de detecció semiautomàtica, Barcelona, Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, 2004.
- [10] Gómez González-Jover A, Vargas Sierra C., Córpora comparables y paralelos para la detección de terminología bilingüe: su explotación y uso con herramientas informáticas, VIII Simposio RiTerm, Cartagena de Indias, 2002.
- [11] Barzilay R, Elhadad N. Sentence, Alignment for Monolingual Comparable Corpora, EMNLP, Japan, 2003.
- [12] <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?DB=pubmed>
- [13] Zweigenbaum P, Grabar N., "Automatic acquisition of morphological knowledge for medical language processing", en *Lecture Notes in Computer Sciences*, Vol. 1620, 416-420, 1999.
- [14] Baud RH, Nystrom M, Borin L, Evans R, Schulz S, Zweigenbaum P., "Sentence. Interchanging lexical information for a multilingual dictionary", en *American Medical Informatics Association*, Simposio Anual 2005, 31-5
- [15] Roseblat G, Tse T., "Study of a Spanish Language ClinicalTrials.gov Prototype System", en *American Medical Informatics Association*, Simposio Anual 2006, 659-664.





# Análisis de los tipos de definiciones terminológicas que aparecen en varias áreas del lenguaje especializado

ÉRIKA EHNIS DUHNE

*Centro de Enseñanza de Lenguas Extranjeras  
y Centro de Enseñanza para Extranjeros,  
Universidad Nacional Autónoma de México*

## 1. INTRODUCCIÓN

La acción de definir consiste en identificar el objeto, palabra o concepto que nos ocupa, así como diferenciarlo de otros parecidos a él. Identificar ese elemento —o “*definiendum*” (lo que va a ser definido)—, como lo llamaba Aristóteles en la definición real, se lleva a cabo por otro término o concepto que delimita el significado de ese elemento, que generalmente es su genérico. Dependiendo del grado de complejidad que los términos o conceptos tengan, en la segunda parte de la definición —o “*definiens*” (cómo va a ser definido ese elemento)— estará la cantidad y complejidad de las referencias o elementos cercanos que se utilicen con el fin de llevar a cabo la definición. En el presente estudio se presenta una tipología de las definiciones, que está determinada por sus objetivos y por la naturaleza misma del término, y del área de especialidad donde se presentan. Este análisis es el objetivo primordial del presente estudio.

## 2. LA DEFINICIÓN LEXICOGRÁFICA Y TERMINOLÓGICA

A pesar de que la definición terminológica (DT) trabaja con los lenguajes de especialidad y la definición lexicográfica con la lengua ge-

neral, ambas comparten muchos elementos. Rickert (1960:370-71) aclara a este respecto:

... la tarea de la definición consiste —según Aristóteles— en *concluir* la investigación y en fijar de modo definitivo la “esencia” del correspondiente objeto de investigación. (...) La definición consta del género y la diferencia. El género indica justamente la “esencia” y, por lo tanto, a él debemos subordinar el objeto por definir para conocer su naturaleza o su esencia. La adición de la diferencia sirve para designar el modo particular en que se presenta fenoménicamente la esencia”.

En general, una definición es la expresión en palabras de un concepto desconocido, mediante conceptos y palabras ya conocidos. Para que la definición sea efectiva debe haberse delimitado con anterioridad el concepto perfectamente. La definición debe cumplir con toda precisión el objetivo para el cual fue creada.

Éstas son algunas de las diferencias que existen entre las definiciones lexicográficas y las DT (Lorente 2001:86):

Definición Lexicográfica	Definición Terminológica
Representa significados lingüísticos	Representa conceptos especializados
Semasiológica	Onomasiológica
Significado polisémico de las palabras	Denominación unívoca
Presenta significados por aproximación	Delimita e identifica conceptos
La aproximación del significado puede expresarse a través de prototipos, que conllevan sesgos culturales, que forman estereotipos	La identificación de conceptos se realiza a partir de la representación ordenada de características identificadoras
El contenido de las definiciones de palabras proviene de la competencia lingüística y de la experiencia	El contenido de las definiciones de conceptos proviene del conocimiento especializado
La función prescriptiva de algunos diccionarios no afecta a la definición	La terminografía prescriptiva fija conceptos y denominaciones

### 3. CARACTERÍSTICAS DE LAS DEFINICIONES:

Es necesario establecer características diferenciadoras del término que va a ser definido con otros que compartan conceptos cercanos, para que ese término sea susceptible de confundirse con los otros. Veamos un ejemplo.

Una definición general de “triángulo” es:

“...figura plana limitada por tres líneas rectas.” (García Máynez 1958:46)

En donde “triángulo” es el término que va ser definido (o *definiendum*), y lo que resta es su definición, (o *definiens*), en donde primero se ubica al término en su categoría (su género, que es figura plana), y luego se dan sus características (limitada por...) de tal manera que este elemento no pueda ser confundido con ninguno otro cercano a él.

### 4. TIPOLOGÍA DE LAS DEFINICIONES

Hay muchos recursos y maneras para hacer una definición. El tipo de herramienta y metodología que se utilice al construir una definición depende, por un lado, de la intención y objetivo del autor, y por el otro, de la naturaleza misma del término y del discurso de especialidad que lo generó.

La ecuación sémica de la definición puede utilizar diferentes predicados (como *llamarse, se dice, confiere, donar, dar, etc.*), para llevar a cabo la reconstrucción del significado (Lara 1997:220). Kant cambió la definición de la “esencia de las cosas” —como Aristóteles lo había descrito— hacia su “concepto” (ídem:209).

Hasta aquí nos hemos referido a la definición con un criterio amplio, dentro del lenguaje general. Ahora pasaremos a analizar la definición dentro del contexto de los lenguajes de especialidad.

Hay muchos autores que han establecido diferentes tipología de las DT, con base en diferentes criterios.:

- Mönke (1978)<sup>1</sup> analizó las definiciones y saca como conclusión que se dividían en más de 70 tipos.
- Alan Rey (1977)
- Rey-Debove (1971)
- Robinson (1962)
- etcétera.

He aquí algunos tipos de definiciones en los que coinciden muchas escuelas y autores:

- a) Es la denominación de un concepto (*cefalea/dolor de cabeza*)
- b) Explicación de la clasificación nominal de un concepto (*amigdalitis es la inflamación de las amígdalas*).
- c) Clasificaciones en general (*Los términos complejos se clasifican en términos derivados, términos compuestos y sintagmas terminológicos*).
- d) Indicaciones de algunas características materiales o físicas (*la colitis se manifiesta por diarreas intermitentes, dolor abdominal, flatulencia y, ocasionalmente, náuseas y vómitos*).
- e) Por la función que cumplen (*el acelerador lineal se emplea para el tratamiento radioterapéutico de pacientes con cáncer*).
- f) Por comprensión, se parte de un concepto superordinado que es superior, y se dicen características distintivas (*nogal/planta arbórea que tiene un tamaño...*), (Lorente 2001:92)
- g) Por extensión, se enumeran todos los elementos que forman esa unidad (*Los planetas son Mercurio, Venus, Tierra...*).
- h) Por contexto, que funciona más como ampliación o expansión de la definición, que como una definición misma.
- i) Nominal, a través de sinónimos relativos (*aspirina/calmante*).
- j) Por otros medios, como el uso de dibujos, gráficas, símbolos, fórmulas, fotografías, etcétera.

---

1 Del curso "La arquitectura del texto universitario", por la Dra. Melva Márquez Rojas, de la Universidad de los Andes en Venezuela, profesora invitada en el IINGEN, UNAM.

Hoy en día se analizan las DT y sus contextos definitorios en gran medida debido a un interés mayoritario de poder automatizar la identificación y extracción de éstos.

Una clasificación muy difundida de las DT , que está mezclada con la anterior, es la siguiente:

Existen dos tipos de definición, dependiendo de la combinación de los conceptos que se lleven a cabo, que ya se dijo anteriormente: por *comprensión* y por *extensión*. La primera se refiere a la definición tradicional, que indica el contenido del concepto. El procedimiento consiste en definir el concepto y en diferenciarlo de los demás conceptos de la misma serie, para lo que se parte del concepto genérico más próximo que sea conocido. En el segundo tipo de definiciones se hace la enumeración de todos los conceptos específicos del mismo nivel del que estamos definiendo, aunque se entiende que la lista no puede ser exhaustiva. A veces este tipo de definición resulta más clara por la enumeración, y a veces el primero, pero se pueden combinar ambos tipos para lograr una mayor claridad y precisión.

Las definiciones también se pueden clasificar según su propósito en definiciones descriptivas y prescriptivas. En la primera definición se constata el significado. También es conocida como la definición analítica o real, que vimos brevemente al principio, por los postulados de Aristóteles. El segundo tipo de definición también es conocido en algunos contextos como definición nominal o sintética, que es para fijar el significado.

Gladis Almeida (2007:3-4) y “El grupo de estudio y pesquisas en Terminología” tiene la siguiente clasificación de definiciones:

- Definición mediante análisis
- Definición mediante sinónimos
- Definición mediante paráfrasis
- Definición mediante síntesis
- Definición mediante implicaciones

La tipología de definiciones que hemos encontrado en el análisis de definiciones lexicográficas y terminológicas en el área de la Fonética son:

1. intencional
2. extensional
3. por contexto definitorio
4. descripción
5. explicación
6. una definición dentro de otra definición
7. ostensiva
8. descripción enciclopédica

## 5. LA DEFINICIÓN TERMINOLÓGICA O TERMINOGRÁFICA

Esta definición comparte las mismas características que la definición lexicográfica, pero trabaja con el lenguaje de especialidad. Se trabaja con conceptos y con sistemas de conceptos interrelacionados entre sí en redes de información. De ahí que sea relevante no sólo dar su DT, sino también su ubicación dentro de la red del concepto que representa (Sager 1990:48).

De acuerdo con los principios terminológicos hay normas o criterios establecidos que se siguen en la elaboración de las definiciones, debido a la armonización terminológica que debe existir. La Norma ISO 137 fue elaborada en 2004 por el comité (Cabré, Pozzi, etcétera). Estos son los principios de una definición terminológica:

- El conocimiento del concepto se debe presentar con orden y precisión.
- El concepto es atribuido a una clase (*Venus, sabiduría*).
- El concepto se agrupa por categorías de rasgos distintivos (*cuadrúpedos*).
- Los conceptos y grupos de conceptos pueden ser diferenciados por un proceso de discriminación entre las categorías (*sillas, relación de mesas a muebles*).
- La interacción de las categorías de conceptos puede ser afectadas en su funcionamiento (*captura-almacenamiento-recuperación, reacción química-compuesto químico*).

Tipos de definiciones según los criterios actuales de Terminología en la ISO 137:

### 1. Definiciones intensionales:

Son las definiciones que requieren el mínimo de información:

- indican el concepto inmediatamente superior o inferior
- indican características que diferencian a este concepto de otros parecidos
- señalan los elementos partitivos, sólo si forma parte de la comprensión global del concepto (como el mouse es a la PC)
- llevan marcadores del discurso tales como: es parte de, es una sección/componente/etcétera de...

### 2. Definición extensiva:

- se da cuando el número de elementos que forman el concepto es limitado (elementos finitos que integran ese universo, como especies en peligro de extinguirse, etcétera).

### 3. Definición ostensiva:

- es la definición que señala o muestra los elementos del concepto.
- no es sólo con imágenes o señalamientos reales (gráficas, bosquejos, etcétera). (*ISO 37, ISO/CD 704rev, etcétera*).

## 6. EJEMPLOS DE DT EN VARIAS ÁREAS

A continuación se muestran otros ejemplos de diferentes ámbitos especializados

### a). Derecho ambiental:

*Fuente: Estudio de impacto ambiental  
DIDAM98:167[1] (Krieger, M. Gracia 1998:167)*

LgP D1 109 de 15/03/91, art. 2º, d  
Environmental impact assessment  
ONU92#2056  
estudio de impacto ambiental FAV94 :17

estudio de impacto ambiental FAV94 :17

Estudio sob responsabilidade do proponente, contendo informações significativas, provocadas por esse projecto a curto pu a longo plazo, sobre o ambiente, nas suas componentes biofísicas, económicas, sòcio-culturais e humanas e suas inter-relações.

- LgP
- ESTABLECIMIENTO INDUSTRIAL; RELATORIO DE IMPACTO AMBIENTAL

b) Ejemplo de Química:

Término “Cobalto”

FUENTE: TRIVON96:86[1]

[1] Trilingual Vocabulary of Nuclear Waste management (1996)  
Unión Latina, Boletín de terminología 231, Canadá.

**Cobalt; Co            cobalt(n.M.)    cobalto (n.m.)**  
**(CUB,ESP; MEX)**  
**Co (ESP; MEX)**

**EN A thoug , lustrous, nickel-white or silvery-grey, metallic element in Group VIII of the periodic system.**

**FR Élément métallique de nombre atomique 27, de masse atomique 59.**



c). Otros ejemplos:

**Fonética 1:** Fuente: Guitart, J. M. (2004) *Sonido y sentido*. Washington: Georgetown University Press (p. 288, Glosario

- **Alveolar** (alveolar) adj. Se dice de: a) la región de la cavidad oral construida por la **cresta alveolar**; b) todo sonido que se articula haciendo que la lengua toque o se acerque a dicha región.
- [**cresta alveolar** (alveolar ridge) n. Protuberancia rugosa del cielo de la boca que está situada antes del paladar duro y constituye la región alveolar. Se llama así por estar detrás de los alvéolos de los dientes superiores. p. 291]

**Fonética 2:** Fuente Trask, L. R. (1996) *A Dictionary of Phonetics and Phonology*. USA/Canada: Routledge, p. 18.

- **Alveolar** 1. adj. Pertaining to the **alveolar ridge**. 2. adj. (formerly also **gingival**) (of an articulation), involving the alveolar (Trask:18).
- Ridge as the upper articulator. In this usage, often short for one of the more specific terms **apico-alveolar** and **lamino-alveolar**.
- 3. n. A consonant articulated in this way, such as [s], [d] or [n]. 4.
  - In the **Ladefoged** and **Williamson feature systems**, one of the eleven possible values of the feature **articulatory place**. Ultimately from Latin alveolus 'small cavity', referring to the tooth sockets.

## 7. CONCLUSIÓN

En general, las definiciones mostradas aquí como ejemplos son intencionales, aunque incluyen diferentes tipos de información: acepciones, tipo de fuente, idioma, equivalentes en países (Canadá, etcétera).

Sin embargo, se siguen diferentes políticas en el tipo de definiciones, dependiendo de:

- los organismos que las crean:
- la política que siguen,

sus objetivos,  
sus necesidades y  
el tipo de usuarios.

Pero siempre el objetivo es la claridad y precisión en la definición del término.

## BIBLIOGRAFÍA

Alpizar Castillo, Rodolfo (1997), *¿Cómo hacer un diccionario científico-técnico?*, Buenos Aires, Ed. Memphis.

Arntz, Reiner y Heribert Picht (1989), *Introducción a la terminología*, trad. del alemán por Irazazábal, Jiménez, Schwarz y Yunquera, Madrid, Fundación Germán Sánchez, (col. Biblioteca del libro núm. 64).

Almeida, Gladis Maria de Barcellos (2007), “A definição nos dicionários especializados: proposta metodologica”, en *Debate Terminológico* Núm. 3, en [http://www.riterm.net/revista/n\\_3/Art\\_Barcellos\\_Almeida.pdf](http://www.riterm.net/revista/n_3/Art_Barcellos_Almeida.pdf) (última consulta 13 febrero 2008).

Cabré Cabré, M. Teresa (1993), *La terminología: teoría, metodología, aplicaciones*, prólogo de J.-C. Sager, Barcelona, Ed. Antártida.

Finatto, María José Bocorny (2001), *Definición terminológica: fundamentos teórico-metodológicos para su descripción y explicación (en portugués)*, Brasil, tesis doctoral presentada en abril del 2001 en la Universidad Federal do Rio Grande do Soul.

García Máynez, Eduardo (1958), “Análisis crítico de algunas teorías sobre el concepto de definición”, en *Dianoia: Anuario de Filosofía*, año IV, núm. 4, México: UNAM, Centro de Estudios Filosóficos, 44-63.

ISO 137 (ISO 137, ISO/CD 704rev, etc.).

Lara, Luis Fernando (2004), *De la definición lexicográfica*, México, El Colegio de México, (Col. *Jornadas* Núm. 146).

Lara, Luis Fernando (1997), “Capítulo V. El contenido proposicional del acto: la definición lexicográfica”, en *Teoría del diccionario monolingüe*, México, El Colegio de México, (Serie Estudios de Lingüística y LiteraturaXXXIII).

Lorente, Mercé (2001), “Teoría e innovación en terminografía: la definición terminográfica”, en *La Terminología científico-técnica*, Barcelona, IULATERM, pp. 81-112.

Rey-Debove, Josette (1971), “Capítulo 6: La définition”, en J. Rey-Débove *Étude Linguistique et Sémiotique des Dictionnaires Française Contemporains*, París, Mouton, 180-257, (col. *Approaches to Semiotics*, editado por Thomas A. Sebeok, núm.13).

Rickert, Heinrich (1960), *Teoría de la definición*, México, Centro de Estudios Filosóficos, UNAM, (col. *Cuaderno* 9).

Sager, Juan C. (1990), *A Practical Course in Terminology Processing*, Amsterdam / Filadelfia, John Benjamins Publishing Co., 254 pp.



# La gestión documental aplicada a la elaboración de un banco de conocimiento especializado

ANNA AGUILAR-AMAT, BARTOLOMÉ MESA LAO,  
PILAR SÁNCHEZ-GIJÓN

*Departamento de Traducción e Interpretación  
Universitat Autònoma de Barcelona, España*

## INTRODUCCIÓN

Las relaciones entre Terminología y Documentación se dan desde la Terminología hacia la Documentación y viceversa. Es sabido que la Terminología de los diferentes ámbitos de especialidad sirve a los documentalistas como instrumento para la descripción, indexación y recuperación de la información. La Documentación, a su vez, está al servicio de los terminólogos como fuente indispensable de información para la elaboración de recursos y aplicaciones. A caballo entre estas dos disciplinas trabajan los profesionales de la Traducción a la hora de transmitir y representar conocimiento especializado.

La responsabilidad de un traductor especializado es muy grande, ya que de su intervención puede resultar la correcta transmisión del conocimiento, su deterioro o, en ocasiones, su optimización; ya que como mediador lingüístico el traductor puede gestionar los recursos expresivos de la lengua de llegada con mayor o menor acierto.<sup>1</sup>

---

1 La posibilidad de cuestionar las opciones del experto, o la de proponer alternativas léxicas autóctonas en el caso de la existencia de un préstamo, es una tarea que deberá llevar a cabo el traductor-terminólogo siempre con el consenso del experto.

Si nos centramos en el ámbito de la Traducción como actividad profesional y, en concreto, en el proceso documental subyacente a todo proceso de traducción especializada, nos daremos cuenta de que —de manera directa y explícita— los traductores desarrollan perfiles profesionales vinculados a la selección, descripción y tratamiento de la documentación.

El propósito de este artículo es doble. En primer lugar presentaremos la metodología para elaborar un banco de conocimiento centrándonos en la fase de búsqueda y gestión documental que realizan los estudiantes dentro de la asignatura de Terminología aplicada a la traducción (cuarto curso de la licenciatura de Traducción e Interpretación de la Universitat Autònoma de Barcelona). En segundo lugar, describiremos las características básicas de BACUS (*Base de Coneixement Universitari*); el banco de conocimiento que recoge todos los trabajos de investigación terminológica que elaboran los alumnos y que puede resultar útil —además de para la consulta telemática manual— en sistemas de traducción asistida por ordenador o para la clasificación automática de documentación multilingüe.

## EL PROYECTO BACUS Y EL DESARROLLO DE COMPETENCIAS DOCUMENTALES

El proyecto BACUS constituye un proyecto universitario de creación de recursos terminológicos compartidos aplicados a la traducción. Este proyecto surge del interés particular de los autores por aunar tres disciplinas complementarias:

- La Terminología
- La Documentación
- La Informática aplicada a la Traducción o Tradumática

Estas tres disciplinas juegan un papel importante en el desarrollo de las competencias profesionalizadoras del traductor especializado y se han visto integradas en un mismo proyecto. El objetivo general de este proyecto consiste en la creación de un banco de conocimiento universitario multilingüe. Como proceso, BACUS ayuda a la formación

del traductor especializado y, como producto, constituye un recurso terminológico útil para detectar equivalencias y relaciones conceptuales multilingües. Después de once años de trabajo con los estudiantes que han pasado por la asignatura se ha conseguido alimentar un banco de conocimiento con más de 23.000 conceptos en más de diez lenguas de trabajo.

Para construir de este banco de conocimiento multilingüe, los estudiantes han trabajado la terminología de diferentes ámbitos temáticos en tres lenguas. Las fases de la investigación terminológica multilingüe que siguen todos los alumnos los obligan a desarrollar competencias que después necesitarán en el mundo profesional. Estas competencias pueden agruparse en torno a tres grandes ejes:

1. Desarrollo de competencias documentales basadas en estrategias para la búsqueda y recuperación de información tanto general como especializada.
2. Adquisición de la metodología básica del trabajo terminológico multilingüe.
3. Desarrollo y uso de recursos informáticos de ayuda a la traducción.

El producto final que realizan los alumnos es utilizado y comparado por todos los miembros de la comunidad universitaria (alumnos y profesores). En este artículo nos centraremos en la primera competencia que citábamos: el desarrollo de competencias documentales.

## TRADUCCIÓN ESPECIALIZADA, TERMINOLOGÍA Y DOCUMENTACIÓN

La lengua —y en concreto el uso especial que de ella se hace en cada ámbito temático— es el instrumento básico de comunicación que emplean los especialistas para vehicular conocimiento especializado. La Terminología, por su parte, es el elemento más importante para fijar cognitivamente el modo de denominación en cada ámbito de especialidad. A través de la Terminología, además de ordenar el pensamiento, los especialistas transfieren el conocimiento sobre una materia en una o más lenguas.

La Terminología en tanto que herramienta cognitiva resulta un elemento clave en la formación del traductor especializado. El traductor especializado debe ponerse en la piel quien emite el mensaje (un especialista en un determinado ámbito) y emular sus mismas competencias. Si no lo hace, la traducción difícilmente resultará adecuada. Emular la competencia cognitiva y discursiva del especialista presupone conocer la materia que se traduce, saber expresarla adecuadamente, y hacerlo como lo haría espontáneamente el emisor del texto como experto en un tema y como hablante nativo de una lengua.

En este sentido, la formación de traductores especializados debe estar orientada a la adquisición de competencias que ayuden a simular el rol de un especialista que conoce la materia y emplea la terminología propia del ámbito. Para alcanzar ambos objetivos, el trabajo terminológico y, en concreto, el proceso de búsqueda y gestión documental vinculado a éste se convierte en un factor clave.

## Documentación de trabajo y el material adicional

Los documentos a los que recurre un traductor para resolver los problemas que pueda plantearle su traducción pueden clasificarse según su naturaleza. Para la elaboración del banco de conocimiento, los estudiantes recurren a este tipo de fuentes:

- *Documentación de consulta*: fuentes que contienen información de base sobre la temática en la que centran el banco de conocimiento (bibliografía básica).
- *Documentación de trabajo*: fuentes que constituyen la base material del trabajo terminográfico a partir del cual construyen el banco de conocimiento.
- *Documentación de ayuda o soporte*: fuentes que facilitan y complementan el desarrollo del proceso de elaboración del banco de conocimiento.

Estas fuentes de consulta pueden clasificarse en cuatro tipos:

- *Fuentes secundarias*: bases de datos documentales, centros de gestión terminológica, bibliografías, etcétera.



- *Fuentes de carácter teórico*: obras sobre Terminología, sobre la materia especializada y sobre la lengua o lenguas de trabajo.
- *Fuentes con información procesada*: diccionarios generales y especializados, enciclopedias, bases de datos, léxicos, etcétera (sólo como material de referencia)
- *Fuentes con información metodológica*: normas y orientaciones sobre el método terminográfico.

Los materiales específicos que constituyen el *corpus* de trabajo para la elaboración del banco de conocimiento son siempre textos especializados. Estos textos pueden ser tanto orales (conferencias, clases, presentaciones, etcétera) como escritos (artículos científicos, informes técnicos, manuales, etcétera), a pesar de que siempre se priorizan las fuentes escritas. El conjunto de textos seleccionados es denominado *corpus* de vaciado del trabajo terminológico.

Desde la perspectiva de la lingüística, un *corpus* se define como un conjunto de enunciados susceptible de ser analizado. En Terminología, el *corpus* está formado por un conjunto cuantitativamente representativo de textos relacionados con el tema sobre el cual se trabaja. Este conjunto de textos se utiliza para establecer la nomenclatura del ámbito temático. Los criterios principales a partir de los cuales se construye el *corpus* de vaciado son tres:

- *Criterio temático*: sólo se incluyen textos sobre un tema determinado.
- *Criterio lingüístico*: sólo se incluyen textos en unas lenguas determinadas.
- *Criterio textual*: sólo se incluyen textos correspondientes a unas determinadas situaciones comunicativas y a unos usos lingüísticos determinados.

El *corpus* de vaciado debe ser representativo del campo de estudio. Por ello, el terminólogo debe diversificar sus fuentes con objeto de cubrir el conjunto de nociones propias del campo. En la práctica, la exhaustividad es prácticamente imposible de alcanzar, sobre todo por la dificultad de accesibilidad a las fuentes y por el tamaño del trabajo que los alumnos llevan a cabo.

Entre todo el conjunto de fuentes documentales que se utilizan para elaborar el *corpus* de trabajo, las más utilizadas son las obras especializadas. En la categoría de obras especializadas incluimos los textos que los especialistas producen en el desarrollo de su actividad profesional para comunicarse tanto entre ellos como con las personas de su ámbito profesional (alumnos, clientes, etcétera). La tipología de textos puede ser muy variada: manuales y obras de teoría, actas de congresos, tesis doctorales, artículos publicados en revistas especializadas sobre el tema de estudio, catálogos, guías de usuario, instrucciones de uso, inventarios, normas nacionales e internacionales, leyes, patentes, reglamentos, formularios, etcétera. Las revistas y las publicaciones periódicas especializadas resultan especialmente útiles a la hora de proporcionar información sobre la evolución más reciente del conocimiento en un área determinada.

El *corpus* debe presentar un cierto grado de homogeneidad y las fuentes estudiadas deben responder a un mismo estado sincrónico de la lengua. Se les pide a los alumnos que sólo trabajen con textos comparables cronológicamente. De la misma manera, los diferentes textos del corpus deben representar niveles de lengua similares.

La obtención de información terminológica a partir de un corpus es de gran utilidad para el traductor especializado, ya que representa una muestra fidedigna del uso que los especialistas hacen de la terminología y de las reflexiones que hacen sobre la relación existente entre conceptos y denominaciones. Este es el motivo por el que los alumnos trabajan únicamente con textos que no hayan sufrido ningún tipo de «intervención» lingüística más allá de la que pueda haber hecho el propio autor (el especialista). Esto quiere decir que se evita incluir en el corpus de trabajo textos traducidos, ya que no es aconsejable trabajar con las posibles opciones de traducción adoptadas por la persona que ha trasladado el texto a la lengua de llegada.

Puede suceder que la documentación de un tema burocrático como la política comunitaria europea sólo exista en forma de traducciones, pero en general la consigna debe ser siempre la de buscar y comparar textos escritos por autores expertos para cada lengua original. También puede suceder que una de las lenguas implicadas en la investigación terminológica sea una lengua minoritaria con poca presencia

en la base de datos documental de Internet o que posea muy poco volumen de producción escrita. Es el caso, por ejemplo, del catalán. Esto supone un mayor esfuerzo de investigación puesto que el traductor debe realizar consultas orales a los expertos que emplean esta lengua como vehículo de comunicación.

## Fases del trabajo terminográfico sistemático

El proceso del trabajo terminográfico que siguen los alumnos para la elaboración de un banco de conocimiento responde a las siguientes fases:

1. Definición y delimitación del trabajo
  - 1.1. Presentación del tema y adquisición de conocimiento
  - 1.2. Delimitación del trabajo: tema, destinatarios, finalidad y dimensiones
2. Preparación del trabajo
  - 2.1. Ampliación y selección de las fuentes documentales
  - 2.2. Estructuración del conocimiento y estructuración conceptual
  - 2.3. Redacción del plan de trabajo
3. Elaboración de la terminología
  - 3.1. Confección del corpus de trabajo
  - 3.2. Vaciado terminológico
  - 3.3. Elaboración del fichero de vaciado
  - 3.4. Análisis y revisión del fichero de vaciado
  - 3.5. Migración del fichero de vaciado al banco de conocimiento
4. Supervisión del trabajo
  - 4.1. Análisis y revisión del fichero terminológico
  - 4.2. Resolución de casos problemáticos
5. Presentación del trabajo
  - 5.1. Presentación del banco de conocimiento

El trabajo terminográfico sistemático plurilingüe sigue estas fases para cada una de las lenguas de trabajo. Una vez establecido el fichero terminológico para cada lengua (punto 3.3. de esquema) se procede a correlacionar los términos de las distintas lenguas que corresponden a un mismo concepto. La definición o la ilustración son los elementos básicos que permiten la verificación de las correspondencias. De este primer cotejo de equivalentes se obtiene un listado plurilingüe de términos equivalentes. Para cubrir posibles huecos en los listados de equivalentes puede consultarse material complementario, ampliar el *corpus* de vaciado, consultar lexicografía especializada o bien plantear cuestiones puntuales a especialistas en la materia.

### A propósito del vaciado terminológico

El vaciado es la fase de obtención de información terminológica a partir de la documentación recogida en el *corpus* de trabajo. El alumno recoge en esta fase denominaciones, contextos de uso, definiciones y relaciones conceptuales de manera sistemática para luego volcar esta información en el banco de conocimiento. La inclusión de relaciones conceptuales dentro de la base de datos que realizan los alumnos convierte lo que sería simplemente una base de datos multilingüe en una base de conocimiento navegable mediante hipertexto. Dichas relaciones son de carácter conceptual (jerárquicas o no jerárquicas) o lingüístico (argumentales o lexicosemánticas).<sup>2</sup>

## EL DESARROLLO DE LAS COMPETENCIAS DOCUMENTALES DEL ALUMNO

Como ya apuntábamos, la formación del traductor especializado debe tener muy en cuenta el desarrollo de competencias centradas en el uso de fuentes documentales, así como en el uso de las nuevas tecno-

---

2 Para el traductor es muy importante reconocer en una lectura profunda de los textos las relaciones léxicosemánticas que se establecen entre los conceptos que vehiculan conocimiento especializado. Por ello, BACUS incide sobre todo en este tipo de relaciones y no se limita a establecer simplemente equivalentes entre lenguas.

logías de la información y de la comunicación. El desarrollo de la subcompetencia documental va indisolublemente unido al desarrollo de estrategias instrumentales para la comprensión y reformulación de los textos con los que trabaja el traductor. Por este motivo, y como apunta el grupo de investigación PACTE de la Universitat Autònoma de Barcelona, el desarrollo de la subcompetencia instrumental —entendida ésta como la capacidad para buscar y recuperar información— debe considerarse una parte más del proceso de adquisición de la competencia lingüística y comunicativa del traductor, así como de su competencia extralingüística y cognitiva.

Documentarse implica saber identificar problemas de traducción y categorizarlos (problemas de equivalencia terminológica, problemas culturales, etcétera) para poder elegir luego las fuentes de consulta adecuadas para su resolución. En este sentido, entre los objetivos de la asignatura de Terminología está el completar la formación del traductor especializado mediante la adquisición de competencias de diversa índole para el desarrollo profesional de la Traducción.

La formación de traductores especializados exige el desarrollo de competencias instrumentales basadas en la búsqueda y recuperación de información. Desde esta perspectiva, la formación del traductor del siglo XXI nos obliga a ir más allá de la perspectiva que sitúa a la Documentación como disciplina auxiliar y básicamente instrumental, para situarla como motor de cohesión, evaluación y difusión del conocimiento especializado.

Durante su formación, los estudiantes desarrollan competencias específicas del perfil del traductor especializado como, por ejemplo, la competencia de adquisición de información mediante procesos de búsqueda y recuperación de información a través de Internet, así como competencia instrumental en el uso de sistemas de recuperación de información a través de Internet.

## BACUS: DESCRIPCIÓN, FINALIDAD Y ALCANCE DEL BANCO

BACUS (1996-2007) es un banco de conocimiento especializado que resulta de la suma de trabajos de investigación terminológica multilingüe sobre distintos temas y distintas lenguas que realizan los alumnos

de cuarto curso de la Facultad de Traducción e Interpretación de la Universidad Autónoma de Barcelona.

Los trabajos incluidos en BACUS comparten un diseño común y conservan la autoría y la identificación de cada trabajo. En el banco hay trabajos que tienen orígenes, finalidades e informaciones diferentes, pero todos ellos tienen una estructura compartida basada en una serie de campos obligatorios y una codificación estándar para determinadas informaciones. Se calcula un incremento aproximado de BACUS en unos 900 conceptos nuevos por año académico.

Las áreas de trabajo son, básicamente, técnicas y científicas, pero no se excluyen otro tipo de ámbitos temáticos. Los alumnos escogen el área temática libremente, acogiéndose (si lo desean) a las propuestas de los profesores de la comunidad universitaria que se prestan como asesores expertos. La ayuda del experto resulta clave cuando, a partir de la lectura de los textos del corpus de trabajo, surgen dudas sobre el sistema conceptual de un determinado ámbito temático.

Cada concepto sobre el que trabajan los alumnos está orientado temáticamente a través de tres ejes: el área temática, el título del trabajo (que recoge el nivel máximo de especialidad) y el código en el nomenclátor de la UNESCO que más se ciñe a la especialidad seleccionada.

La definición del banco de conocimiento se ha realizado con el gestor terminológico MultiTerm y pretende ser una plataforma para representar conocimiento de manera jerarquizada y relacional. Toda la información terminológica que se recoge en el banco de conocimiento en formato MultiTerm es fácilmente exportable a formato texto, lo que significa que todos los datos pueden migrarse fácilmente a otros sistemas de gestión terminológica.

Fruto del trabajo documental que realizan los alumnos es la construcción de bancos de conocimiento que recogen, como mínimo, los 45 conceptos más representativos de la especialidad que han escogido los estudiantes en grupos de tres o cuatro personas. Cada registro del banco de conocimiento recoge un concepto con las diferentes denominaciones que éste recibe en cada una de las lenguas de trabajo, junto con una serie de relaciones conceptuales (hiperónimos, hipónimos, causa, efecto, material, color, etcétera) vinculadas a cada concepto.

Los registros también incluyen información administrativa (como por ejemplo, autores y asesores del trabajo), información temático-cognitiva de carácter macro y microestructural (por un lado, dominio, título del trabajo y código en el nomenclátor de la UNESCO y, por el otro, delimitación del concepto mediante definición y explicitación de las relaciones con otras unidades conceptuales), e información de carácter lingüístico (denominaciones, contextos de uso, etcétera).

## CONCLUSIONES

Ejercer la traducción especializada supone conocer los elementos metodológicos y los recursos disponibles necesarios para resolver los problemas terminológicos que plantea cada traducción. Este es el motivo por el que el currículum del traductor debería prestar especial atención a la formación metodológica en el ámbito de la Documentación.

En el marco de la asignatura de Terminología aplicada a la Traducción, consideramos que la investigación terminológica multilingüe que realizan los alumnos constituye un buen modo de desarrollar habilidades instrumentales de búsqueda y recuperación de información que, en conjunto, contribuyen al desarrollo de su competencia traductora.

El análisis de textos especializados con el objetivo de detectar y organizar las unidades que vehiculan el conocimiento y su posterior codificación en un banco constituye una buena estrategia de formación. En este sentido, el proyecto BACUS ayuda a desarrollar competencias estratégicas e instrumentales para la formación de un traductor profesional polivalente (capaz de consultar recursos terminológicos, pero también capaz de crearlos y gestionarlos). El resultado de la investigación de los alumnos se convierte así en un recurso útil para toda la comunidad universitaria: estudiantes de traducción, profesorado o aprendices de especialista que necesiten leer bibliografía en otras lenguas.

## BIBLIOGRAFÍA

- Aguilar-Amat, A. (2002), «Traducción, computación, utopía». Conferencia impartida en la sede de la Comisión Europea en Bruselas, <http://europa.eu.int/comm/translation/bulletins/puntoycoma/73/pyc734.htm>, [consulta: 12/02/07]
- Aguilar-Amat, A.; Parra, J.; Piqué, R.; Sánchez-Gijón, P.; Vilar-nau, J. (2000), «Reciclatge del traductor professional», José Chabás *et al.* (ed.), *Actes del I Congrés Internacional de Traducció Especialitzada*, UPF, Barcelona 2-4 de març de 2000, Barcelona, Universitat Pompeu Fabra.
- Aguilar-Amat, A.; Mesa, B.; Sánchez Gijón, P. (2006) «La formación del traductor especializado y la creación de recursos terminológicos compartidos», en *Actas del I Congreso Internacional de Traducción Especializada*, Buenos Aires, Colegio de Traductores Públicos de la Ciudad de Buenos Aires, ISBN-10: 987-96910-7-5, ISBN-13: 978-987-96910-7-6
- Cabré, M. Teresa (1999), *La terminología: representación y comunicación*, Barcelona, IULA, UPF.
- Cabré, M. Teresa (2000), «Elements for a Theory of Terminology: Towards an Alternative Paradigm», en *Terminology International Journal of Theoretical and Applied Issues in Specialized Communication*, 6, 1, pp: 1-23.
- Cabré, M. Teresa (2000), «Terminología y Documentación», Gonzalo García, Consuelo; García Yebra, Valentín (eds.), en *Documentación, terminología y traducción*, Madrid, Síntesis, Fundación Duques de Soria, pp: 31-43.
- Faber, P. (2002), «Investigar en Terminología», Pamela Faber y Catalina Jiménez Hurtado (eds.), *Investigar en Terminología*, Interlingua, Granada, Editorial Comares, pp: 3-23.



Pacte (2005) «Primeros resultados de un experimento sobre la Competencia Traductora», en *Actas del II Congreso Internacional de la AIETI (Asociación Ibérica de Estudios de Traducción e Interpretación) Información y documentación*, Madrid, Publicaciones de la Universidad Pontificia Comillas, pp. 573-587.

Sánchez-Gijón, Pilar. (2004), *L'ús de corpus en la traducció especialitzada*, Barcelona, Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, 352 (Serie Materials, 4) ISBN: 84-96367-08-8



## El tratamiento terminológico en la Lista de Encabezamientos ARMARC para Bibliotecas Pequeñas

GLORIA ROJAS LEÓN

*Consultores Nacionales Asociados, Colombia*

AGEO GARCÍA BARBABOSA

*Tulane University, E.U.A.*

Desde comienzos de los años 70, los medios para generar riqueza han venido trasladándose de los sectores industriales a los sectores de servicios, como resultado de la llamada *Sociedad de la Información*. Las sociedades modernas se caracterizan por tener la mayor parte de los empleos en la generación, almacenamiento y procesamiento de todo tipo de información y parece que ya no estarán asociados a las fábricas de productos tangibles, sin que esto signifique la desaparición de dichas fábricas.

La *Sociedad de la Información* se ve afectada por nuevos entornos económicos, sociales y tecnológicos. Algunos fenómenos propios de esta “nueva configuración” de sociedad, tales como la “explosión de la información” y la aparición de las llamadas nuevas tecnologías de información y comunicación (TIC), han hecho necesaria la reorientación de procesos bibliotecológicos que antes eran considerados intocables y por consiguiente inmejorables.

Los efectos asociados a estos fenómenos tales como el incremento en la disponibilidad de bases de datos computarizadas, la evolución de tareas manuales en procesos automatizados, la aparición de nuevas técnicas de comunicación y el aumento en el volumen de información manejado por el común de las personas, han hecho que la creación de

vocabularios normalizados sea necesaria para recuperar y administrar esa cantidad de información existente.

Por otra parte problemas bibliotecológicos comunes al crear lenguajes estructurados como la ambigüedad en la terminología dada por características lingüísticas propias de cualquier idioma tales como la sinonimia, la homonimia, los acrónimos, las diferencias en la pronunciación, la forma de las palabras y el tratamiento jerárquico, se suman a los factores anteriormente mencionados como problemáticas que deben resolverse mediante lenguajes formales que tiendan puentes dinámicos para la comunicación a alta velocidad, el almacenamiento de la información y la estructuración de sistemas de recuperación eficientes.<sup>1</sup>

Conscientes de esta problemática y su impacto en la actividad diaria de las bibliotecas, Rojas Eberhard Editores y Consultores Nacionales Asociados, CNA LTDA., empresas del Grupo Rojas Eberhard de Colombia, se unieron desde fines del 2005 para desarrollar un proyecto bandera en la conformación de un Archivo de Autoridades bajo el formato MARC (ARMARC), que en principio surgió como un subproducto del trabajo de Outsourcing en Catalogación que CNA Ltda., ha desarrollado para varios programas y bibliotecas en Colombia, trabajo que a partir del segundo semestre del año 2006 se convirtió en un proyecto permanente como resultado de un esfuerzo de investigación de fuentes nacionales e internacionales, realizado por un equipo de bibliotecarios colombianos, con la asesoría internacional del Maestro Ageo García, bibliotecario mexicano, experto internacional en las áreas de catalogación y clasificación, quien ha estado colaborando con Rojas Eberhard Editores, como miembro del Comité Editorial de las últimas ediciones del Sistema de Clasificación Dewey (SCDD) en español, y como revisor y traductor de obras como las Reglas de Catalogación Angloamericanas y el Formato MARC, además de estar trabajando en varios esfuerzos internacionales en esta área de encabezamientos y autoridades de materia, en programas de IFLA y OCLC

---

1 *Thesaurus of scientific, technical, and engineering terms*, Cambridge, Hemisphere Publishing Corporation, 1988, p. vii.

Este proyecto ha sido bautizado con la sigla ARMARC (Archivo de Autoridades de Materia en Formato MARC), porque se está desarrollando con base en la normatividad del formato MARC21. Este proyecto contempla cuatro productos: 1. Las Autoridades de Materia para Bibliotecas Pequeñas, en formato impreso, 2. Las Autoridades de Materia para Bibliotecas Mayores, en formato impreso, 3. El ARMARC-en línea para consulta vía Internet y 4. La base de Autoridades de Materia de ARMARC.

Para el desarrollo del proyecto ARMARC, también se consideró la problemática resultante de las dificultades de la mayoría de las bibliotecas en América Latina, para mantener una actividad continuada y consistente en el desarrollo de las autoridades de materia, debido principalmente a razones presupuestales.

La experiencia separada de CNA Ltda y Rojas Eberhard, ha permitido descubrir que las bibliotecas pequeñas, es decir las que tienen colecciones menores o iguales a 20,000 volúmenes, necesitan una lista de autoridades de materia que sin ser tan desarrollada se ajuste a sus necesidades para el análisis y recuperación de la información de sus colecciones.

**ARMARC PARA BIBLIOTECAS PEQUEÑAS**, nace de las necesidades antes mencionadas, pero parte de la experiencia en el desarrollo de ARMARC para Bibliotecas Mayores y de una lista de encabezamientos para bibliotecas pequeñas de hace 20 años creada por estudiantes de bibliotecología de la Pontificia Universidad Javeriana bajo la dirección de Octavio Rojas León, en ese entonces director del Departamento de Bibliotecología. Esta lista debió pasar por un proceso de actualización y ampliación que se ajustara a las necesidades actuales de las bibliotecas pequeñas, las cuales han sido impactadas por los cambios en la ciencia, la tecnología y los cambios sociales, culturales, políticos y económicos de nuestra región (América Latina). También se debió ajustar a los cambios en el sistema educativo, el cual ha venido incliniéndose cada vez más hacia la promoción de la educación tecnológica (y comercial) caracterizada por promover competencias en los estudiantes que les permitan orientarse hacia una carrera profesional e ingresar al mundo laboral con unas destrezas desarrolladas desde el colegio.

**ARMARC PARA BIBLIOTECAS PEQUEÑAS®** está diseñado para un mercado hispanohablante centrado en las llamadas Bibliotecas Pequeñas, y desde luego hace uso de la experiencia acumulada (Véase la Bibliografía Consultada), pero evalúa e incorpora diferencias idiomáticas (vocabulario, sintaxis, etcétera) y culturales (especialmente en anatomía, arquitectura, derecho, etnias, historia, medicina, religión, etcétera), como un punto fundamental para la consistente estructuración de esta herramienta.<sup>2</sup>

Como es normal, **ARMARC PARA BIBLIOTECAS PEQUEÑAS®** es una herramienta cuyo objetivo fundamental es proporcionar el término más apropiado para un concepto dado, con el fin de ser utilizado como “encabezamiento de materia” en procesos de análisis documental en bibliotecas pequeñas; tiene 5,313 términos autorizados y las áreas con más desarrollo en la lista son: agricultura, anatomía, arquitectura, derecho, educación, etnias, historia, medicina y tecnología de la información.

“Entendiéndose por encabezamiento de materia la palabra o palabras con que se define el contenido intelectual de la obra. Incluye tópicos, personas, organizaciones y lugares. De esta manera, además de las entradas de autor, título, etcétera, las obras quedan representadas en el catálogo público mediante los temas que tratan; esto les permite a los usuarios saber qué obras posee la biblioteca sobre un asunto determinado. Esta función de los encabezamientos de materia los convierte en importantes auxiliares de los servicios que presta la biblioteca, puesto que le permiten al usuario y al bibliotecario localizar en forma rápida y segura el material necesario para contestar preguntas sobre un tema dado representado en un lenguaje controlado.”<sup>3</sup>

En general el proyecto ARMARC se ha previsto como un proceso de continuo desarrollo a cargo del personal que ha venido trabajando en su producción hasta ahora, con la participación activa del grupo que labora en la línea de Outsourcing de Catalogación en CNA Ltda. ([www.cnaltda.com.co](http://www.cnaltda.com.co)). En una segunda etapa se contará con la parti-

---

2 *Taller de creación de registros de autoridad con RCA2 y MARC21. (2006)*, [en línea], San Luis Potosí, México, Universidad Autónoma de San Luis Potosí. Sistema de Bibliotecas. recuperado en 2007, de <http://cictd.uaslp.mx/autoridades/>

3 COLEGIO DE MÉXICO. BIBLIOTECA DANIEL COSIO VILLEGAS (2000), [en línea], México, El Colegio, Recuperado en 2006, de <http://biblio.colmex.mx/nuevo/dcbpol3.htm>

cipación de algunas bibliotecas de la región, para convalidar, corregir y enriquecer la base de términos autorizados. Desde luego la coordinación del programa estará siempre atenta a todas las observaciones y sugerencias provenientes de la comunidad bibliotecaria de habla hispana.

El proyecto ARMARC consolida el trabajo del Grupo Rojas Eberhard, con la editorial a la cabeza, para producir y tener las herramientas más importantes necesarias para el desarrollo consistente y normalizado del trabajo de procesamiento técnico en las bibliotecas.

De hecho, el Armac para Bibliotecas Pequeñas, en formato impreso, ha sido pensado como acompañante obligado a la edición 14 abreviada del Sistema de Clasificación Dewey, obra que también debe aparecer en el segundo semestre del año en curso.

## PROCESO PARA LA VALIDACIÓN DE MATERIAS O ENCABEZAMIENTOS

### Equipo de Desarrollo

El equipo de trabajo de **ARMARC PARA BIBLIOTECAS PEQUEÑAS®** se ha ido fortaleciendo con personal profesional de un alto perfil y experiencia en el desarrollo y fortalecimiento de lenguajes documentales. Este perfil contempla aspectos fundamentales para el desarrollo y validación de autoridades de materia tales como: el manejo del idioma inglés, conocimiento experto del castellano (gramática, taxonomía, sintaxis, etcétera), conocimiento de obras de consulta (generales y especializadas), experticia en lenguajes controlados (LCSH, Tesoros, etcétera), amplia cultura general, raciocinio (capacidad de argumentación y negociación) y por supuesto sentido común.

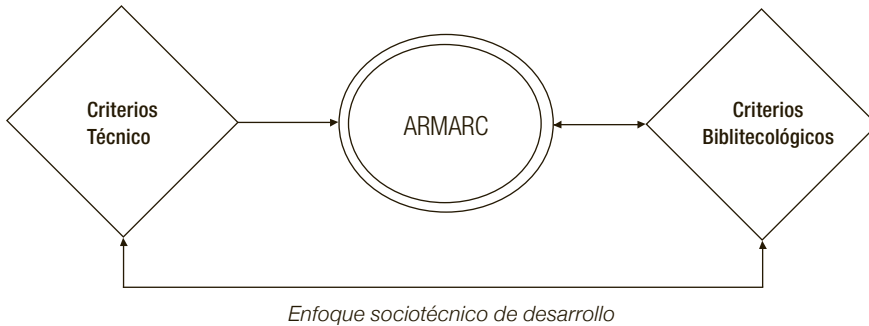
### Validación De Materias: Criterios De Desarrollo<sup>4</sup>

La *validación* es un proceso manual o automático en el que los datos y los indicadores de contenido de un registro coinciden con estándares y se notifica sobre cualquier error o inconsistencia que se

---

<sup>4</sup> Documento validación: guía para la aplicación de los criterios para validación de autoridades de materia. (2006), [en línea], San Luis Potosí, México, Universidad

detecte”<sup>5</sup> **ARMARC PARA BIBLIOTECAS PEQUEÑAS®** contempló dos procesos de validación interrelacionados de tipo bibliotecológico y técnico para asegurar la calidad de la lista.



### Criterios Bibliotecológicos

La totalidad de los 5313 términos autorizados que componen **ARMARC PARA BIBLIOTECAS PEQUEÑAS®** fueron cuidadosamente seleccionados y desarrollados a partir de los siguientes criterios bibliotecológicos, fundamentales en procesos de validación de materias.

CRITERIO	TIPO	EXPLICACIÓN	EJEMPLOS
FORMALES	Corrección lingüística	<p><b>Ortografía</b> Forma correcta del término bajo las normas que regulan la escritura de la lengua, cuidando la aparición de errores de tipografía.</p>	<p>Sicología = Psicología Naylón = Nailon</p>
		<p><b>Sintaxis</b> Términos sintácticamente correctos, contruidos bajo las reglas que gobiernan la estructura del lenguaje. (Coordinación y unión de palabras)</p>	<p>Crítica del arte = Crítica de arte</p>

Autónoma de San Luis Potosí, Sistema de Bibliotecas, recuperado en 2007, de <http://cictd.uaslp.mx/autoridades/>

5 *Ídem.*



**Traducción para transferir información**

CRITERIO	TIPO	EXPLICACIÓN	EJEMPLOS
FORMALES	Forma adecuada del encabezamiento	<p><b>Ordenación</b> Clarificación en el orden de la carga semántica dentro del encabezamiento (Invertido o directo)</p>	Energía, Consumo de = Consumo de energía; Estado, Jefes de = Jefes de estado
		<p><b>Subdivisiones</b> Construcción de subdivisiones estructuradas:</p> <ul style="list-style-type: none"> <li>• Relaciones de subordinación correspondidas</li> <li>• Pertinencia y utilidad</li> <li>• Especificidad</li> <li>• Alcance</li> </ul>	Enfermedades transmisibles--Prevención; Medicina--Investigaciones; Química--Aparatos e instrumentos
		<p><b>Paréntesis</b> Utilidad de la palabra entre paréntesis para la aclaración y/o especificación de significados en términos confusos o poco comunes</p>	Motores (Mecánica), Valores (Filosofía)
SEMÁNTICOS	Traducción	<p><b>Calidad de la traducción</b> Los términos conservan el mismo significado que el atributo en inglés.</p>	Accidentes nucleares (Nuclear accidents) ; Acelgas (Swiss chard) ; Construcciones metálicas (Building, Iron and steel) ; Baladas (Ballads) ; Delfines (Dolphins) ; Latinoamericanos (Latin Americans)
	Correspondencia conceptual	<p><b>Precisión y claridad</b> Los términos corresponden exactamente a lo que se quiere expresar, sin posibilidad de ambigüedad, vaguedad o tergiversación.</p>	Actividades atléticas ; Deportes acuáticos ; Clavados
LÓGICOS	Consistencia	<p><b>Equilibrio terminológico</b> Consistencia terminológica con otros términos usados en el desarrollo. Los términos han sido utilizados de igual manera en otros encabezamientos o como encabezamientos separados.</p>	Indumentaria, Prendas de vestir, Moda; Adolescencia, Juventud, Muchachos, Pubertad
	Coherencia Estructural	<p><b>Jerarquías</b> Estructura de relaciones clara para la identificación de encabezamientos generales y específicos relacionados con el término.</p>	Atletismo: Salto (Atletismo), Lanzamiento de jabalina, Maratones (Atletismo)

CRITERIO	TIPO	EXPLICACIÓN	EJEMPLOS
LÓGICOS	Coherencia Estructural	<b>Voces sinónimas</b> Control lingüístico de otros términos aceptados con el mismo significado.	Dirigentes de empresas, Ejecutivos; Clínica médica, Medicina clínica
		<b>Voces polisémicas</b> Control lingüístico de la multiplicidad semántica de los términos.	Valores ; Mercurio ; Sellos ; Papas ; Americanismos
	Correspondencia interna	<b>Correspondencia</b> Control de las reglas internas de la lista para la formación de encabezamientos.	Ética legal, Abogados- -Ética profesional; --Enfermedades infantiles, Niños—Enfermedades
SOCIALES	Accesibilidad para el usuario	<b>Facilidad de Uso</b> Uso del término • Identificación • Comprensión • Armonía (estructura del término para búsqueda natural por el usuario) • Lenguaje natural vs. Lenguaje documental • Semántica (significado claro para especialistas y usuarios no especializados).	Aves rapaces; Economía--Teorías; Hematoma
	Ausencia de ideología o prejuicios implícitos	<b>Responsabilidad</b> Construcción de términos sin prenociones científicas o sociales ni connotaciones ofensivas para grupos sociales o religiosos.	Arte negro ; Mahometanos ; Marranos (Judíos conversos) ; Trabajo de la mujer
PRAGMÁTICOS	Uso en otros catálogos	<b>Benchmarking investigativo</b> Investigación, comparación y desarrollo de terminología en catálogos generales y especializados nacionales e internacionales.	<b>Birds of prey [LCSH]</b> , traducido como: • ARMARC: Aves rapaces. • LEMB: (No tiene). • CSIC: Aves de presa (Término usado en España, menos común en América Latina).

## Crterios técnicos

Además del proceso de validación bibliotecológico descrito anteriormente, los 5313 términos autorizados de **ARMARC PARA BIBLIOTECAS PEQUEÑAS®** han pasado por un proceso de validación técnica de los registros de autoridad, que ha permitido avalar y respaldar la

información colocada en cada uno de los campos de la estructura del registro MARC21 para autoridades de materia.

El equipo de trabajo de **ARMARC PARA BIBLIOTECAS PEQUEÑAS®** trabajó rigurosamente con base en el formato MARC21 para autoridades de materia, específicamente en las siguientes etiquetas:

Elementos y orden de la revisión:

ETIQUETA	DENOMINACIÓN	TIPO	SC $\alpha$	DEFINICIÓN
1XX	<b>ARMARC</b> Encabezamiento	NR	<u>a</u>	Campos que contienen la forma autorizada utilizada en registros de encabezamientos autorizados o de encabezamientos con extensión.
	<b>MARC 21</b> Encabezamiento (nombre personal, corporativo, reunión o congreso, temático, geográfico, título uniforme)			
360	<b>ARMARC</b> Referencia compleja de "véase además"	R	<u>i</u> <u>a</u>	El texto explicativo y los encabezamientos a los que remite el envío de relación; los cuales se requieren explicitar cuando existen relaciones entre un encabezamiento temático establecido y otros encabezamientos temáticos autorizados, que no pueden ser expresadas adecuadamente mediante una o más referencias simples de "véase además", generadas a partir de campos 5XX.
	<b>MARC 21</b> Referencia compleja de "véase además" - materia			
4XX	<b>ARMARC</b> UP (Usado Por)	R	<u>a</u>	Se utilizan en registros de encabezamientos autorizados (o de encabezamiento con subdivisión) para crear un envío de la forma no utilizada hacia la forma autorizada correspondiente.
	<b>MARC 21</b> Referencia de envío (nombre personal, corporativo, reunión o congreso, temático, geográfico, título uniforme)			
5XX	<b>ARMARC</b> Término General (TG)	R	<u>a</u> <u>w</u> <u>g</u>	Se utilizan en registros de encabezamientos autorizados (o de encabezamiento con subdivisión) para crear un envío de la forma autorizada hacia un encabezamiento establecido relacionado.
	<b>MARC 21</b> Referencias de envío "Véase además" para nombre personal, nombre corporativo, reunión o congreso, título uniforme, temático y geográfico			

*Primer Simposio Internacional sobre Organización...*

ETIQUETA	DENOMINACIÓN	TIPO	SC#	DEFINICIÓN
5XX	<b>ARMARC</b> Término Específico (TE)	R	<u>a</u> <u>w</u> <u>h</u>	
	<b>MARC 21</b> Referencias de envío "Véase además"			
5XX	<b>ARMARC</b> Término Relacionado (TR)	R	<u>A</u>	
	<b>MARC 21</b> Referencias de envío "Véase además"			
670	<b>ARMARC</b> Fuente productiva de información	R	<u>a</u>	La cita de una fuente consultada, en la cual se localizó información relativa a un encabezamiento 1XX; contenida en un registro de encabezamiento establecido (o de encabezamiento con subdivisión), en un registro de subdivisión o referencia. El campo puede contener también la información localizada en la fuente.
	<b>MARC 21</b> Fuente en la que se localizaron los datos			
680	<b>ARMARC</b> Nota general de Usase	R	<u>i</u> <u>a</u>	Información general acerca de un encabezamiento 1XX para el cual no se ha definido una nota especializada. Indica cómo debe usarse un encabezamiento.
	<b>MARC 21</b> Nota general con despliegue al público			
681	<b>ARMARC</b> Nota Envío a otra nota	R	<u>i</u> <u>a</u>	Una nota que documenta el uso de un encabezamiento temático 1XX (o un encabezamiento de subdivisión) utilizado como un ejemplo (o referencia) en los campos 260, 360 y/o 680 de otro registro de autoridad.
	<b>MARC 21</b> Nota de envío a ejemplos de encabezamiento temático			
7XX	<b>ARMARC</b> Traducción y fuente	R	<u>a</u>	La inclusión de un campo ligado dentro de un registro de encabezamiento establecido (o de encabezamiento y subdivisión autorizados) puede utilizarse para nombres equivalentes en tesauros en idiomas múltiples, encabezamientos equivalentes de términos temáticos en diferentes sistemas de autoridades, un término temático (campo 150) y el mismo término o uno similar que se utiliza como subdivisión temática (campo 78X), un encabezamiento de nombre geográfico (campo 151) y el mismo término o uno similar que se utiliza como subdivisión
	<b>MARC 21</b> Asientos ligados al encabezamiento (nombre personal, nombre corporativo, reunión			

ETIQUETA	DENOMINACIÓN	TIPO	SC $\alpha$	DEFINICIÓN
	o congreso, título uniforme, temático y geográfico)	R	<u>a</u>	geográfica (campo 751), un término de género/forma (campo 155) y al mismo término o uno similar que se utiliza como subdivisión de forma (campo 755).

▣ **Subcampos.** *El formato MARC 21 para autoridades de materia maneja múltiples subcampos para cada etiqueta. Por motivo de espacio se citan los más importantes en el desarrollo. Para consultar la versión completa en español visite <http://www.loc.gov/marc/authority/ecadspsa.html>*

## BIBLIOGRAFÍA

Cardona, Silvia; Obando, María Margarita; Rodríguez, Lourdes, *Tesaurus del sistema financiero colombiano, Medellín*: Grupos de información especializada, 1993, 119 p.

Dávila Pessagno, María Inés, *Tesaurus de Trabajo Social*. 2a.ed. actualizada y ampliada, Bogotá, Rojas Eberhard, 2005, 188 p.

*Ei Thesaurus*. Hoboken, NJ., Engineering Information, 1992, ix, 790 p.

Formato MARC 21 para Registros de Autoridades: versión concisa / MARC Network Development and Standards Office, traducción y adaptación de Ageo García, Bogotá. Rojas Eberhard Editores, 2002.

Gobierno de la ciudad de Buenos Aires; Ministerio de cultura ; Programa de tecnologías aplicadas a la Cultura, *Lista de Términos Museológicos de Arte del Catálogo Colectivo Accede [en línea]*, Buenos Aires, ActitudBsAs, 2006 [fecha de consulta: septiembre de 2006], disponible en: <http://www.buenosaires.gov.ar/areas/cultura/dgmcultura/index.php>

***Primer Simposio Internacional sobre Organización...***

Jiménez Piano, Marina, *Control de autoridades: encabezamientos de materia. Normativa para su redacción (Resumen)*, Castilla-Lamancha: Universidad de Castilla-La Mancha, 2001, [fecha de consulta: octubre de 2006], disponible en: <http://www.geocities.com/cienencinas/resumenmaterias.htm?200631>

Library of Congress Rule Interpretation, 2nd ed., Washington, D. C., Cataloging Distribution Service, 1999, 2v.

Library of Congress, Subject Cataloging Manual : Subject headings, 6th ed., Washington, D.C., Cataloging Distribution Service, 1999, 4 v.

MARC Standards  
<http://lcweb.loc.gov/marc/>

Oficina Internacional de Educacion, *Tesaurus de la educación UNESCO:OEI*, 5a.ed. París, IBEData, 1991, xii, 144 p.

Organización de las Naciones Unidas para la Agricultura y la Alimentación, AGROVOC: *Tesaurus agrícola multilingüe*. 4a.ed. Roma, FAO, 2000, xix, 701, B15 p.

Programa de Cooperación en Catalogación (PCC) de la Biblioteca del Congreso (LC) y sus componentes:  
NACO (Base de registros de autoridades de nombres)  
SACO (Base de registros de autoridades de materia)  
<http://www.loc.gov/catdir/pcc/pccinfsp.html>

Reglas de Catalogación Angloamericanas, 2a ed., Rojas Eberhard Editores, 1998.

*TALLER de creación de registros de autoridad con RCA2 y MARC21*, [en línea], San Luis Potosí, México, Universidad Autónoma de San Luis Potosí, Sistema de Bibliotecas, 2006 [fecha de consulta: 2007], disponible en: <http://cictd.uaslp.mx/autoridades/>

## Procesos en la traducción de “Subject Headings for School and Public Libraries”

JOANNA F. FOUNTAIN

*Sam Houston State University, E.E.U.U.*

**E**n 1990 recibí una llamada de una bibliotecaria en el distrito escolar de Eanes, en Austin, Texas. Habían usado la lista *Sears List of Subject Headings* para controlar los encabezamientos en los aproximadamente veinte catálogos escolares durante muchos años. Pero habían descubierto que les costaría menos trabajo si usaran la lista de la biblioteca del Congreso, *Library of Congress Subject Headings (LCSH)*. Estos términos son los que se usan en CIP (Cataloging in Publication), la catalogación en la fuente que se encuentra en la mayoría de libros que se compran para las bibliotecas escolares y públicas en los Estados Unidos. Porque en esos años no era tan común comprar los registros MARC directamente de los vendedores de libros, y todavía no se podían descargar fácilmente de la Web, los bibliotecarios tenían que buscar cada encabezamiento que encontraban en la CIP en su copia de *Sears*, y cambiar los que no eran iguales.

El superintendente del Distrito había decidido comenzar un proyecto central de automatización que incluiría las funciones de las bibliotecas, y eso haría posible, entre otras ventajas, comprar registros ya en formato MARC. La mayoría de los registros MARC, especialmente los que se usan en bibliotecas escolares y públicas, vienen originalmente de la Biblioteca del Congreso de los Estados Unidos (LC). Los

vendedores de libros los dejan tal como los encuentran, o les agregan datos por los cuales cobran aparentemente poco o nada. En aquella época era más obvio el costo de agregar encabezamientos de varias listas; hoy ya no es tanto así, porque si no hay cargo evidente se ha incluido el costo de servicios de catalogación en el costo del libro y otros materiales que se compran. En el distrito de Eanses decidieron no pagar el costo adicional y aceptar los registros como venían originalmente de la LC, haciendo localmente cambios mínimos.

Para que esto funcionara, tenían primero que unir todos sus registros en una sola base de datos, o catálogo unido, y ver lo que tenían de encabezamientos. Resultó una lista bastante buena que se necesitaba redactar más o menos a nivel mínimo. Lo más importante quedaba —la comparación de cada encabezamiento— y cada subdivisión se comparaba con la de la LC. Lo que complica esta tarea es la lista de excepciones y reglas especiales producida para catálogos juveniles como parte del “Annotated Card Program” (*LCSH/AC*). El equipo encargado de la implementación de este programa forma parte de la división de catalogación de la historia y la literatura de la LC, y desde los años sesenta ha creado y a veces redactado una lista de términos más o menos comunes que sustituyen por otros que los niños y jóvenes no suelen saber o usar en sus búsquedas para libros e información. Además, siguen reglas que reflejan una manera juvenil de pensar; por ejemplo, no usan términos o subdivisiones geográficas para el país relacionado; —en este caso, los Estados Unidos.

El proceso de comparación, redacción y cambios (afortunadamente globales) al archivo de encabezamientos siguió por aproximadamente un año, y se hizo en dos fases. La primera fue la de comparar la copia impresa con la lista en los libros de la LC, y la segunda fue hacer el teclado y las funciones relacionadas en la computadora. Cuando terminé tuve una reunión con la bibliotecaria que había coordinado el proyecto, y le comenté que había muchas bibliotecas que harían lo mismo, pero había falta de fondos o personal para llevarlo a cabo. Al decir eso se me ocurrió pedirle permiso al distrito y preparar una lista comparando entre las dos listas, desde el punto de vista de usuarios de la lista *Sears*. No hubo inconveniente y en unas semanas recibí una carta de sus abogados dándome el derecho de usar la



lista de encabezamientos que resultó de la unión y la que resultó de la comparación, para publicarla como se pudiera. Felizmente empecé a redactar la lista, y a buscar una casa editorial que la publicara. Primero se la ofrecí a H.W. Wilson, la editorial de *Sears*, pero lo rechazaron porque no les interesaba el tipo de cambio que sugiere la lista. La Biblioteca del Congreso no publica ningún manual aparte de los que ellos mismos producen, así que me decidí a llamar a las editoriales de bibliotecología. *Libraries Unlimited* lo aceptó sin vacilar y me ayudó en cuestiones del formato, que resultó ser muy parecido al de las listas de la LC. Cuando llamé a la bibliotecaria principal del distrito Eanes para informarle del plan, me dijo que en su última reunión las bibliotecarias habían comentado que lo que les hacía falta a ellas era un sistema de referencias, y querían que lo agregara en la lista final de ser posible. Esto resultó en una lista aún más parecida a la de la LC, que usa un sistema de referencias que se parece al estilo de un tesaurus; es decir, con siglas que representan términos no usados, equivalentes o relacionados, más amplios, y con notas para el usuario. El libro salió en 1993 bajo el título de *Headings for Children's Materials: An LCSH/Sears Companion*.

Resultó ser un buen producto para la compañía y cuando parecía que se agotaría la primera impresión ofrecieron un contrato para una segunda edición. La realidad es que no hay muchos cambios en la lista de AC de un año al otro, ni en los términos más usados para catalogación en bibliotecas generales. Quería que una nueva edición agregara más nombres propios, ya que las bibliotecas escolares y públicas —con excepción de las más grandes— no tenían acceso a los archivos autorizados, y un mismo catálogo estaba usando varias formas del nombre de las mismas personas. Mientras pensaba cómo seleccionar cuáles nombres deberían estar en la lista, recibí una invitación por parte de la biblioteca pública de Wichita Falls (Texas) para organizar un taller para catalogadores en la región. Me dijeron que aunque el título de la primera lista no mencionaba a bibliotecas públicas, ellos habían comenzado a usarla para uso diario porque es más pequeña, menos costosa que el libro completo de cuatro volúmenes, y así cada persona podía tener una copia en su propio escritorio, pero también porque la mayoría de los encabezamientos que necesitaban

a diario se encontraban en la lista. Es más, me ofrecieron una copia impresa de la lista de encabezamientos que ellos usaban, que incluía nombres propios para usar en el desarrollo de la segunda edición. Así surgió la ampliación del contenido de la lista y el cambio de título a *Subject Headings for School and Public Libraries: An LCSH/Sears Companion* (SHSPL) cuando se publicó en 1996.

Para la tercera edición (2001) se agregaron referencias bajo nombres propios —especialmente nombres personales— puesto que normalmente se requiere un encabezamiento de campo además del encabezamiento con el nombre de personas biografiadas, que constituyen la mayoría de los nombres en la lista. Para ayudar en la creación de registros en formato MARC, las etiquetas no obvias se agregaron a encabezamientos de nombres propios de todo tipo: nombres de individuos, grupos, personajes, barcos, edificios, eventos, etcétera.

Durante el desarrollo de cada edición me puse en comunicación con Forest Press, la organización que mantiene el sistema decimal de Dewey, para pedirles permiso para agregarles clasificaciones cortas a los encabezamientos. No me dieron permiso porque la organización OCLC, que estaba pagando los gastos de mantenimiento, estaba preparando otra lista de encabezamientos, basada en registros del catálogo unión que tienen un código de audiencia juvenil: *Subject Headings for Children*. Los editores del sistema Dewey agregarían esos números a dicha compilación automática para producir una nueva lista para usarse en bibliotecas juveniles. Esa valiosa lista, de unos 20,000 encabezamientos para títulos juveniles, editada por Lois Winkel, se publicó en 1994 y más tarde la segunda edición en 1998. No se ha redactado ninguna otra desde ese tiempo, ni existen planes para hacerlo. Por lo tanto, los editores del sistema Dewey están preparados para agregar las clasificaciones a la siguiente (cuarta) edición de *SHSPL*. Este cambio ayudará a usuarios de *SHSPL* que la usan en combinación con la lista *Sears* para términos más específicos en una variedad de categorías, *p. ej.*, nombres propios y de flores, industrias, lugares geográficos, etcétera.

La idea de traducir alguna lista menos comprensiva que *Bilindex*, que se basa en la lista completa de la LC, comenzó hace unos años. Las casas editoriales estadounidenses no se atrevían a vender

tal lista, en parte por el costo de entrar a lo que percibían como un mercado nuevo —el de la América Latina— pero también por razones lingüísticas, políticas, y sociales. Por lo tanto nadie desarrolló este tipo de lista. Lo único que había —especialmente en los Estados Unidos— era la traducción y adaptación de Carmen Rovira al español de 1984 de la lista *Sears*, a la que le faltaban términos recientes y muchos de los términos específicos que requieren los catálogos. La mayoría de los usuarios de esta lista en los Estados Unidos saben algunas palabras en español pero no son hispanoparlantes, y por lo tanto necesitan una lista más comprensiva. Aún así, la compañía H.W. Wilson estaba preparada para considerar una nueva traducción de esa lista, (lo cual fue mi plan original) trabajando con un traductor no bibliotecario en California. Pero después de un tiempo ya no quiso hacer el trabajo por ser demasiado detallado, y lo comencé a hacer yo misma aunque mi español no llega al nivel que se necesita para ese tipo de trabajo. Por supuesto, ese plan falló, y ahora el proyecto *Sears* lo está llevando a cabo uno de nuestros excelentes colegas.

Cuando comencé mi empleo como profesora de bibliotecología con la Universidad Estatal de Sam Houston, se nos ofreció a los nuevos profesores la oportunidad de solicitar una subvención interna para hacer alguna encuesta u obra académica que tendría el beneficio de ayudar a algunos profesores, ser publicado y darle más reconocimiento público a esta universidad estatal que está creciendo enormemente. Propuse hacer una traducción de la lista *SHSPL*, de la cual estaba por preparar la cuarta edición. Contrataría con dos traductores no estadounidenses para hacer traducciones paralelas de la lista, con el doble propósito de obtener lenguaje de dos países de habla española sin norteamericanización, y de buscar algún consenso para los encabezamientos en un lenguaje que se entendería en la mayoría de casos, aunque no fueran los términos preferidos en algunos lugares. Finalmente contrataría un editor bilingüe profesional para resolver conflictos que seguramente se encontrarían en las dos traducciones. Con el apoyo de bibliotecarios escolares y públicos, y la promesa de que buscaría una casa editorial que lo publicara, obtuve una subvención de dos años para completar el proyecto. Estamos en el segundo año.

Como primer paso me puse en comunicación con varios bibliotecarios bilingües para encontrar traductores apropiados. Aunque buscaba dos personas, resultó una sola traducción, que llevó a cabo un equipo de Buenos Aires dirigido por la doctora en bibliotecología Elsa Barber. Ellos comenzaron a partir de una copia en papel de la tercera edición en inglés, que incluye no sólo términos usados sino también algunos no usados, y una selección de encabezamientos relacionados, o más amplios o estrechos, además de algunas notas para el usuario. También usaron una copia electrónica del archivo que se empleó en la imprenta para no tener que copiar el texto original. Decidimos simplemente agregar la traducción al español después de cada elemento en inglés, precedido por la marca de equivalencia (=) acostumbrada en catálogos para indicar paralelismo en dos idiomas.

Como la traducción debería ser de la cuarta edición, que todavía no existía, se contrataron dos estudiantes de la universidad Sam Houston para verificar cada término en la tercera edición, comparándolas con entradas en el sitio Web de las fichas de nombres y temas que mantiene la LC con la ayuda del proyecto internacional NACO (Name Authority COoperative). Usando copias en papel del libro como base, las estudiantes crearon una ficha con la forma nueva de cada entrada que había cambiado después del año 2000 —la última vez que se habían verificado. Durante esos años la LC había cambiado su política en cuanto al uso de fechas finales en entradas de nombres de personas individuales, con el resultado de que hay un gran número de cambios que se tienen que hacer, no sólo en esta nueva edición de *SHSPL* sino también en miles de registros bibliográficos MARC. El impacto de este cambio todavía no se ha sentido en muchísimas bibliotecas porque hay muchos catálogos sin sistemas automáticos de control de autoridades y grandes cambios manuales de este tipo son infrecuentes. En las miles de bibliotecas que usan herramientas condensadas como *Sears* y *SHSPL*, también puede ser que este cambio de política y de entradas en CIP y registros MARC no se conozca bien, y que lo irán reconociendo con las nuevas ediciones que adquieran.

Además de los cambios en encabezamientos para nombres personales, hay relativamente pocos cambios en los términos usados en la tercera edición. Por lo tanto, otros tipos de cambios que se le manda-

ron al equipo traductor también fueron pocos. Sin embargo ellos sugirieron referencias y términos adicionales para una edición en español para el público general, como por ejemplo nombres de individuos, teatros, guerras, etcétera. Por el momento se van agregando referencias sugeridas, pero no muchos términos nuevos, lo que se hará hasta una última fase u otra edición, por cuestiones de tiempo y fondos.

El equipo traductor mandó archivos individuales de borradores de cada letra del alfabeto inglés, comenzando en junio de 2006 y terminando en junio de 2007; un plazo muy corto para unas 25,000 entradas con referencias. Hay centenares de entradas que comienzan con la letra A en inglés; uno se siente al terminar cualquier trabajo relacionado, que ha terminado un proyecto muy grande. Y luego siguen las otras letras, algunas cortas y otras más largas, pero ninguna como la primera. Durante este año la traductora principal mandó electrónicamente los archivos basados sólo en el libro durante el año, y finalmente la traducción de términos cambiados de cada letra, se añadirían en Texas.

Para obtener un segundo borrador los cambios traducidos por separado al final se van haciendo poco a poco en el primer borrador. Las referencias se van verificando o agregando según las necesidades de la nueva edición, y se van haciendo las correcciones a los términos y el formato, lo que llevará esta nueva versión a un nivel más cercano a lo que requiere la casa editorial Libraries Unlimited.

Todavía se está reexaminando la lista *Sears* para identificar cambios en las referencias que se requieren para el *SHSPL*. El trabajo continúa actualmente, y cualesquier cambios que hayan se agregarán al segundo borrador, o quizás sea el tercero, basado en lo que se encuentre en este trabajo. El índice de *Abridged Dewey Decimal Classification and Relative Index, Edition 14* también debe compararse con el propósito de agregarle referencias. Esa tarea quizás se deje para otra edición, ya que esté hecho el primer borrador de la traducción, si es que éste tiene éxito.

El director de Libraries Unlimited quiere publicar la cuarta edición de la versión en inglés, pero tardó muchos meses en hacerse la versión en español. Durante el verano pasado se encontró con un grupo al que le mencionó la idea, y éste lo convenció de que sería buena

idea. Ya tenemos un contrato para la versión en inglés, pero una versión en español todavía no está contratada. Mientras tanto, tienen los resultados del cuestionario que se describe aquí.

Tengo muchos estudiantes que viven cerca de la frontera de los Estados Unidos con México. Alguna vez les mencioné el proyecto y les gustó mucho la idea de una versión bilingüe. En realidad nunca lo había yo pensado, pero los estudiantes creían que sería un producto que les ayudaría, especialmente porque algunos saben hablar español pero no lo han estudiado y no conocen las reglas gramaticales del idioma.

Para averiguar si tal idea sería de interés para otros bibliotecarios, desarrollé una pequeña encuesta en Survey Monkey y dejé un mensaje en dos grupos de discusión electrónicos: AUTOCAT (grupo internacional de catalogadores) y TLC (grupo tejano de bibliotecarios escolares). La encuesta pidió saber el país de la persona que había llenado el cuestionario y el tipo de biblioteca en la que trabajaba. De las 88 personas que respondieron, 86 fueron de los Estados Unidos (97.7%), y dos de otros países no identificados. La mayoría, el 47.7%, eran bibliotecarios escolares; el 26.1% de bibliotecas públicas, y los demás variaban en el tipo de biblioteca que representaban, o eran profesores o directores que personalmente no usaban herramientas para catalogar. No se preguntó nada sobre qué se podría usar para combinar estos sencillos elementos de identificación.

El cuestionario planteó cinco preguntas [véase el Apéndice]. La primera buscaba saber si la persona usaría una lista de encabezamientos basada en *LCSH* o *LCSH/AC* en un sólo volumen y solamente en español. El grupo más grande de 38.6% respondió que no, mientras que el 25 % dijeron que quizás, el 20.5% que sí, y el 12% respondieron que lo usarían a veces. Aproximadamente el 58% dijeron que no, o que probablemente no comprarían tal libro si costara \$100 (USD) o menos, que fue la segunda pregunta. La tercera pregunta, quiso saber si usarían una versión bilingüe (español/inglés=) de ese mismo libro, y resultó muy diferente. Cuarenta y tres, es decir el 48.9% respondieron que sí, aproximadamente el 35% que a veces o que quizás, y el 16% que no. En la cuarta pregunta, cuarenta y ocho personas, o el 55% dijeron que sí, o que probablemente lo comprarían; el 20% que

posiblemente, y aproximadamente el 20% respondieron que no, o probablemente no. Para confirmar sus preferencias, la última pregunta buscó saber si usarían o comprarían listas separadas en español e inglés si ambas estuvieran disponibles. Cincuenta, o el 59% dijo que no, o probablemente no comprarían las dos herramientas; dieciseis (el 19%) que sí o probablemente, y 14 (el 16%) estaban indecisos.

Durante el proceso de traducir se tomaron algunas decisiones, siempre basadas en dos principios principales: uno, usar términos más o menos comunes, en paralelo con la versión en inglés, que sustituye los términos seleccionados para niños y jóvenes; y dos, usar términos que comiencen con la palabra más usada, que sería la primera en la entrada, cuando esto fuera posible.

Una de las decisiones pasó por dos fases: el término “americano” tiene una variedad de usos, y es lingüísticamente parecido a “American”, que en la mayoría de los casos se usa para describir a una persona o cosa de los Estados Unidos de América. Ni en inglés ni en español quiere decir siempre eso, pero esto no se explica. Al principio se sustituyó “estadunidense”, para ser más correcto, pero en realidad ésta no es la palabra en que uno piensa ni busca primero. Ni la palabra “norteamericano” —que es un poco menos correcta— se entiende como solamente relacionada a los EE.UU, y en realidad se usa menos que “americano”. Se resolvió, entonces, cualificar la palabra “americano/a” con (EE.UU.), abriendo la posibilidad de agregar el nombre de otro país, en algunos casos. Esto satisface los dos principios descritos, y ojalá lleve al buscador al tema necesitado más pronto que a las alternativas.

Otra decisión se basa en la gran probabilidad de uso en registros récords de catálogos en los Estados Unidos u otros países donde se encuentran concentraciones de usuarios hispanohablantes. Lo que anticipamos es que se agreguen encabezamientos en español. Para hacer esto en forma paralela, algunos encabezamientos que normalmente no se encontrarían en catálogos latinoamericanos o españoles quedaron en la lista. Sería maravilloso agregar a la lista otras entradas de más interés en más de los países hispanos, pero eso requiere más que la simple traducción y son tareas que no forman parte del trabajo para el cual se dio esta primera subvención. Para esta nueva edición,

*Primer Simposio Internacional sobre Organización...*

quedan en la lista bastante entradas que no serán de gran uso en muchos catálogos fuera de los Estados Unidos, aun para los de habla inglesa. Una ventaja de esta decisión se verá en la catalogación de libros y otras obras traducidas del inglés. Lo contrario también será facilitado para bibliotecas en países angloparlantes que usan la herramientas *Bilindex* y *Sears* para buscar equivalentes en español de términos que ya se encuentran en los registros de sus catálogos. Ninguna de estas listas incluye los términos excepcionales de la lista “juvenil” LCSH/AC, y esta traducción promete ayudarles en el expresado objetivo de darle mejor servicio a los usuarios de habla hispana.

A esta compilación se le han agregado las etiquetas MARC y más referencias. Más que nada, se espera que el uso de términos comunes tendrá el efecto de simplificar la tarea de catalogación, especialmente para el bibliotecario principiante.

APÉNDICE

**Date: Sun, 1 Apr 2007 18:09:55 -0500**

**Subject: Spanish Subject Headings Survey**

I am considering compiling a Spanish-English bilingual (dual-language) list of subject headings based on /Subject Headings for School and Public Libraries/. This book, currently in its 3rd edition, is a subset of about 10,000 terms from LCSH and the LC/NACO Name Authority File that are widely used in public and K-12 libraries in the United States.

If you would like to express your opinions about this, please click on the following link to fill out a short, anonymous survey: <http://www.surveymonkey.com/MySurveys.asp?Rnd=0.1639003>

Your time and thoughts will be very much appreciated!

Joanna F. Fountain

Assistant Professor, Dept. of Library Science

Sam Houston State University

[jff001@shsu.edu](mailto:jff001@shsu.edu)



1. Where are you located?			
		Response Percent	Response Total
United States		97.7%	86
Other Country		2.3%	2
<b>Total Respondents</b>			<b>88</b>
(skipped this question)			0
2. In what type of library do you work?			
		Response Percent	Response Total
School (K-12)		47.7%	42
Public (All levels)		26.1%	23
Academic (Post-secondary and above)		10.2%	9
Special (Public or private organization)		3.4%	3
Church or Seminary		1.1%	1
Other (please specify)		13.6%	12
<b>Total Respondents</b>			<b>88</b>
(skipped this question)			0

*Primer Simposio Internacional sobre Organización...*

3. Do you think that you would use a Spanish-only edition of a one-volume subject heading list based on LCSH or LCSH/AC? Note: This is an anonymous survey; no commitment is implied in your choice.

		Response Percent	Response Total
Yes		20.5%	18
Sometimes		12.5%	11
Maybe		25%	22
<b>No</b>		<b>38.6%</b>	<b>34</b>
Other (please specify)		11.4%	10
<b>Total Respondents</b>			<b>88</b>
(skipped this question)			0

4. Do you think that you would purchase (or recommend for purchase) a Spanish-only edition of a one-volume subject heading list based on LCSH or LCSH/AC if it were available for less than \$100 (U.S. dollars)? Note: This is an anonymous survey; no commitment is implied in your choice.

		Response Percent	Response Total
Yes		19.3%	17
Probably		3.4%	3
Possibly		26.1%	23
Probably not		14.8%	13
<b>No</b>		<b>33%</b>	<b>29</b>
Other (please specify)		8%	7
<b>Total Respondents</b>			<b>88</b>
(skipped this question)			0

5. Do you think that you would you use a bilingual (dual-language) English/Spanish edition of a one-volume subject heading list based on LCSH or LCSHAC? Note: This is an anonymous survey; no commitment is implied in your choice.

		Response Percent	Response Total
Other (please specify)		6.8%	6
<b>Total Respondents</b>			<b>88</b>
(skipped this question)			0

6. Do you think that you would you purchase (or recommend for purchase) a bilingual (dual-language) English/Spanish edition of a one-volume subject heading list based on LCSH or LCSHAC if it were available for less than \$100 (U.S. dollars)? Note: This is an anonymous survey; no commitment is implied in your choice.

		Response Percent	Response Total
Yes		40.9%	36
Probably		13.6%	12
Possibly		20.5%	18
Probably not		5.7%	5
No		14.8%	13
Other (please specify)		6.8%	6
<b>Total Respondents</b>			<b>88</b>
(skipped this question)			0

7. Would you prefer to use/purchase separate Spanish and English subject heading lists if both were available? Note: This is an anonymous survey; no commitment is implied in your choice.	
<b>Total Respondents</b>	<b>86</b>
(skipped this question)	2

8. Please make any comment* you wish. If you would like to ask a question and get a reply, but still protect your anonymity, please email Dr. Fountain at: jff001@shsu.edu.	
<b>Total Respondents</b>	<b>34</b>
(skipped this question)	54

**COMMENTS:**

1) I do purchase a significant quantity of Spanish books which I attempt to catalog in Spanish with the assistance of a long-suffering business manager who is fluent and of course a bi-lingual dictionary.

I did purchase a CD several years ago that was somewhat helpful with subject headings.

Sounds like a great idea for those of us that have bilingual students.

2) Library Resources, XX School District

We have two Spanish immersion elementary schools and I deal with nearly 109 languages groups in our library index.

We are in the business of teaching English people a foreign language in the immersion programs: Spanish, Japanese, Russian, and German.

Under other circumstances we are busy teaching 109 language group people to assimilate English speaking skills.

I have established that all MARC fields are in English: subject headings, annotations, etc.

Table of contents, if used, is in the language of the book.

3) I'd have LOVED a dual language subject heading book!—especially for under \$100.

4) If someone (like me) doesn't anticipate using such a tool but recognizes its value for others, we need a better response option for the use question.

5) What a wonderful and useful project! Many smaller libraries that cannot afford a subscription to *Bilindex* will find the new edition of *Subject Headings for School and Public Libraries* a worthwhile investment.

The use of Spanish and English subject headings represents the philosophical point of view for the library's public services. Over the last couple of years I conducted several research projects related to the topic you are proposing. Although the idea of bilingual access appeals to many librarians, the number of tools that they have to use during the cataloging and/or bibliographic enhancement process is severely limited. And the individual bibliographic records in OCLC do not provide consistent guidance. Your contribution would be one of the only tools affordable for smaller libraries. Go for it!

6) I would prefer a bilingual edition. My reasoning is that such a list would be useful to both

Spanish speakers and English speakers. If you want to target medium and small libraries, they probably do not have the resources to hire a Spanish-speaking cataloger - they are using what they have. Many Spanish records in OCLC have the LCSH assigned. Many more are translations of English language items where the appropriate LC-SH are easily found/determined. With a bilingual edition, the English speaking cataloger simply needs to find the appropriate English heading and accept the correlated Spanish heading.

I catalog the Spanish language items at X Public Library using *Bilindex* (bidex) and when necessary the *Lista de Encabezamientos de Materia para Bibliotecas* (lomb). The chief improvements I would recommend for both resources are an upgrade to make them fully bilingual, close attention paid to including all diacritics (because En-

glish speakers will not know when/where to insert them), and greater comprehensiveness.

Great project, I'm excited to see the result!

7) Yes! A bilingual list would be much more useful. While a separate Spanish edition would be useful to fluent Spanish users, until that day when all of us librarians and technicians are all nearly as fluent in Spanish as native speakers there will be a lot of us having to do cataloging or reference in Spanish who are lucky to find our way to the bathroom (even if we were fluent in French, Russian, and Chinese; and running on and on in that dark stormy night a bilingual edition would help us.

## BIBLIOGRAFÍA

*Abridged Dewey Decimal Classification and Relative Index*, Edition 14, Dublin, Ohio, OCLC, 2004.

“Annotated Card Program”, *Library of Congress Subject Headings, volume 1*, Subject Cataloging Division, Processing Department, Washington, D.C., Library of Congress, 1975 [-fecha].

*Bilindex: a bilingual Spanish-English subject heading list: Spanish equivalents to the Library of Congress subject headings = Una lista bilingüe en español e inglés de encabezamientos de materia : equivalentes en español de los encabezamientos de la Biblioteca del Congreso de Estados Unidos de Norteamérica*, Berkeley, Calif., Floricanto Press, 1983-1986.

*Library of Congress Subject Headings (LCSH)*, Subject Cataloging Division, Processing Department, Washington, D.C., Library of Congress, 1975 [-fecha].

MARC 21 concise formats, Network Development and MARC Standards Office, Library of Congress, Washington: Library of Congress, 2006.

Sears, Minnie Earl, 1873-1933, *Sears list of subject headings*. 18th ed. / ed. Joseph Miller; Joan Goodsell, New York, H. W. Wilson, 2004.

Sears: lista de encabezamientos de materia: traducción y adaptación de la 12a ed. en inglés, editada por Barbara M. Westby, y Carmen Rovira, New York, H. W. Wilson, 1984.

Subject headings for children: a list of subject headings used by the Library of Congress with Dewey numbers added / ed. Lois Winkel, Albany, N.Y., Forest Press, 1994.

Subject headings for children : a list of subject headings used by the Library of Congress with abridged Dewey numbers added / ed Lois Winkel, 2nd ed. Albany, N.Y., Forest Press, 1998.

Fountain, Joanna F., *Headings for children's materials: an LCSH/Sears companion*, Englewood, Colo., Libraries Unlimited, 1993.

Subject headings for school and public libraries: an LCSH/Sears companion, 2nd ed., Englewood, Colo., Libraries Unlimited, 1996.

Subject headings for school and public libraries: an LCSH/Sears companion, 3rd ed. Englewood, Colo., Libraries Unlimited, 2001.





SECCIÓN IV  
Semántica en la información



## La Biblioteca Semántica

JUAN VOUTSSÁS MÁRQUEZ

*Centro Universitario de Investigaciones Bibliotecológicas,  
UNAM, México*

Las bibliotecas digitales están conformadas por tres elementos básicos: las colecciones —*obviamente digitales*—, los servicios —*por supuesto en red*— y los usuarios —*naturalmente diseminados a lo largo del ciberespacio*. De acuerdo con lo anterior, una biblioteca digital les ofrece a sus usuarios servicios en red a través de sus colecciones digitales, al margen de todo lo que pueda poseer en sus acervos tradicionales. Por lo mismo es fundamental saber desarrollar una colección digital adecuadamente. Para empezar, es necesario establecer que, al igual que en una biblioteca tradicional, no existe una única colección documental, sino que en realidad está formada por una variedad de colecciones parciales que engloban distintos tipos de materiales y formatos los cuales conformarán lo que la biblioteca y sus usuarios conocen como “la colección” en términos generales. De esta forma, al referirme a la “colección digital” de una biblioteca estoy diciendo que el conjunto de ellas conforman esa colección. Por lo mismo y, como en toda colección bibliográfica que se respete, debe establecerse un objetivo y una política de desarrollo para la colección —o como hemos explicado— colecciones; es decir, debe haber una *selección*. Ésta es una condición imperativa y que sin embargo es soslayada frecuentemente procediéndose a crear

colecciones digitales de manera directa, sin ninguna planeación. La importancia de esto es tan significativa que sin duda se vuelve el primer principio de la preservación digital.

Al igual que en una biblioteca tradicional, no puede desarrollarse adecuadamente una colección en el entorno digital si no se ha establecido previa y claramente cuál es el objetivo que se persigue al crear esa colección, a quién va dirigida, cuál es su nivel, alcance, cobertura, política de adquisición, permanencia y descarte de materiales, etcétera. Aquí es donde la capacitación y la experiencia del bibliotecario profesional es capitalizada por ser sumamente útil en el momento de establecer esa política de manera adecuada lo que, por supuesto, debe quedar por escrito en un documento al efecto como guía de largo alcance para la biblioteca. Es de suma importancia que en estos objetivos y esta selección, distingan cuáles partes de la colección son para distribuir, cuáles son para preservar y cuáles para ambas cosas.

El deseo de preservar proviene, como ya hemos establecido, de la voluntad de heredar los documentos a futuras generaciones. Según sea la fragilidad del soporte en el que se encuentre el documento, el número de copias existentes, su edad y su estado de conservación, decidiremos que tal documento puede estar en riesgo de destrucción inminente, y deberemos entonces preocuparnos de “fijarlo” en otro soporte para evitar, si no la destrucción del documento-soporte como se conoce, cuando menos sí la de su contenido. Tal sería, por ejemplo, el caso de periódicos de finales del siglo XIX o principios del siglo XX. En otros casos la preservación puede darse evitando su manipulación aunque su estado de conservación sea satisfactorio, para alargar la vida del soporte actual; por ejemplo el caso de un incunable: por lo general, si éste está más o menos conservado, su edad no lo pone en un riesgo inminente de destrucción, pero sabemos que cada vez que ese libro se abre y cierra y que sus páginas son pasadas, lo sometemos a un desgaste físico que si fuese frecuente, en algunas décadas, el esfuerzo mecánico, la exposición a la luz, etcétera, lo deteriorarían inexorable y gravemente, cosa que, por supuesto, no sería deseable. Ésta es la primera decisión de fondo que el bibliotecario debe tomar al momento de pensar en crear y operar colecciones digitales; digitalizar una colección es una decisión que depende de muchas cosas.

Las bibliotecas que sólo crean colecciones para distribuir carecen de responsabilidades de preservación a largo plazo, tienen una tarea un poco más sencilla, ya que el material efímero de esas colecciones no implica tanto trabajo adicional. Pero cuando el objetivo es la preservación, la responsabilidad crece proporcionalmente con el número de variables a ser consideradas. Estas variables obedecen a una serie de principios perfectamente establecidos en la actualidad y que deben ponderarse antes de desarrollar la colección para ver el grado en que debe integrarse cada principio a cada colección de la biblioteca.

Así, aunque los elementos de los documentos digitales están identificados no están agrupados de la misma forma. Por lo mismo, he tratado de establecer una resultante de lo que todo documento que se pretenda preservar debe cumplir y la he agrupado en mi versión de cinco principios generales:

- Selección
- Calidad
- Permanencia
- Accesibilidad
- Funcionalidad

El principio de la calidad es directamente proporcional al grado en que el documento digital emula al documento original o representado, y le otorga por tanto un valor de uso equivalente al de ese documento. Las características que conforman la calidad incluyen conceptos como la apariencia respecto de un original; es decir, ¿qué tanto se parece el documento digital a lo que pretende representar? Esta apariencia a su vez está definida por elementos tales como la resolución, o grado de minuciosidad que tienen los elementos que conforman el documento; tono, color e incluso textura; su escala de —preferencia uno a uno—, su secuencia original, su integridad, un identificador único, etcétera.

La integridad es sumamente importante. Algunos autores la consideran incluso como uno de los elementos básicos de la preservación, y para el mundo de la archivística es indispensable y fundamental. Este concepto se refiere al grado en que el documento digital refleja al

original, no tanto en su apariencia física, sino en su esencia. Un documento íntegro es el que refleja totalmente la esencia del original; es decir, no ha sido corrompido en su contexto: alterado, mutilado, interpretado, aumentado, recortado, deformado, etcétera; es confiable y por tanto aceptable. Su mensaje, autoría, fechas asociadas, lugares, etcétera, son en realidad las consignadas en el documento; es auténtico. En esencia, refleja de manera completa lo que el autor original estableció. Un objeto digital debe poder ser autenticable en al menos tres aspectos. Primero, el usuario debe poder determinar que el objeto es lo que pretende ser. Segundo, el usuario debe poder ser capaz de determinar el origen, estructura, versión, y porción del objeto. Y tercero, el usuario debe poder determinar que el objeto no está corrompido o alterado de forma intencional o accidental.

Los objetos digitales de calidad están nombrados con un identificador único y permanente de acuerdo con un esquema bien documentado. Nunca de acuerdo con una referencia absoluta a su nombre de archivo o dirección —como el caso de su URL— ya que éstos tienden a cambiar con el tiempo. En lugar de ello, un identificador estable debe poder ser resuelto o mapeado siempre a su dirección actual.

El principio de la *permanencia*, tiene que ver con el concepto de que el documento estará disponible durante un lapso considerable. Está asociado con su presencia, su seguridad y, por supuesto, con la duración de su soporte. Muchas personas tienden a confundirlo con el principio de la accesibilidad, el cual tiene que ver con que el documento, existiendo, pueda ser accesible. Son dos conceptos distintos; trataré de explicar brevemente esta diferencia.

Si tenemos en nuestras manos un disco fonográfico de acetato de 78 r.p.m., un cassette con una película “Beta”, una microficha de 4 X 6 pulgadas, una película en 8 mm., o un archivo en *Lotus 1-2-3* en un disquete de ocho pulgadas; un archivo de *Word 5.1* en una cinta magnética de computadora de media pulgada, 7 o 9 canales, y todos estos documentos se encuentran en perfecto estado de conservación, tenemos entonces un conjunto de documentos que han permanecido a lo largo del tiempo: existen sin duda. Han tenido “permanencia”. Pero una cosa es que existan, que hayan permanecido en buen estado, y otra es que podamos acceder a ellos. Para cada uno de ellos reque-

rimos de un accesorio tecnológico del cual dependemos para lograr esa acción. En el caso del archivo *Lotus*, además de un lector de disquetes de ocho pulgadas conectado a una computadora, necesitamos del programa *Lotus* o alguno que pueda acceder a ese archivo; doble requerimiento. Lo mismo sucede con el archivo en *Word* en una cinta magnética: necesitamos el dispositivo lector de la cinta y el programa que pueda abrir e interpretar el archivo, sin mencionar el sistema operativo, que para ese procesador de texto, era MS-DOS. Dependiendo de nuestra capacidad de disponer de esos artefactos, tendremos acceso a esos documentos. Habrá o no “accesibilidad”, independientemente de su permanencia.

Hoy en día ya se maneja el concepto de *archivo permanente*, el cual consiste en una serie de estrategias y técnicas tendientes a lograr que la arquitectura de los sistemas de archivos de información digital para preservación sea independiente de la tecnología usada para crearlos, precisamente para reducir este problema de la accesibilidad.

El concepto de *interoperabilidad* significa que el documento será accesible a lo largo de varias plataformas y programas de computadora. Esto es, el documento no debe estar atado a formatos específicos de un proveedor de *software*, a una marca de computadora o de cámara fotográfica, etcétera. Un ejemplo muy ilustrativo lo vemos cuando intentamos abrir un documento en una página Web y nos aparece el mensaje “este documento sólo puede verse en *explorer* versión tal...”. Ese documento no es interoperable. Un documento en verdadero formato html es visible en prácticamente cualquier navegador: *explorer*, *netscape*, *mozilla*, *firefox*, *ópera*, etcétera; incluso a lo largo de varias versiones de los mismos. El secreto de la interoperabilidad es preferir siempre los *estándares abiertos* y evitar los *estándares propietarios*, ya que a veces, en aras de una cierta comodidad de creación, nos atamos a una plataforma de equipo o de programas.

El último principio mencionado en la preservación es el de la *funcionalidad*. Hablando formalmente, la funcionalidad son *las características interconstruidas en una interface de búsqueda que determinan la facilidad con que los usuarios pueden formular búsquedas y obtener resultados*.<sup>1</sup> Tratemos de explicar este concep-

---

1 ODLIS, Online Dictionary for Library and Information Science.

to: su trasfondo está asociado a la facilidad con que los usuarios puedan acceder a un documento y a sus documentos relacionados. Pero esa facilidad depende de las capacidades que puedan alcanzar las interfaces de búsqueda de documentos, las que a su vez dependen de las características de información que puedan obtener del interior de los propios documentos. En conjunto, la funcionalidad se mide por la suma de las capacidades que tengan los documentos, programas, sistemas, etcétera; si observamos que una parte muy importante de esta funcionalidad proviene del propio documento, podemos entender la importancia de que un documento digital tenga funcionalidad, pues de otra forma se afectaría negativamente al resto de los componentes informáticos que accederán a él.

De ahí la importancia que tienen los metadatos de un “objeto digital”. Sin metadatos el documento digital no podría ser accedido en forma alguna. No tendría funcionalidad. Algunos metadatos sencillos van incrementando paulatinamente esa funcionalidad y ésta es la medida en que el documento puede ser recuperado y asociado con otros. Buenos metadatos nos van aproximando a una buena funcionalidad, como el grupo de metadatos “típico” o “estándar”: metadatos descriptivos, técnicos, estructurales y legales. Obviamente deseamos acercarnos a una óptima funcionalidad para que el documento tenga óptima recuperación y acceso.

Pero para poder entender más a fondo este concepto de “óptima funcionalidad” es necesario explicar primero qué son la *Biblioteca Semántica* y la *Web Semántica*.

La *Biblioteca Semántica* es la tendencia evolutiva de las bibliotecas; las bibliotecas digitales se están convirtiendo poco a poco en *bibliotecas semánticas* y ése será su siguiente estadio. Este concepto se basa a su vez en el principio de la *Web Semántica*. En una explicación sencilla, la *Web Semántica* es una *Web* extendida, dotada de mayor significado en la que cualquier usuario de Internet podrá encontrar respuestas a sus preguntas de forma más rápida y sencilla gracias a una información mejor definida. Hasta ahora las páginas *Web* han estado hechas para ser leídas por seres humanos, no por máquinas. Al dotar a la *Web* de más semántica para sí misma, para sus programas y buscadores tendrá a su disposición más significado, y gracias a ello se-



rá posible obtener soluciones a problemas habituales en la búsqueda de información gracias a la utilización de una infraestructura común, mediante la cual, permitirá compartir, procesar y transferir información de forma sencilla. Esta *Web* extendida y basada en el significado no tan sólo humano sino computacional, se apoya en lenguajes universales que resuelven los problemas ocasionados por una *Web* carente de semántica en la que, en ocasiones, el acceso a la información se convierte en una tarea difícil y frustrante.

Una *biblioteca semántica* es una biblioteca digital cuyas colecciones están formadas por *documentos funcionales*. Un documento funcional, comienza con un objeto digital, como en una colección de biblioteca digital, documento que, como ya mencionamos, consta de una “reproducción digital fiel”, con sus características de calidad, permanencia, integridad, interoperabilidad, etcétera, pero no se queda en sólo algunos simples metadatos agregados para identificarlo. Un *documento funcional* es un objeto digital de calidad, semánticamente rico, el cual, además de los cuatro tipos de metadatos estándar y gracias a otros metadatos llamados de hipervínculo, ha sido definido y ligado de tal forma con otros documentos funcionales que lo han dotado de un óptimo reuso, búsqueda, integración, minería de datos e interoperabilidad, y lo han provisto también de una lista de sus contenidos que permiten las interfaces con otros programas externos que puedan explotarlos. No se trata de una inteligencia artificial mágica que les permita a las máquinas entender las palabras de los usuarios, es sólo la habilidad programada en una máquina para resolver problemas bien definidos, a través de operaciones bien definidas que se llevarán a cabo sobre datos existentes bien definidos.

Para obtener esa adecuada definición de los datos, la *Web Semántica* utiliza herramientas cuyas primeras versiones de prueba ya se han empezado a definir y utilizar, tales como el “marco de descripción de recursos” —*Resource Description Framework*— —*RDF*— y —*Web Ontology Language*— —*OWL*—, dos estándares que ayudan a convertir la *Web* en una infraestructura global con la que es posible compartir, y reutilizar datos y documentos entre diferentes tipos de usuarios. *RDF* proporciona información descriptiva simple sobre los recursos que se encuentran en la *Web* y que se utiliza, por ejemplo, en catálogos de libros, directorios, colecciones personales de música,

fotos, eventos, etcétera. OWL es una herramienta para desarrollar temas o vocabularios específicos que pueden asociar esos recursos. Lo que hace OWL es proporcionar un lenguaje que puede definir ontologías estructuradas para ser utilizadas a través de diferentes sistemas. Las ontologías son ni más ni menos que la definición moderna y para computadora del concepto de tesoro: incluyen las definiciones de conceptos básicos y subordinados en un campo determinado y la relación entre ellos, y se encargan de definir los términos utilizados para describir y representar un área de conocimiento. Estas ontologías son utilizadas por los usuarios, las bases de datos y las aplicaciones que necesitan compartir información específica; es decir, la de un campo determinado como puede ser el de las finanzas, medicina, deporte, bibliotecología, etc etcétera. Véase el diagrama de *Web Semántica en el Anexo*.

En resumen, la funcionalidad de un documento está dada por características inherentes de calidad aunadas a metadatos adecuados, sobre todo los de hipervínculo los cuales los relacionan con otros documentos semejantes —es decir, funcionales— de manera tal que el conjunto documental así relacionado *tenga un significado* procesable por computadora, y esto permita que los sistemas puedan darle a los documentos prelación, pertinencia, pertenencia, jerarquía, relación, etcétera, y lograr con ello una búsqueda inteligente, y por tanto una recuperación más precisa y útil para el usuario. Esto debe incrementar sensiblemente el reuso y hallazgo de esos documentos, la adecuada minería de datos y la integración, todo bajo un esquema de interoperabilidad entre sistemas.

En conclusión, dentro de la preservación digital, todo documento que se pretenda preservar debe cumplir lo más que sea posible con los cinco principios de los objetos digitales para ser preservados. En la medida en que cumpla con esto, ese documento será un buen objeto digital. Resumiendo, un **buen objeto digital deberá:**

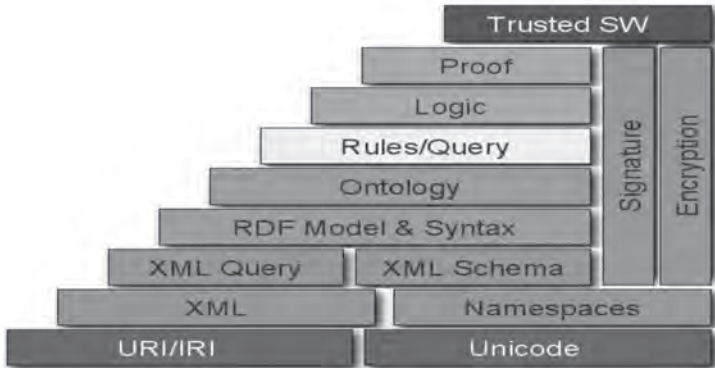
- 1) Provenir de una política específica de selección; y por tanto cumplir con las prioridades y objetivos de la colección.
- 2) Tener calidad intrínseca: apariencia adecuada, resolución, tono escala, secuencia, integridad, comprensión sin pérdida, identificador único, etcétera.

- 3) Ser permanente; esto es, debe existir la intención y responsabilidad de una persona o institución identificados, de que ese objeto permanecerá accesible a lo largo del tiempo. Y deberá estar representado en un formato digital que soporte el uso actual para el que fue diseñado así como usos futuros, y ser susceptible de ser copiado a otros formatos.
- 4) Tener accesibilidad; esto es, la institución que lo custodia tendrá acceso a las tecnologías que le permitan acceder a ese documento, o bien el objeto digital habrá sido modificado, actualizado, emulado, etcétera a nuevas tecnologías al alcance de la institución. Tendrá por tanto interoperabilidad —será intercambiable entre plataformas—, habrá sido hecho bajo algún estándar o buena práctica reconocidos y sólo se alejará de ellos debido a razones claras y bien documentadas.
- 5) Será funcional; esto es, tendrá una riqueza notable de metadatos, tanto descriptivos como técnicos, estructurales, legales y de hipervínculo, de tal forma que será semánticamente rico, y por tanto estará ligado de tal forma con otros documentos funcionales de manera tal que permitirá su óptimo reuso, búsqueda, integración, minería de datos e interoperabilidad y les dará a otros programas, sistemas e interfaces una lista completa de sus contenidos.

Como habrá podido observarse con estos principios, es sumamente importante que toda biblioteca digital establezca con precisión sus objetivos para desarrollar sus colecciones. Las bibliotecas deben crear buenos objetos digitales, pero el concepto de lo que significa un buen objeto digital depende en mucho de para qué se crean esas colecciones y sus objetos. Por tanto, la primera decisión a tomar, como ya hemos mencionado, es establecer si vamos a crear objetos digitales para preservarlos, para distribuirlos, o para hacer ambas cosas, ya que como hemos empezado a establecer, las características y los costos de cada uno de ellos son sumamente distintas. Nunca debemos emprender proyectos de preservación con especificaciones de distribución ni viceversa. Las consecuencias pueden ser lamentables.

## ANEXO

Diagrama de la *Web Semántica*.



**Unicode** es una codificación del texto que permite utilizar y visualizar en pantalla los caracteres de diferentes idiomas. Esto permite expresar información en la Web Semántica en cualquier idioma.

**URI** es el acrónimo de *Uniform Resource Identifier*, o en castellano, *Identificador Uniforme de Recursos*. Es un poco confusa la diferencia entre URI y URL, por lo que diremos que un URL es un recurso que puede ser accedido via Internet.

**XML + namespaces + esquema xml + XML Query**; la capa más técnica de la Web Semántica. En esta capa se agrupan las diferentes tecnologías que hacen posible que los agentes puedan entenderse entre sí.

**Reglas / Query** - Reglas que deben aplicarse durante la búsqueda para favorecer la recuperación del documento.

**Modelo RDF + sintaxis**; directamente basada y apoyada en la capa anterior, esta capa define el *lenguaje universal* con el cual podemos expresar diferentes ideas en la Web Semántica.

**Ontología ó Lenguaje de Ontologías** - nos permiten extender la funcionalidad de la Web Semántica, y agregar nuevas clases y propiedades para describir los recursos.

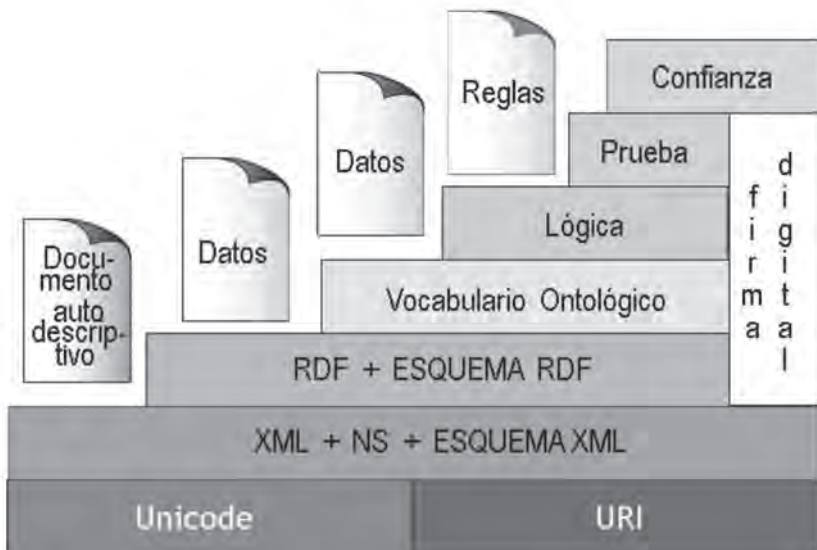
**Lógica** - Lógica asociada a las búsquedas o recuperaciones del documento.

**Pruebas** - pruebas aplicables al documento.

**Software de confianza** - software confiable asociado para explorar el documento.

**Firma Digital** - firma digital que puede agregarse al documento para autenticarlo si se desea.

**Encriptado** - llaves de encriptado del documento si se desea, esto.





# El trabajo terminológico en la Escuela Interamericana de Bibliotecología de la Universidad de Antioquia

MARÍA TERESA MÚNERA TORRES  
CARLOS ARTURO MONTOYA CORREA  
*Escuela Interamericana de Bibliotecología  
Universidad de Antioquia, Colombia*

## INTRODUCCIÓN

La ponencia que se presenta a continuación tiene como propósito mostrar un panorama de lo que ha sido la evolución del trabajo terminológico en la Escuela Interamericana de Bibliotecología en sus 50 años de existencia. Para comenzar se hace referencia a la etapa inicial de la Escuela y a la realización de documentos cuyo contenido se centraba en apuntar hacia un desarrollo de la terminología bibliotecológica de aquel entonces, justamente cuando los primeros docentes de la Escuela provenían de los Estados Unidos y de otros países del continente. Luego, en la segunda etapa del desarrollo de la Terminología en la Escuela, se da cuenta de ésta como disciplina estrechamente relacionada con el campo de la organización de la información, los lenguajes documentales, los lenguajes especializados, así como la introducción en la conceptualización de la Terminología, concebida desde tres perspectivas: La primera perspectiva, centrada en la disciplina que estudia los términos y sus relaciones semánticas dentro de un contexto determinado del conocimiento, la segunda está enfocada en el ejercicio de una actividad destinada a elaborar lenguajes controlados de carácter especializado, y la tercera relacionada

con los mismos productos terminológicos, tales como diccionarios, tesauros, léxicos, ontologías, entre otros. Además en esta segunda etapa se resaltaba la proyección que iniciaba la Escuela en conjunto con otras instituciones y universidades que se enfocaban también en el estudio terminológico a nivel local, nacional e internacional, para luego dar cuenta de la forma en la cual se abordaba dicho contenido dentro de la estructura curricular de la E.I.B., y culminar luego con la última etapa de consolidación de la Terminología en los primeros años del siglo XXI, época en la que se iniciaba la estructuración oficial del Grupo de Investigación EIBTERM.

## 1. ETAPA INICIAL DE LA TERMINOLOGÍA EN LA E.I.B.

Indiscutiblemente la Terminología ha estado presente en las diferentes etapas de desarrollo de la Escuela Interamericana de Bibliotecología desde la década de los años 50, cuando inicia su labor como formadora de bibliotecólogos y empieza a trabajar de manera especial en los procesos de organización de la información representados por las técnicas de catalogación y clasificación y el manejo de los lenguajes documentales, además de constituirse en un componente fundamental para la labor de los profesionales de la bibliotecología que se formaron en la naciente E.I.B., aspecto sobre el cual el profesor Uriel Lozano se refiere cuando precisa cómo; *desde el primer año de labores académicas —1958— se toma la iniciativa de incentivar el estudio de la terminología técnica. Los profesores Gaston Litton, Victoria Fisher de Duque, Emma Corsslan Simonson y Edith Schwarz, publican los primeros trabajos terminológicos en la historia de la bibliotecología del país. Ellos, compilan cuatro manuales, publicados por la Escuela, con el fin de propiciar que el alumno adquiriera el correcto uso del vocabulario técnico tanto en inglés como en alemán. Pues esta última lengua era considerada en su momento como la segunda en importancia bibliotecológica en el mundo. Los títulos de los manuales que fueron ampliamente usados y difundidos no sólo en Colombia sino en otros países de América Latina, fueron:*



1. *Terminología profesional, 1958. Que fue una lista de términos de uso frecuente en la literatura profesional en inglés.*

2. *Professionelle terminologie der bibliothekswissenschaft, 1960. Que abarcó siete categorías de términos en alemán relacionados con la bibliotecología, las bibliotecas y las librerías.*

3. *Anthology of readings in library science, 1961. Que en dos volúmenes y bajo veinte categorías temáticas sobre las bibliotecas y el mundo del libro pretendía enseñar no solo el uso de la terminología técnica sino el del mismo inglés.<sup>1</sup>*

Más adelante se introduce el tesoro como una herramienta de análisis para organizar el conocimiento especializado y sistematizar los términos con base en la definición de relaciones semánticas que sirvieran de fundamento para precisar los niveles de significación pertinentes, acordes con los contextos y orientaciones disciplinares. Y fue precisamente en los albores de la década de los años 70, cuando se empieza a desarrollar la teoría de los tesauros como un aporte al desarrollo de los lenguajes documentales, aspecto sobre el cual el profesor Uriel Lozano escribe:

*Una nueva corriente documental, que apenas despuntaba en Europa y en Estados Unidos, se hizo presente en el programa de estudios de la EIB. Nos referimos a la teoría de los nuevos lenguajes documentales, llamados tesauros, y a la indización como forma de comunicación de la ciencia, y a los estudios terminológicos, para que, de esta manera, los nuevos conceptos y vocablos, originados por la constante renovación de la ciencia y la tecnología, tuvieran su proceso de validación y le permitieran al estudioso uniformar y fijar los términos científicos.*

*Al tenor de lo que rezan los cánones curriculares, la Escuela envió a Europa a uno de sus docentes, el profesor Uriel Lozano,*

---

1 Uriel Lozano Rivera, *El estudio de la terminología en la escuela interamericana de bibliotecología: una aproximación a sus momentos históricos*, Medellín, UdeA, E.I.B, 2007, p. 1.

*para que se capacitara, bajo la tutela del profesor F.W. Lancaster, en aquellas nuevas herramientas terminológicas aplicadas a los servicios documentales. La EIB fue, en consecuencia, la primera institución en Colombia en enseñar la nueva teoría de Tesoros. Arrojando, en consecuencia, que sus egresados no sólo se prepararan en esta área de los lenguajes documentales, tan prioritaria para la organización del conocimiento, sino que además, serían los multiplicadores de la idea, pues iban a ser los docentes en otras escuelas de bibliotecología del país y del exterior, con el agregado de que muchos de ellos fueron partícipes y propiciadores de toda una generación de tesoros que se hicieron en aquella época en Colombia.<sup>2</sup>*

El mismo profesor Lozano, en representación de la Escuela, entre diciembre de 1973 y diciembre de 1983, fue presidente-fundador del Subcomité C.30.1 de Terminología de la Documentación del Instituto Colombiano de Normas Técnicas (ICONTEC), máximo organismo para la normalización del país.

Tanto la capacitación en herramientas terminológicas de sus egresados como la participación de sus profesores en los organismos nacionales de normalización de la Terminología, se pueden considerar como respuestas acertadas al nuevo contexto y a las políticas gubernamentales de ciencia y tecnología que se instauraron en Colombia en aquella década, lideradas por el recién creado Instituto Colombiano para el Desarrollo de la Ciencia (COLCIENCIAS), organismo que se esforzó por desarrollar el Sistema Nacional de Información (SNI), hoy conocido como el Sistema Nacional de Ciencia y Tecnología e Innovación.

Paralelamente a la enseñanza del tesoro en la Escuela Interamericana de Bibliotecología, se formaba a los nuevos profesionales en el manejo de las listas de Encabezamientos de Materia de Carmen Rovira y Jorge Aguayo, además de las de la Biblioteca del Congreso (Library Congress Subset Headings) y las de la Biblioteca Nacional de Medicina de los Estados Unidos, MESH. También, se hace necesario mencionar en este proceso de evolución de la terminología en la E.I.B., a algunos

---

2 Uriel Lozano Rivera, *Ibíd.* p. 2.

de los docentes que participaron desde la actividad lectiva en esta etapa aplicativa de los lenguajes documentales, a profesores como Luis Florén Lozano, Irma Isaza, Iván Rúa, Libia Lotero, Uriel Lozano y Olga Cecilia Velásquez, entre otros.

## 2. ETAPA DE IMPULSO DE LA TERMINOLOGÍA EN LA E.I.B.

Como se puede apreciar, en las primeras décadas de trabajo formativo de la Escuela Interamericana de Bibliotecología, la Terminología se manifestó de manera más aplicativa que conceptual y fue sólo hasta la década de los años 90, cuando empezó a perfilarse la teoría de la Terminología, como una disciplina que contenía no solamente el componente práctico, sino también el teórico, sobre el cual se había generado una serie de reflexiones y aportes desde lo investigativo y lo epistemológico, según lo manifiesta la profesora Bertha Nelly Cardona en una de sus intervenciones académicas : *Es en la Escuela Interamericana de Bibliotecología donde se han formado los primeros trabajadores de esta disciplina, ya que desde los lenguajes documentales llámense tesauros, sistemas de clasificación, glosarios, nomenclaturas y todo tipo de taxonomías, han abordado la Terminología para la solución de sus problemas prácticos relacionados con la indización de los documentos, su organización y posterior recuperación.*<sup>3</sup>

Durante los años noventa se impulsa el estudio de la Terminología, en la E.I.B. especialmente con la introducción de dicho contenido como parte de las asignaturas de organización del conocimiento, tales como descripción bibliográfica y análisis y representación de la información. Es importante resaltar a la profesora Bertha Nelly Cardona como impulsadora del estudio específico de la Terminología, no sólo en la Escuela Interamericana de Bibliotecología sino también en el país. Con su liderazgo, nació la Red Colombiana de Terminología COLTERM, sobre la cual el profesor Uriel Lozano, sostenía: *Se genera con la vinculación a la Escuela como Directora primero y luego*

---

3 Bertha Nelly Cardona R., Presentación Revista Interamericana de Bibliotecología, Medellín, Vol. 22, núm. 1 (ene-jun. 1999); p. 5.

*como profesora de la Licenciada Betha Nelly Cardona, una apasionada por la terminología, miembro de la Red Colombiana de Terminología (COLTERM), del Consejo Directivo de la Red Iberoamericana de Terminología (RITERM) y de la Unión Latina. Le cabe el honor de introducir los estudios de la terminología como asignatura independiente dentro del currículo, así como también de llevar la representación de la Escuela, presentar ponencias y artículos, y organizar numerosos eventos nacionales e internacionales sobre esta temática. Su paso por la Escuela dejó una semilla importante para el desarrollo de los estudios terminológicos.<sup>4</sup>*

Se hace entonces necesario destacar en esta fase del desarrollo terminológico de la E.I.B., el papel que jugó la profesora Bertha Nelly Cardona en la valoración de esta importante disciplina, el cual fue decisivo para su desarrollo, y el legado que ella dejó a sus alumnos y docentes, tanto de la Escuela Interamericana de Bibliotecología como de la Universidad de Antioquia y otras universidades de Colombia.

Adicionalmente y en consonancia con las afirmaciones anteriores es menester hacer una especial mención al trabajo académico que desarrolló la E.I.B. en torno a la Terminología en la década de los años 90, así como la proyección que esto provocó en los ámbitos locales, nacionales e internacionales.

Durante 1995 se llevaron a cabo diferentes eventos académicos en torno a la Terminología y bajo la coordinación de las Escuelas Interamericana de Bibliotecología y de Idiomas de la Universidad de Antioquia, entre los cuales destaca el curso taller sobre terminología, en el cual participaron los reconocidos expertos Heribert Picht y Gerard Budin y cuyo contenido central se basó en temas como la teoría, métodos y aplicaciones de la terminología, la comunicación profesional, los lenguajes profesionales, la Terminología y disciplinas afines, la investigación en lenguas profesionales, las bases teóricas de la terminología, la lexicografía del lenguaje profesional, la Terminología y los bancos de datos terminológicos, la Terminología, la Información y la Documentación. De igual manera se resalta la creación de la Red Colombiana de Terminología en mayo de ese mismo año.

---

4 Uriel Lozano Rivera, *op. cit.*, p.2

Luego, en 1996, se realizan además el I Seminario Internacional sobre pedagogía de la terminología, con la participación de Heribert Picht y Klaus-Dirk Schmitz, el I Seminario Internacional sobre lenguajes profesionales y terminología asistida por computador, el I Taller Colterm y la II Asamblea General de Colterm, y en 1997, la Escuela Interamericana de Bibliotecología participa en la I Escuela de Verano de Terminología, que se lleva a cabo en la Universidad Pompeu Fabra de Barcelona con la presencia de dos estudiantes y una bibliotecóloga del Departamento de Bibliotecas de la Universidad de Antioquia. También a finales de ese mismo año, la E.I.B. tiene la gran satisfacción de contar con la profesora María Teresa Cabré de la Universidad Pompeu Fabra de Barcelona, quien orientó el curso sobre fundamentos de terminología, y cuyo fin principal era iniciar a los participantes en los conceptos más elementales, pero cruciales de la Terminología, concebida como una materia interdisciplinaria, introducirlos en su metodología y mostrarles algunas técnicas de trabajo, además de contrastar aproximaciones teóricas diferentes a quienes ya contaban con formación previa en Terminología.<sup>5</sup>

Para 1998 se participa en el I Coloquio Nacional de didáctica de la Traducción y la Terminología, en la ciudad de Cali, así como en el Taller Colterm. También se crea el Comité Interinstitucional de Terminología, con la Universidad Pontificia Bolivariana, ESUMER y la Universidad de Antioquia, y se desarrolla la investigación “Tesoro Colombiano de Seguridad Social”, por parte de la Profesora Bertha Nelly Cardona R., en colaboración con el bibliotecólogo, Orlando Tobón B., el estudiante Milton Alexander Ayala E. y la abogada María Isabel Lopera V., quien sirvió como asesora del grupo dada su amplia experiencia en el campo laboral y en seguridad social en la misma universidad.

Se realizaron pasantías académicas, como la del estudiante de doctorado de Filología de la Universidad de Lérida, España, José Luis Orduña, quien estuvo en la Escuela Interamericana de Bibliotecología entre agosto y septiembre de 1998. En octubre del mismo año, se participó en el III Seminario Nacional de Terminología en Santa Marta,

---

5 María Teresa Cabré Castellví, *La Terminología. Marco teórico, metodología, consecuencias prácticas*. [Curso-Taller], Medellín, Universidad de Antioquia, E.I.B., E.I., 1997.

así como en el VI Simposio Iberoamericano de Terminología en la Habana, Cuba.

Es importante también traer a colación otra manifestación de la consolidación de la Terminología como componente fundamental de la formación de los profesionales de la Ciencia de la Información en la E.I.B., y es la edición, el primer semestre de 1999, de la *Revista Interamericana de Bibliotecología*, con un número especial dedicado exclusivamente a tratar temas terminológicos y relacionados. Es así como en dicho número se destacan contenidos tales como: La relación entre objeto, concepto y símbolo por parte de Helmut Felber y con la traducción de John Jairo Giraldo. También se abordan los tesauros conceptuales como herramientas de precisión en los sistemas de organización científica, por parte de Miguel Angel López, y esto es seguido por el artículo titulado La Terminología en algunos diccionarios de bibliotecología, a cargo de José Luis Orduña. Entre otros de los tópicos que se tratan en este número, está el relacionado con la terminología de la seguridad social colombiana para construir un tesoro y glosario especializado, fruto de la investigación llevada a cabo por parte de la profesora Bertha Nelly Cardona y un equipo interdisciplinario.

También hacen parte del contenido de la revista "El resumen como habilidad esencial para el trabajo académico" con la autoría de los profesores Luz Stella Castañeda y José Ignacio Henao. Se presentan además estudios técnicos en los que se aborda la catalogación descriptiva: aspectos que intervienen en la calidad procesual de Antonio Bereijo. Igualmente la profesora Bertha Nelly Cardona hace un recuento histórico de lo que es La Red Colombiana de Terminología, COLTERM, y el profesor Guillermo Correa describe un modelo para la evaluación integral de tesauros.

En términos generales, la Escuela tiene presencia en diversos eventos académicos, cuyo contenido se basa en la Terminología como objeto de estudio y como complemento fundamental de la formación de los bibliotecólogos y documentalistas. Es así como se participa en El II Coloquio Nacional sobre didáctica de la traducción y la terminología, el Curso-taller Colterm para capacitación interna de los grupos de investigación en Terminología, en las Primeras Jornadas Iberoamericanas de Terminología en el Instituto Caro y Cuervo de la ciudad

de Bogotá, en Coordinación con las Escuelas Interamericana de Bibliotecología y de Idiomas de la Universidad de Antioquia, así como con la Pontificia Universidad Javeriana de Bogotá y con el apoyo de Unión Latina y la Red Iberoamericana de Terminología, Riterm. En esta oportunidad se dan cita, las profesoras María Teresa Cabré y Mercé Lorente por España, Ana María Cardero y Leticia Leduc, por México, Bertha Nelly Cardona y María Cecilia Plested por Medellín, y Lucy Espinosa por Bogotá. Se tratan temas relacionados con la síntesis histórica y el panorama actual de la Terminología, La Terminología en España, La Terminología en América Latina, Terminología y discursos especializados, Terminología, inteligencia artificial e Ingeniería lingüística, Terminología y Lexicografía, De los corpus textuales a la Terminología, Terminología y Documentación, Documentación en Terminología, la Terminología propiamente dicha, la Informática en el trabajo terminológico, sistemas de gestión y tratamiento de bases de datos, y se impartieron talleres cuyas temáticas principales se centraron en ¿Cómo obtener información terminológica en Internet?, ¿Cómo extraer textos y glosarios de Internet?, ¿Cómo tratar un texto para fines terminológicos? ¿Cómo analizar los textos usando programas informáticos? ¿Cómo crear una base de datos terminológicos?

### 3. ETAPA DE CONSOLIDACIÓN, PRIMEROS AÑOS DEL SIGLO XXI

A partir del año 2000 se vincula con la Escuela Interamericana de Bibliotecología la profesora María Teresa Múnera, con el interés de desarrollar un importante trabajo en el campo de la Terminología y de continuar el trabajo iniciado en años anteriores, por los profesores Uriel Lozano y Bertha Nelly Cardona.

En ese año la Escuela participa de manera dinámica en la elaboración del Tesoro Colombiano en Familia y Género, investigación coordinada desde el Departamento de Trabajo Social de la Facultad de Ciencias Sociales de la Universidad de Antioquia, con la coordinación de la profesora Angela María Quintero y con la participación de las profesoras Bertha Nelly Cardona y María Teresa Múnera. Con esta investigación, la Escuela Interamericana de Bibliotecología participa en la III Escuela de Verano de Terminología, en junio de 2001, con un



*poster* en el cual se expone la estructura y resultados de la investigación sobre el Tesoro Colombiano en Familia y Género. En ese mismo año se inicia el grupo de investigación en Gestión del Conocimiento GECOB, con la investigación denominada :”Incidencia de la Gestión del Conocimiento en el desarrollo y avance de las empresas de servicios de Medellín entre 1995 y 2001”, uno de cuyos objetos de estudio, es precisamente la Terminología, como la herramienta que contribuye a estructurar los procesos de gestión del conocimiento en todas las instituciones. Esta investigación da origen a diferentes artículos y ponencias de carácter local, nacional e internacional. Es así como se realizan publicaciones en la *Revista Interamericana de Bibliotecología*, tales como: “La terminología: un campo del saber transdisciplinario”, “La gestión del conocimiento en la empresa: terminología y documentación, elementos para su medición”, “la gestión del conocimiento y su relación con la documentación”, “La terminología como elemento de la gestión del conocimiento en las organizaciones: Empresas Públicas de Medellín-EEPPM”. Y a nivel internacional se publica el artículo titulado “La gestión del conocimiento en las empresas de servicios de Medellín” en la *Revista Investigación Bibliotecológica* del Centro Universitario de Investigaciones Bibliotecológicas de la UNAM, en México, en el segundo semestre de 2003.

De igual forma se tiene la oportunidad de participar en diferentes eventos académicos, en los que se presentan ponencias relacionadas con el trabajo terminológico en la Escuela Interamericana de Bibliotecología, entre las cuales se pueden mencionar las siguientes : El Seminario Taller sobre Terminología e interpretación de conferencia, el III Coloquio Nacional de Didáctica de la Terminología y Traducción, el III Seminario de Lenguajes profesionales, terminología y disciplinas afines. Y en el año 2001, la Escuela había participado en la III Escuela Internacional de Verano de Terminología y en el segundo semestre se desarrolló la pasantía del estudiante Francisco Antonio Mateos de la Universidad de Granada, quien desarrollaba un módulo sobre introducción al conocimiento de la Terminología. Este año se cumple una importante actividad académica en torno al desarrollo de la terminología en la E.I.B. y se participa con ponencias en los siguientes eventos : 7º Congreso Nacional de Bibliotecología y 1º. de Archivística, en la ciudad de Bogotá.



En el segundo semestre del mismo año, la Escuela Interamericana de Bibliotecología, organiza en coordinación con otras universidades de Cartagena y Bogotá, así como de Riterm, y Colterm, entre otras, el VIII Simposio Iberoamericano de Terminología: la Terminología entre la globalización y la localización, en la ciudad de Cartagena de Indias, entre el 28 y 31 de octubre del 2001. En esa oportunidad se dieron cita varios expertos de la Terminología de toda Ibero América, entre los cuales se pueden resaltar: Ieda María Alves de Brasil, la entonces presidenta de Riterm, Bertha Nelly Cardona y Ana María Cardero, Vicepresidentas de Riterm, Daniel Prado, secretario y Carlos Tebé, Tesorero de Riterm y otros expertos internacionales tales como Rosa Estopá, Alicia Fedor de Diego, Mercé Lorente, entre otros. Allí la Escuela Interamericana de Bibliotecología aportó la presencia de docentes y estudiantes que de manera entusiasta participaron tanto en calidad de asistentes como de ponentes. Algunas de las temáticas que se trataron en dicho evento fueron: Terminología y meta cognición, el uso de la Terminología en textos, neologismos terminológicos en el área de economía, la variación terminológica en las aplicaciones terminográficas, la Terminología y las TIC, nuevos retos, la Terminología y su relación con la gestión del conocimiento, mapas autoorganizativos de kohonen para la representación del conocimiento, etcétera.

En el 2003 se participa en el IV Coloquio Internacional de ISKO, realizado en la Universidad de Salamanca, España. Y el mismo año se participa en el Primer Congreso Internacional de Gestión del conocimiento y calidad, organizado por la Universidad Industrial de Santander en Cartagena de Indias. De igual forma, la Escuela Interamericana de Bibliotecología, empieza a formar parte del Comité Editorial de la *Revista Debate Terminológico de la Red Iberoamericana de Terminología Riterm*, cargo que se ocupó hasta mediados del año anterior.

En noviembre de 2004, organiza el IX Simposio Iberoamericano de Terminología, en la ciudad de Barcelona, con la presentación de un *póster* sobre la investigación de la “Incidencia de la Gestión del conocimiento en las empresas de servicios de Medellín” y en el 2005, se estructura el grupo de estudio en Terminología de la Escuela Interamericana de Bibliotecología, como grupo en formación para convertirse en Grupo de Investigación EIBTERM, a partir del 2006. En la actuali-

dad dicho Grupo de Investigación está conformado por docentes, estudiantes y egresados de la Escuela Interamericana de Bibliotecología y bajo la coordinación del profesor Carlos Arturo Montoya. Por otra parte se participa en el Primer Seminario Nacional de Terminología y Discurso Especializado en la Universidad Autónoma de Manizales, en el mes de septiembre de 2005.

En la actualidad el grupo de investigación EIBTERM es reconocido por Colciencias, máximo organismo de investigación en Colombia, y se espera que muy pronto obtenga la clasificación que lo ubica como uno de los más importantes grupos de investigación de la E.I.B. En la actualidad el interés académico del grupo es profundizar en el estudio de la relación terminología-documentación, así como en la definición de la Unidad Informacional como componente vital de la Terminología dentro de los procesos de organización de la información y la documentación.

#### 4. CONCLUSIONES

Para concluir es importante considerar lo importante que fueron las dinámicas sobre la Terminología por parte de los primeros docentes que se desempeñaron en la Escuela Interamericana de Bibliotecología, así como el aporte y legado que dejaron como semilla de lo que sería el trabajo terminológico en la Escuela de manera especial en la década de los años 90 del siglo XX.

Con el pasar del tiempo se aprecia la evolución y el desarrollo de la actividad terminológica en la Escuela Interamericana de Bibliotecología, hasta convertirse en uno de los contenidos que conforman el núcleo de organización de la información de la estructura curricular. Asimismo se resaltan los aportes de la producción académica de los docentes y estudiantes, traducidos en la elaboración de libros, artículos de revistas, investigaciones, actividades académicas y en la conformación de grupos de investigación en esta área del conocimiento.

Cabe resaltar que el trabajo terminológico ha sido aceptado y solicitado por otras dependencias de la Universidad, en el caso concreto por la U de A, y adicionalmente otras dependencias están solicitando apoyo de las Escuelas de Bibliotecología e Idiomas para que los asesoren y participen en trabajos terminológicos.

“La colaboración de Riterm, el apoyo académico de varias universidades, la cualificación y capacitación de los miembros, han sido indudablemente elementos muy positivos para el desarrollo de Colterm, lo cual esperamos que se siga dando y convertirnos si no en un modelo a imitar, sí en una buena experiencia a analizar y tener en cuenta”.<sup>6</sup>

## 5. BIBLIOGRAFÍA

*Anthology of Readings in Library Science, una antología de lecturas en Bibliotecología en inglés par el estudio de la terminología profesional*, compilación de Victoria Fisher de Duque y Gaston Litton, ed. Preliminar, Medellín, UdeA, E.I.B., 1961, T. 1.

“Diccionario Especializado de Trabajo Social: un aporte a la profesión desde la metodología”, Gloria H. Montoya Cuervo y Bertha Nelly Cardona Rave y Cecilia Inés Zapata López, en *Revista Colombiana de Trabajo Social*, (Colombia), Núm. 14, 2000; pp. 143-152.

*El Estudio de la Terminología en la Escuela Interamericana de Bibliotecología: una aproximación a sus momentos históricos*, Uriel Lozano Rivera. Medellín, UdeA, E.I.B., 2007, 3 p.

Jornadas Iberoamericanas de Terminología (1:1999: Bogotá) panorama general, fundamentos y aplicaciones, Bogotá, UdeA, PUJ, Riterm:Colterm, 1999, pag. Variada.

[PRESENTACIÓN REVISTA INTERAMERICANA DE BIBLIOTECOLOGÍA] / Berta Nelly Cardona R., Medellín, Vol. 22, núm. 1 (ene-jun, 1999) , p. 5.

PROFESSIONELLE TRMINOLOGIE DER BIBLIOTHEKSWISSENSCHAFT  
: siete categorías de palabras y términos afines en biblio-

---

6 Bertha Nelly Cardona Rave, “Red Colombiana de Terminología Colterm”, en *Revista Interamericana de Bibliotecología*, Medellín, Vol. 22, Núm. 1 (ene-jun 1999), pp. 129-138.

***Primer Simposio Internacional sobre Organización...***

tecnología en alemán, con sus equivalentes en español e inglés, compiladas por Edith Schwarz y Gaston Litton, Medellín, UdeA, EIB, 1960.

“Red Colombiana de Terminología COLTERM”, Berta Nelly Cardona Rave, en *Revista Interamericana de Bibliotecología*, Vol. 22, núm. 1 (ene-jun. 1999), pp. 129-138.

*Revista Interamericana de Bibliotecología*, Universidad de Antioquia, Escuela Interamericana de Bibliotecología, Medellín, Vol. 22, Núm.1 (ene-jun. 1999), 166 p.

Seminario Nacional de Terminología (2:1995: Medellín), Medellín, UdeA, E.I.B., 1995, 69 p.

Seminario de Terminología (1996:Medellín), Sistematización y representación del conocimiento en la terminología : estructuración y representación del conocimiento terminológico, Medellín, UdeA, 1996, 16 p.

La Terminología: macro teórico, metodología, consecuencias prácticas. [curso-taller], Medellín, Universidad de Antioquia, E.I.B : E.I., 1997.

La Terminología en Algunos Diccionarios de Bibliotecología, José Luis Orduña López, en *Revista Interamericana de Bibliotecología*, Medellín, Vol. 22, núm. 1 (ene-jun. 1999), pp. 37-60.

La Terminología en el Nuevo Orden de la Información: globalización e internacionalización [curso-taller], Medellín, UdeA, E.I.B., 1995.

Terminología Profesional, compilada por Victoria Fisher de Duque, Emma Crosland Simonson, Gaston Litton, ed. Preliminar, Medellín, Universidad de Antioquia, EIB, 1958.

# Ontología sobre el desplazamiento forzado en Colombia<sup>1</sup>

SANDRA ARENAS GRISALES  
*Escuela Interamericana de Bibliotecología  
Universidad de Antioquia, Colombia*

## INTRODUCCIÓN

**E**n 1995 la Conferencia Episcopal de Colombia publicó los resultados de su investigación “Derechos Humanos: Desplazados por Violencia en Colombia”, en ella puso en evidencia ante el país la tragedia de las personas que huían en medio del conflicto armado colombiano. Las cifras eran alarmantes, para el periodo comprendido entre 1985 y 1995 el estudio de la Conferencia estimó un promedio de **586.261** personas en situación de desplazamiento<sup>2</sup>. A pesar de la contundencia de las cifras, la reacción de la sociedad, el Estado y la comunidad internacional osciló entre la indiferencia, la incredulidad

---

1 Investigación en curso, financiada por el CODI, Universidad de Antioquia. Participan en ella: Sandra Arenas Grisales, docente Universidad de Antioquia, investigadora principal; Ana María Tangarife, jefe del Centro de Documentación del Instituto de Estudios Políticos, coinvestigadora; Jaime Guzmán Luna, docente Universidad Nacional sede Medellín, asesor académico; Wilmar Arley Martínez Márquez, auxiliar de investigación; Néstor Eduardo Rivera Vidal y Grissa Vianney Maturana González estudiantes Escuela de Sistemas, Universidad Nacional de Colombia sede Medellín.

2 Conferencia Episcopal de Colombia, *Derechos humanos: desplazados por la violencia en Colombia*, Bogotá, Conferencia Episcopal de Colombia, 1995.

y el espanto. Sin embargo, el principal logro del estudio de la Conferencia Episcopal fue presentar el desplazamiento forzado como una violación masiva y múltiple de los derechos humanos y una infracción grave del Derecho Internacional Humanitario. Desde ese momento el fenómeno no pudo ser ignorado por la sociedad colombiana, el tema del desplazamiento se tornó un debate público de interés para investigadores sociales, académicos, organizaciones no gubernamentales, entidades estatales y organismos internacionales. No obstante, el éxodo continúa, según cálculos del Codhes, entre 1985 y septiembre de 2005, el acumulado alcanza la cifra de 3,662.842 personas.<sup>3</sup>

Hoy los desafíos en relación con el desplazamiento forzado son enormes, pero existe una mayor conciencia nacional y un acumulado de experiencias, desarrollos institucionales y legales, análisis e investigaciones desde la academia y las ONG, que permiten pasar de la denuncia del fenómeno a un intento por comprender las múltiples causas y consecuencias asociadas a él. Sin lugar a dudas, la suerte que corran los desplazados será también la suerte que corramos como país y de la forma como logremos enfrentar y resolver este problema —entre muchos otros— dependerá en buena medida nuestro futuro como nación.

En este proceso es central el papel que desempeña la academia en la comprensión del problema y su visualización. Existe hoy una amplia producción académica que aborda el tema desde distintas perspectivas y enfoques teóricos y metodológicos, pero sin que se presente aún entre los investigadores y analistas claridad y unificación en relación con los conceptos y los términos que utilizan, por el contrario hay dispersión, ambigüedad y poca rigurosidad en el manejo de éstos. Y si bien existe en la normatividad internacional un conjunto de términos claramente definidos y utilizados para explicar fenómenos sociales de este tipo,<sup>4</sup> los mismos no pueden ser mecánicamente

---

3 Conferencia Episcopal de Colombia, CODHES, *Desafíos para construir nación. El país ante el desplazamiento, el conflicto armado y la crisis humanitaria 1995-2005*, Bogotá, Conferencia Episcopal de Colombia, 2005, p. 19.

4 *Tesaurus de Derechos Económicos, Sociales y Culturales*, [en línea] disponible en Internet: : <http://ip.aaas.org/sthesaurus.nsf> [consulta: 3 de agosto de 2007].

impuestos para analizar el desplazamiento en Colombia, pues la particularidad y complejidad del conflicto que vivimos hace necesaria su creación, contextualización o resignificación.

Por otra parte se presenta el problema de la dispersión de la información y de la invisibilización de la misma, que en lo macro implica un efecto político y social, pues hablamos de un problema que afecta a buena parte de la población pero sobre el que no se discute ampliamente y en relación con el cual no existen cifras claras acerca del número de víctimas, la atención que reciben y las posibilidades de reparación y de su retorno. Todo ello en una sociedad como la nuestra que hoy más que nunca precisa de la verdad para salir del círculo de violencia y exclusión en el que está inmerso. En el nivel micro, en relación con aquello que profesionalmente nos compete como documentalistas, el problema que genera la dispersión de la información se expresa en la dificultad para ser ésta descrita y recuperada y en consecuencia la poca difusión de los estudios e informes de investigación que se realiza.

Para intentar resolver este problema desde el aspecto documental, el Grupo de Terminología de la Escuela Interamericana de Bibliotecología, EIB-Term, en asociación con la Unidad de Documentación del Instituto de Estudios Políticos de la Universidad de Antioquia, y con la asesoría y el acompañamiento del Grupo de Inteligencia Artificial de la Universidad Nacional de Colombia, Sede Medellín, desarrolla actualmente la investigación *Ontología sobre el desplazamiento forzado en Colombia integrado a un sistema de información*.

El objetivo general es construir una Ontología donde se represente conceptualmente la terminología sobre el desplazamiento forzado en Colombia, de manera que permita la representación y recuperación de la información. Dicha Ontología se constituye en el soporte estructural de un sistema de documentación en red sobre desplazamiento forzado, que permitirá el acceso a información referencial y en texto completo (investigaciones, informes, memorias de eventos, legislación, políticas públicas, fotografías, mapas, etcétera). Se espera crear, a partir del consenso en relación con la ontología y su uso, una red colaborativa para intercambiar, divulgar y recuperar la producción académica sobre el desplazamiento forzado en Colombia.

La investigación dio inicio en abril del presente año y esta ponencia tiene como objetivo presentar un avance del enfoque metodológico utilizado en las primeras etapas de la investigación, en relación con la definición del *corpus* bibliográfico, la selección de términos y el análisis de la información; el uso y la aplicación de software de análisis lexicográfico para la selección y el análisis de los términos, las estrategias utilizadas en su validación con los especialistas y las decisiones en relación con el software que se va a utilizar.

Vamos a iniciar con una definición de ontología y una de las más citadas es la de Gruber, de acuerdo con él:

“Ontología es una especificación formal y explícita de una conceptualización dada, lo que existe es aquello que puede ser representado [...] Cuando un conocimiento de un dominio es presentado en un formalismo declarado, el conjunto de los objetos que pueden ser representados son llamados del universo del discurso. Ese conjunto de objetos, y las relaciones descritas entre ellos, son reflejados en un vocabulario representacional con el cual un programa de conocimiento de base representa el conocimiento”.<sup>5</sup>

Se deriva de esta definición lo siguiente: ontología es un conjunto de términos que pueden ser usados para describir un área del conocimiento o construir su representación. Estos términos definen un dominio dado y pueden ser usados por las máquinas para responder consultas o preguntas de búsqueda. Sin embargo, más que el vocabulario en sí mismo, en la ontología son esenciales los conceptos asociados a los términos, de ahí su papel central en el desarrollo de la Web semántica, pues intenta proporcionarle a las máquinas información amplia, de manera que, a partir de las relaciones entre los conceptos y de las inferencias lógicas, se puedan resolver consultas de los usuarios con mayor precisión.<sup>6</sup>

---

5 Thomas R. Gruber, “*Toward principles for the Design of Ontologies Used for Knowledge Sharing*”, en *Formal ontology in conceptual analysis and knowledge representation*, March 1993, Padova, Italy. Available as Technical Report KSL 93-04, Stanford University, citado por: Carlan, Eliana, *Ontologia e web semântica*, Monografía Departamento de Ciência da Informação e Documentação, Brasília, junio de 2006.

6 *Ibid.*



Por otra parte, las ontologías, según Eliana Carlan “...presentan un modelo relacional de entidades en un dominio particular del conocimiento. El objetivo de su construcción es la necesidad de un vocabulario compartido donde la información pueda ser intercambiada y también usada por los usuarios de una comunidad, sean ellos humanos o agentes inteligentes”.<sup>7</sup>

En resumen y en palabras de Lozano Tello, ontología es una herramienta que define conceptos y relaciones propias de un dominio de forma compartida y consensuada, representada de manera formal, legible y utilizable por el computador.<sup>8</sup>

La ontología sobre el desplazamiento forzado en Colombia especifica los términos, los define y relaciona formalmente con otros, proporcionando una mayor variedad de relaciones que luego podrán ser utilizadas por los sistemas automáticos para resolver consultas.

## METODOLOGÍA

La metodología desarrollada en la construcción de la ontología retoma elementos propios de la terminología, en tanto que busca denominaciones conceptuales. Se establece una lista de conceptos que forman parte de una estructura nocional, los cuales se encuentran interrelacionados lógicamente y ontológicamente y cuyo conjunto constituye el sistema conceptual de una disciplina o actividad especializada. A cada concepto se le atribuye una denominación correspondiente a la utilizada por el especialista, este proceso se llama onomasiológico y tiene dentro de la documentación una aplicación particular que es la recuperación de la información.

Para construir la Ontología seguimos una metodología que integra la sugerida por Noy y McGuinness, quienes proponen un modelo para el desarrollo de ontologías,<sup>9</sup> y la propuesta metodológica propia de la

---

7 *Ibíd.*, p. 60.

8 Adolfo Lozano Tello, Ontologías en la Web semántica, [en línea] disponible en Internet: <http://www.informandote.com/jornadasIngWEB/articulos/jiw02.pdf> [Consulta: 1 de julio de 2007].

9 Natalya F., NOY, MCGUINNESS, Deborah L., *Ontology Development 101: A Guide*

Terminología para el desarrollo de tesauros. Los pasos que hemos establecido son los siguientes:

## 1. Identificación del problema

En este primer punto debe responderse a preguntas tales como: ¿cuál es el dominio? ¿Para qué será usada la Ontología? ¿Qué tipo de preguntas deberá responder la Ontología? ¿Quiénes serán los encargados de mantener la Ontología?

En relación con la primera pregunta, el tema a cubrir, como se dijo antes, es el desplazamiento forzado en Colombia, e intenta abarcar todo el fenómeno: causas, consecuencias, características de la población desplazada, asistencia humanitaria que se brinda a estas personas; pero también incluir las formas de investigar el fenómeno y las metodologías utilizadas. El problema que se busca solucionar es la dispersión y la dificultad para recuperar la información que se produce, debido a la diversidad, incoherencia y poca precisión de los términos con los cuales son descritos los trabajos.

La Ontología será utilizada, en primer lugar, para estructurar un sistema de información documental en la Unidad de Documentación del Instituto de Estudios Políticos de la Universidad de Antioquia. En segundo lugar se espera generar consenso entre las instituciones dedicadas a la investigación sobre el tema, para que en la descripción de sus documentos utilicen la Ontología de manera que los agentes puedan recuperar el mayor número de información posible. En tercer lugar, se les quiere proporcionar a bibliotecólogos y documentalistas una herramienta terminológica para describir el material sobre el desplazamiento para que los catálogos y bases de datos de las bibliotecas tengan una recuperación eficiente del material. En síntesis, procuramos crear un lenguaje de recuperación común a todos, que facilite la comunicación.

---

to *Creating Your First Ontology*, Stanford University, Stanford, 2001, [en línea] disponible en Internet: <[http://www.ksl.stanford.edu/people/dlm/papers/ontology101/ontology101-60\\_noy-mcguinness.html](http://www.ksl.stanford.edu/people/dlm/papers/ontology101/ontology101-60_noy-mcguinness.html)> [Consulta: 23 de julio de 2007].

## 2. Definición de criterios para la construcción de la Ontología

Se retomaron los criterios propuestos por Gruber.<sup>10</sup>

**a. Claridad:** se espera que los términos comuniquen y signifiquen algo; que sus definiciones sean objetivas, completas, expresadas en un lenguaje natural y documentadas.

**b. Coherencia:** las inferencias deben ser consistentes con las definiciones, si una sentencia inferida a partir de los axiomas de la Ontología, contradice una definición, entonces la Ontología es incoherente. La coherencia debe también aplicarse a los conceptos, tanto aquellos definidos informalmente como los bien documentados.

**c. Aplicabilidad:** toda Ontología debe ser planeada teniendo en cuenta la posibilidad de ser aplicada y extendida a otras áreas. Nuestra Ontología se convierte en la estructura base de un sistema de información documental en línea sobre el desplazamiento forzado. Por otra parte se crea con la intención de ser aplicada en otras instituciones o incluso extendida a otras áreas.

**d. Compromiso ontológico y con la implementación:** una Ontología debe hacer pocas imposiciones sobre el dominio en el que se basa, tener flexibilidad y permitir la aplicación en otras áreas. Como lo planteamos en los objetivos, esperamos generar consenso alrededor de la Ontología entre especialistas e investigadores, pero también entre aquellos que desarrollan bases de datos, observatorios, páginas Web. Para lograr esto se contempla la validación por parte de los expertos, de los términos y sus definiciones, y una amplia difusión cuando se tenga el resultado final.

---

<sup>10</sup> Tom Gruber, What is an Ontology? [en línea], disponible en Internet: <http://ksl.stanford.edu/kst/what-is-an-ontology.html> [consultado. 23 de agosto de 2007], citado por. Elina Carlan, *Op. Cit.*

### 3. Definición del corpus bibliográfico

En vista de la considerable producción académica se hace necesario reducir el *corpus* bibliográfico sobre el cual se trabaja. En relación con el Desplazamiento en Colombia, la cantidad considerable de instituciones de educación superior, ONG y entidades estatales dedicadas al estudio de la cuestión, hizo necesario reducir el *corpus* bibliográfico y retomar la documentación producida por las principales instituciones: ACNUR (Alto Comisionado de las Naciones Unidas para los Refugiados), Universidad de Antioquia, Universidad Nacional de Colombia, Universidad de los Andes, CODHES (ONG), la Universidad Javeriana y la Organización Internacional para las Migraciones. El criterio para elegir estas instituciones fue, por una parte, su amplio recorrido académico, por la otra, la variedad de enfoques y metodologías utilizadas que van desde ver el fenómeno como un tema de derechos humanos, hasta analizarlo en la perspectiva de los mecanismos de producción y acumulación del capitalismo y las estrategias de dominación que imponen ciertos sectores en nuestro país.

Se seleccionaron entonces investigaciones, informes, tesis, artículos, boletines, un total de 121 textos. El periodo comprende los años 1996 y 2006 y se limita a la producción que sobre el tema se hace en Colombia.

### 4. Elaboración del mapa conceptual o árbol de dominio

Debido a lo vasto del tema y a las múltiples aristas que sobre el fenómeno se presentan, nos dimos a la tarea de elaborar un mapa conceptual sobre él que sirviera inicialmente de guía para la comprensión cabal del fenómeno y sus interrelaciones, pero también como estructura básica de la Ontología. Es así como se hizo una lectura analítica de los textos más representativos en cada una de las instituciones, intentando rastrear allí las distintas hipótesis interpretativas, las características propias del desplazamiento, los actores involucrados, las acciones políticas y militares de esos actores, la intervención estatal en el fenómeno y la acción social derivada de las formas organizativas de los desplazados.

Luego de elaborar el borrador inicial, el resultado fue puesto a consideración de los expertos en el tema y una vez acatadas sus observaciones se elaboró el siguiente mapa:

**FACETA A : CAUSAS**

CONFLICTO ARMADO  
MODELO ECONÓMICO  
ACTORES

**FACETA B: CARACTERÍSTICAS**

POBLACIÓN  
TRAYECTORIAS  
TIPOLOGÍAS

**FACETA C: CONSECUENCIAS**

CRISIS HUMANITARIA  
DESINTEGRACIÓN Y REARTICULACIÓN  
IMPACTOS SOCIOECONÓMICOS  
RECOMPOSICIÓN DEL TERRITORIO

**FACETA D: ATENCIÓN A LA POBLACIÓN**

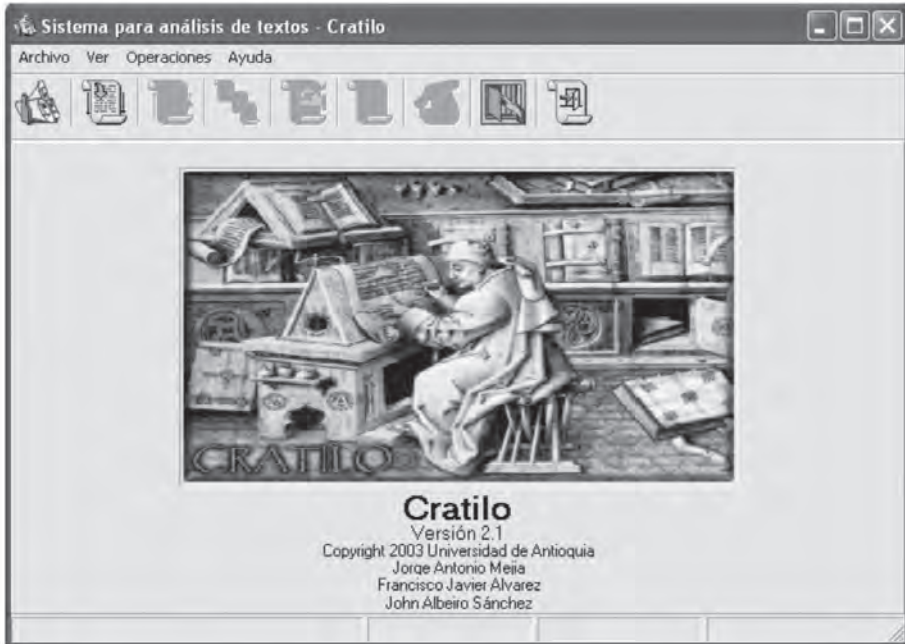
PREVENCIÓN  
ASISTENCIA HUMANITARIA  
RETORNO  
ESTABILIZACIÓN  
Y CONSOLIDACIÓN SOCIOECONÓMICA

**FACETA E: METODOLOGÍAS DE INVESTIGACIÓN**

CUANTITATIVA  
CUALITATIVA

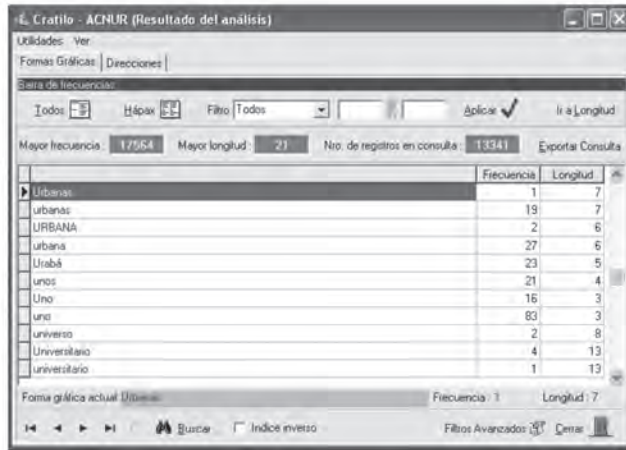
## 5. Rastreo de los términos dentro de los textos especializados

Para seleccionar los términos preliminares utilizamos una herramienta llamada Cratilo.

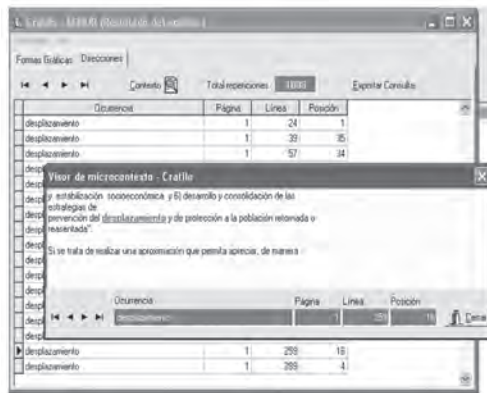


El programa Cratilo 2, fue desarrollado en el Instituto de Filosofía de la Universidad de Antioquia. Cratilo es un sistema para el análisis lexicográfico de textos, que identifica las palabras que aparecen en éstos, su frecuencia y su ubicación en el texto físico.

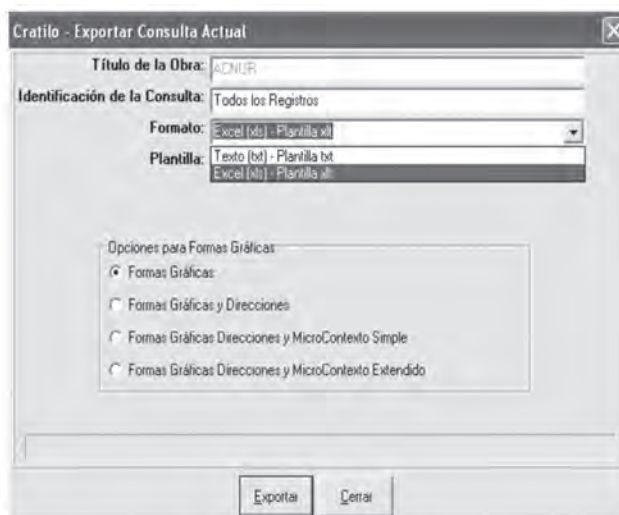
Mediante un proceso de reconocimiento, Cratilo identifica todas las palabras presentes en el texto y construye una base de datos que se convierte en el proyecto de análisis de la obra. Cada proyecto lo constituye el conjunto de textos seleccionados en una institución por la investigación.



Una ventana de consultas permite explorar la base de datos para establecer las palabras, su frecuencia o longitud. Las consultas se pueden realizar en el Índice Alfabético o en el Índice Inverso. Igualmente Cratilo provee un Visor de Formas Gráficas; es decir una interfaz que muestra las formas gráficas encontradas en un texto con sus direcciones y el contexto en el cual aquéllas se presentan.



La información que se va almacenando en la base de datos de Cratilo puede ser exportada a formatos de texto y de hoja electrónica utilizando plantillas definidas para ello.



El resultado de la consulta arroja la siguiente hoja de cálculo:

	A	B	C	D
1	Título de la Obra			
2	ACNUR			
3	Identificador de la Consulta			
4	Todos los Registros			
5	Número de Registros			
6	13341			
7	No.	Forma Gráfica	Frecuencia	Longitud
13235	8931	desde	170	5
13236	12637	acceso	171	6
13237	3503	políticas	175	9
13238	5029	mayor	176	5
13239	10145	conflicto	176	9
13240	4593	necesidades	177	11
13241	7694	está	178	4
13242	11444	Atención	178	8
13243	4808	mismo	179	5
13244	9502	cuenta	179	6
13245	137	violencia	179	9
13246	9002	desarrollo	179	10
13247	9561	cual	180	4



Luego se realiza una depuración de estas formas gráficas, se eliminan términos no representativos (cifras, conectores, artículos, etcétera) y posteriormente se seleccionan aquellas palabras que posean un valor semántico para los fines de la investigación.

## 6. Elaboración del sistema de conceptos

El primer paso para la implementación de ontologías, es definir los elementos que conformarán el tesoro, para luego establecer sus relaciones en el dominio específico del desplazamiento forzado.

Para conformar el tesoro utilizaremos una versión modificada de la ontología Skos-Core, la cual permite modelar explícitamente los conceptos y sus relaciones de jerarquía, asociación y equivalencia al igual que permite desarrollar nuevas relaciones inferidas a partir del uso de axiomas. El vocabulario de Skos-Core es una aplicación de RDF<sup>11</sup> y provee algunos formalismos para hablar de términos, sus propiedades, interrelaciones y categorías (clases).

En la investigación retomaremos las siguientes clases de Skos-Core:

- Skos:Concept: término de entrada
- Skos: Alterlabel: use
- Skos: broader término amplio
- Skos narrower: término específico
- Skos prelabel: término preferido
- Skos related: término relacionado
- Skos Top concept: faceta

Además de las relaciones que acabamos de mencionar, que son las tradicionales de un tesoro, la ontología skos-core puede ser extendida agregándole nueva información asociada a los conceptos. En el caso particular del proyecto, extenderemos la caracterización de los términos con los siguientes atributos:

---

11 RDF (*Resource Description Framework*) / RDFS (*RDFSchema*): desarrollado por W3C tiene como objetivo representar el conocimiento por medio de la idea de redes semánticas. Son lenguajes que permiten la representación de conceptos, taxonomías de conceptos y relaciones binarias, tomado de Eliana Carlan, p. 38.

Skos:definition: declaración o especificación formal del significado del concepto o término.

Skos editorial note: nota del editor, traductor o encargado del mantenimiento del tesauro.

Skos-example: un ejemplo de uso del concepto.

Skos history notes: una nota acerca del estado anterior o cambios en el uso o significado del concepto.

Skos note: cualquier comentario que pueda surgir dentro del concepto.

Skos-scope note: fuente del término.

Término restringido y término amplio, el primero se refiere a aquél propio del dominio, utilizado exclusivamente en él o creado sólo para uso en el área del desplazamiento forzado; el segundo, a aquellos términos tomados de otras áreas del conocimiento.

Hasta aquí hemos hablado del tesauro y su implementación. En cuanto a su configuración en el dominio del desplazamiento forzado, cada uno de los términos rastreados en el *corpus* bibliográfico, se implementará sobre esta ontología mediante el uso del Protégé,<sup>12</sup> que es un ambiente interactivo para desarrollo de ontologías. Ofrece una interfase gráfica para la edición de las ontologías y una arquitectura para la creación de herramientas basadas en conocimiento

## 7. Validación con los especialistas

La validación con los especialistas es fundamental para el desarrollo de cualquier Terminología; en buena medida de tal validación dependerá el consenso que pueda lograrse en el uso de esa Terminología. En la investigación se tiene previsto que en cada una de las etapas se realizará una sesión de trabajo con los especialistas, en la cual se validarán los términos, sus definiciones y sus relaciones.

---

12 Protégé: es un ambiente interactivo para desarrollar Ontologías. Ofrece una interfaz gráfica para editar las ontologías y una arquitectura para crear herramientas basadas en el conocimiento.

En esta primera etapa de la investigación se convocó a cuatro especialistas que trabajan en el tema a partir de diversos enfoques (político, sociológico, psicológico y de intervención) para discutir el mapa conceptual que desarrollamos. A cada especialista se le envió previamente una presentación del proyecto y del mapa conceptual, la intencionalidad del mapa y la bibliografía utilizada para realizarlo. En una sesión conjunta se analizó el contenido, el enfoque, la organización y el lugar que cada uno de los términos ocupaba. Al final se logró un consenso en relación con la estructura del mapa conceptual.

### POSIBILIDADES Y VENTAJAS EN EL USO DE LA ONTOLOGÍA SOBRE DESPLAZAMIENTO FORZADO

La Ontología es una herramienta para crear y estructurar sistemas de información documental y para recuperar la información. Esta Ontología permite estructurar un conocimiento acerca de las relaciones que existen entre los términos de un dominio específico. Sobre esta base se pueden utilizar mecanismos de inferencia que permitan expandir este conocimiento estableciendo nuevas relaciones entre los conceptos que no están explícitamente definidos. Lo anterior permite tener estructuras de tesoro mucho más complejas que las tradicionales y en consecuencia, el proceso de la recuperación de información puede ser mejorado.

Otra posibilidad de la Ontología es convertirse en una herramienta de comunicación. En la medida en que se logre en su elaboración definir y relacionar los conceptos de un dominio dado, y generar consenso alrededor de esos conceptos, la Ontología logrará que tanto quien produce la información, como el documentalista, el administrador de los sistemas de información y el usuario final, hablen un mismo lenguaje y que por lo tanto la comunicación y la recuperación de la información sea eficiente.

No obstante, y esta es tal vez una de sus características más interesantes, ese vocabulario no es estático, cambia de la misma forma en que el saber en su dominio lo hace. El documentalista por lo tanto debe estar en permanente comunicación con los especialistas y atento a esos cambios para implementárselos a la ontología.

## PROBLEMAS QUE SE HAN PRESENTADO EN EL DESARROLLO DE LA ONTOLOGÍA

1. Una de las características fundamentales de la Ontología es que debe existir un cierto nivel de consenso entre los especialistas en relación con los términos, los conceptos y las inferencias. Desde la elaboración del mapa conceptual que sirve de base a la Ontología que desarrollamos, puede evidenciarse que hay un enfoque analítico que ve el fenómeno como una estrategia de guerra y como una consecuencia clara de las lógicas del mercado. Si bien es éste un problema de derechos humanos y vivimos una catástrofe humanitaria, el fenómeno no es considerado como un problema de derechos humanos, sino como un problema social que tiene como una de sus consecuencias la violación de los derechos humanos. Esta posición es compartida por aquellas instituciones de las cuales hemos extraído nuestro *corpus* bibliográfico, pero podrá ser cuestionada por otros enfoques y otros investigadores.

2. Esto nos remite a un segundo problema y éste es en relación con el consenso. Si bien hemos implementado en cada uno de los pasos la validación con los especialistas, ello no garantiza que logremos consenso en el medio académico debido a que hay en la Ontología un enfoque particular de análisis del problema, y ese enfoque no es compartido por todas las instituciones dedicadas a su estudio. El consenso puede lograrse en cuanto a las denominaciones y sus definiciones, pero no tanto en las relaciones y en la estructura conceptual de la Ontología. Sin embargo, este problema no descarta las posibilidades de uso de la Ontología como una herramienta que permite la recuperación de documentos.

3. Estamos analizando un fenómeno social relativamente reciente y que tiene características particulares, de ahí la dificultad en la exactitud de los términos y sus definiciones. No se trata de una ciencia exacta o de una área del saber sobre la cual se tenga ya bases sólidas; por el contrario, estamos en el proceso de configuración, análisis y construcción de un corpus terminológico, en relación con el cual muchas de las categorías propias de las ciencias sociales no son aplicables, y por lo mismo surgen otras que deberán esperar un tiempo para su

validación entre los académicos. Sin embargo en la investigación estamos concientes de ello y de la necesidad de adaptarnos, de entender la evolución de los conceptos y las dinámicas de su creación y difusión.

4. Otro de los problemas que enfrentamos es la pregunta sobre ¿cuál es el límite? Si la Ontología se relaciona directa o transversalmente con tantos otros temas, uno de nuestros problemas a resolver será hasta dónde llegar. Lo paradójico es que este aparente problema se relaciona con una de las ventajas de las ontologías y es la posibilidad de expandir su dominio. Resolvimos este problema centrándonos en la descripción del fenómeno del desplazamiento mientras que algunos otros temas relacionados, como el conflicto armado y las violaciones a los derechos humanos, serán abordados en la medida en que éstos estén directamente relacionados.



## Ontologías para relacionar contenidos educativos digitales

**CLARA LÓPEZ GUZMÁN**

*Dirección General de Servicios de Cómputo Académico  
Universidad Nacional Autónoma de México*

**FRANCISCO J. GARCÍA PEÑALVO**

*Facultad de Informática y Automática  
Universidad de Salamanca, España*

**E**n los sistemas de enseñanza y aprendizaje en línea se pretende que los contenidos educativos tengan un alto nivel de interoperabilidad, con ello se puede explotar al máximo el uso de una infraestructura digital para comunicar, crear y ejecutar los programas académicos. Es necesario que los contenidos puedan utilizarse en diversas aplicaciones, como en las plataformas de aprendizaje, en las bibliotecas digitales, en los repositorios, en los sistemas de información e incluso en páginas web sencillas. Aunque hay estándares tecnológicos para que estas herramientas puedan comunicarse, los métodos de descripción utilizados requieren de mayor definición para emparar los significados de los descriptores y realizar un intercambio realmente eficiente de información y contenidos.

Por otra parte también se vislumbran las primeras aplicaciones de la Web Semántica en la educación y ésta requiere no sólo de la descripción de los contenidos, sino también de ubicarlos semánticamente en un dominio de conocimiento, para que las computadoras puedan ejecutar más procesos automáticos de los que realizan actualmente, y principalmente para mejorar la recuperación e indización de los recursos docentes dispersos en Internet, así como aquellos almacenados en las diversas aplicaciones de los entornos digitales de aprendizaje.

Los metadatos han coadyuvado en la tarea de organizar la información en la Web, al hacer descripciones genéricas y utilizar lenguajes apropiados para que las computadoras puedan indizarlos. Sin embargo, las etiquetas que se generan no proporcionan mayor descripción de los elementos mismos, más allá de datos que la computadora puede procesar pero no comprender. Con el uso de los metadatos y de los tesauros se pueden obtener jerarquías temáticas sobre los contenidos pero no es posible, de forma automatizada, que las computadoras puedan crear relaciones entre contenidos, así como tampoco se puede delimitar el dominio de conocimiento que comprenden.

Para obtener una mejor interoperabilidad entre contenidos y potenciar así las aplicaciones de los sistemas educativos se le está dando auge a la descripción y relación de contenidos a través de Ontologías, para indizar documentos e incrementar el desempeño de los servicios Web; otra aplicación relevante relacionada con la educación se encuentra en la ingeniería del conocimiento, que apoya fuertemente la tendencia hacia la reutilización de conocimiento para generar más conocimiento, y también hacia la construcción de cursos de forma colaborativa para desarrollar e integrar mejores sistemas de aprendizaje basados en la Web. Dentro de los sistemas de información, las Ontologías facilitan la interoperabilidad entre sistemas heterogéneos y la distribución de conocimiento dentro y entre las aplicaciones de software (García 2004).

## LA WEB SEMÁNTICA

Los contenidos disponibles en la Web se encuentran codificados en HTML,<sup>1</sup> un lenguaje que, a través de etiquetas predefinidas, sólo permite codificar contenidos por su estructura y con elementos muy limitados. En este tipo de archivos, únicamente se puede indicar qué es un párrafo y qué es un encabezado pero ningún otro tipo de elementos como el nombre de un alumno y una calificación, ya que no hay etiquetas para ello. En las páginas creadas con HTML, estas etiquetas permiten que la información se despliegue en la pantalla de la computa-

---

1 Hyper Text Markup Language, <http://www.3w.org/html/>



dora con ciertas características y que sean entonces éstas legibles para el usuario, pero para las máquinas estas etiquetas son de muy poca utilidad ya que para procesar la información requieren de más elementos. Ante esta limitante aparece XML,<sup>2</sup> un lenguaje que permite incluir etiquetas específicas para diferentes componentes de un documento, y que tienen una sintaxis más detallada y manejable que permite definir documentos con estructuras más descriptivas.

El principal inconveniente es que estos lenguajes no incluyen semántica sobre esas estructuras o sobre el contenido global y que las computadoras no tienen la inteligencia para deducirlo. También se apuntan otros inconvenientes sobre cómo evitar ruido en las búsquedas, navegación jerárquica y poca flexibilidad (García & Sicilia, s/f).

Para atacar estas deficiencias, que repercuten principalmente en la calidad de las búsquedas y las recuperaciones de los contenidos, se está desarrollando una web complementaria en la que la información es definida por su significado y sus relaciones, y no por su estructura morfológica ni por su estilo gráfico. A esta nueva web se le conoce como Web Semántica.<sup>3</sup>

La Web Semántica se propone darle significado al contenido de los sitios web, y permitir un entorno en el que agentes de software puedan ejecutar tareas complejas para leer las páginas y extraer información más sofisticada, lo cual facilitaría las tareas de indización y de búsqueda pero principalmente las haría más precisas y las acercaría más a la forma en que los seres humanos buscamos (Bernes-Lee, Hendler, Lassila, 2001).

El objetivo más general de la Web Semántica es que la información en la Web sea definida y ligada de tal modo que pueda ser utilizada por máquinas no sólo para propósitos de despliegue sino también para la automatización, la integración y la reutilización de los datos entre diferentes aplicaciones.

Para lograr que los contenidos tengan un significado explícito para las máquinas, la Web Semántica utiliza XML como lenguaje que le da

---

2 eXtended Markup Language, <http://www.w3.org/XML/>

3 <http://www.w3.org/2001/sw/>

una sintaxis a la RDF<sup>4</sup> para modelar las relaciones, y las Ontologías para las descripciones y los vocabularios.

Hoy día, las Ontologías se consideran un elemento clave en la arquitectura de la Web Semántica y se vislumbran como un método explícito para hacer la descripción de contenidos educativos que tengan el potencial para relacionarse y reutilizarse más fácilmente.

## DEFINICIÓN DE LAS ONTOLOGÍAS

Las Ontologías comienzan a utilizarse a finales de los años 80, en el campo de la inteligencia artificial, aplicándose como un método para compartir y reutilizar el conocimiento. A finales de los años 90, comienzan a utilizarse ya en un contexto web con la intención de adjuntarle descripciones semánticas explícitas a contenidos y servicios.

La definición más citada es la que da Gruber (1993), proveniente de la filosofía, que dice: “Una Ontología es una especificación explícita de una conceptualización”. Una conceptualización es una abstracción de algo que existe; es una forma simple de representación. Una especificación explícita es un conjunto de declaraciones sobre esa representación de lo existente. Con los sistemas informáticos, estas declaraciones se llevan a cabo a través de métodos y lenguajes de computadora.

Uschold y Gruninger (1996) proponen que una Ontología es el conjunto de términos utilizado para referirse a un entendimiento compartido en algún dominio de interés, lo cual se utiliza como un marco unificador para resolver problemas de un mismo dominio. Los mismos autores conciben la conceptualización como la definición de conceptos y sus interrelaciones, y concluyen diciendo que asumen una Ontología como un recuento explícito o la representación de una conceptualización, o de parte de ésta.

## ONTOLOGÍAS, TESAUROS Y LOM

La manera más simple de entender una Ontología es considerarla como un vocabulario de términos, y la especificación del significado de

---

4 RDF. *Resource Description Framework* (Estructura para la Descripción de Recursos). <http://www.w3.org/RDF/>

éstos y de sus interrelaciones en un dominio de conocimiento, para hacer búsquedas y otros usos entre agentes. Esta definición se acerca mucho a lo que hemos definido como un Tesouro. García (2004) hace una interesante revisión de diversos autores que analizan las diferencias y coincidencias entre Ontologías y Tesauros, y resalta que por un lado hay Ontologías *Lightweight* (de peso ligero), que incluyen conceptos, las taxonomías de los conceptos, las relaciones entre los conceptos y las propiedades que describen los conceptos; y que este tipo de Ontologías son similares a los tesauros. Pero por otro lado están las ontologías *heavyweight* (de peso pesado) que incluyen axiomas, y éstas superan el alcance de los tesauros.

El principal potencial de una Ontología es contar con reglas de inferencia que pueden facilitar la generación automática de conocimiento, propiedad de la que carecen los tesauros. Además al describir contenidos educativos difícilmente se recurre a los tesauros para hacer una descripción general; se podrían utilizar tesauros pero como un tipo de vocabulario para el correspondiente elemento de sus metadatos.

Por otra parte, el modelo de datos para contenidos educativos, usualmente utilizado LOM (*Learning Object Metadata*-Metadatos para Objetos de Aprendizaje) (IEEE, 2001), en la Web tradicional, carece de cualquier formalidad en el vocabulario del contenido de cada elemento del estándar, y aunque hay grupos categorizados de tipos de metadatos no se tiene clara la relación entre éstos; y además, no se hace explícito el tipo de convenciones que se siguen para hacer el llenado de los metadatos.

Algunas iniciativas han desarrollado LOM en RDF (RDF Resource Description Framework-Marco de Descripción de Recursos),<sup>5</sup> con la intención de comenzar a probar este tipo de metadatos en ambientes semánticos y echar a andar el desarrollo de aplicaciones educativas sobre la Web Semántica, basándose en ontologías y en RDF (Lee, Hua & Wang, 2007).

Por encima de los Tesauros y los Metadatos, las Ontologías garantizan consistencia en dominios de conocimiento y hacen explícitos sus

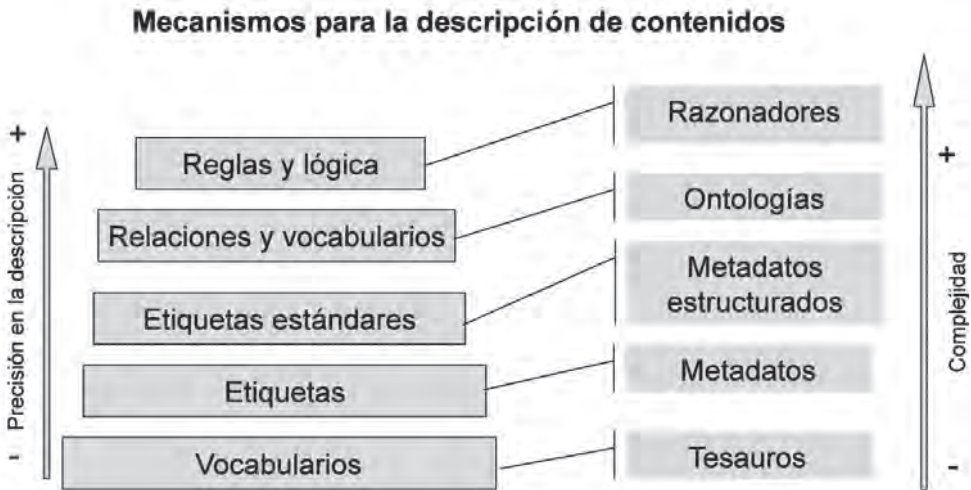
---

5 Busque "LOM" en <http://swoogle.umbc.edu/> y recuperará diferentes versiones y ejemplos.

vocabularios, pero principalmente lo que aportan son las reglas que les permiten a las máquinas interpretar las relaciones entre contenidos. Los Tesauros y los Metadatos estructuran contenidos; las Ontologías estructuran la semántica de los contenidos.

De esta manera la organización o descripción de contenidos puede llevarse a diferentes niveles de complejidad (figura 1); y por supuesto que una mayor precisión en la descripción de un recurso siempre implicaría mayor complejidad, y aunque los Tesauros han jugado un papel fundamental para el almacenamiento y la recuperación de la información ahora se auxilian de los Metadatos para tener estructuras más consistentes que amplían la descripción de un recurso, sobre todo si se hace uso de metadatos estándares. Las Ontologías tenían aún más estas estructuras y la gran ambición es lograr obtener razonadores automáticos que le simplifiquen el trabajo al humano y descarguen en las computadoras las tareas de organizar, pero sobre todo las de relacionar contenidos.

Figura 1. Niveles de descripción de contenidos



## DISEÑO DE ONTOLOGÍAS

El diseño de Ontologías no es una tarea simple, a diferencia de otros mecanismos de descripción, éstas requieren de un profundo conocimiento del dominio que se describe, ya que también hay que definir relaciones; y se pretende que las Ontologías puedan ser reutilizadas y que además puedan ser la base para generar más conocimiento. Dentro de un contexto de conceptualizaciones compartidas, Gruber (1993) dice que los criterios a considerar para el diseño de Ontologías son:

**Claridad.** La definición debe ser objetiva y comunicar de forma eficiente el significado. También debe ser independiente del contexto informático o social.

**Coherencia.** Las inferencias deben ser consistentes con las definiciones.

**Extensibilidad.** El diseño debe permitir diversos usos del vocabulario compartido. Principalmente, una definición debe servir para definir otra.

**Codificación mínima.** La conceptualización debe especificarse sin depender de símbolos de codificación, ya que los agentes que los pueden utilizar pueden ser llevados a cabo en diferentes estilos de representación.

**Mínimo compromiso ontológico.** Una Ontología debe tener el mínimo de casos especiales o pretensiones sobre lo que está representando, con la intención de que la propia Ontología pueda después ser más especificada para usos particulares.

Un factor de relevancia es que para que una ontología resulte de utilidad en la Web Semántica tiene que ser aceptada universalmente, y para que esto sea posible es necesario definirla colaborativamente por la comunidad en la que se quiera utilizar (García & Sicilia, s/f).

## METODOLOGÍAS PARA CONSTRUIR ONTOLOGÍAS

García (2005) define que para desarrollar una Ontología se requiere, de manera general, definir las clases que forman un dominio, organizar las clases en una jerarquía taxonómica, definir las propiedades

de cada clase e indicar las restricciones de sus valores, y asignarle valores a las propiedades para crear instancias. Sin embargo, no existe una metodología única para construir Ontologías dado que el diseño de ellas no es algo reciente como alternativa para la organización del conocimiento. Entre las más destacadas se encuentran las metodologías de Uschold & King, Methontology y Onto-knowledge, aunque también hay otras, como la IDEF5, Gruninger y Fox y The reference ontology, entre otras.

Es importante resaltar que aun cuando se siga una misma metodología para un determinado dominio, el resultado de cada grupo de diseño es diferente porque se puede interpretar y representar de diferente manera el conocimiento de cada individuo o de cada comunidad.

## LINGUAJES PARA LA DEFINICIÓN DE ONTOLOGÍAS

Para definir una Ontología se requiere de lenguajes que traducen lo que se anotó en el lenguaje lógico a lenguaje de máquina. Estos lenguajes permiten el manejo de taxonomías complejas y de relaciones lógicas, que además se estructuran mediante una semántica definida. Dentro de estos lenguajes, entre los más utilizados, se encuentran OIL (Fensel, 2000) y DAML (Hendler & MacGuinness, 2000), actualmente unidos en uno solo llamado DAML+OIL (Connolly, 2001), y OWL (Lacy, 2005).<sup>6</sup> Dado que la incorporación de las Ontologías a la web es relativamente reciente estos lenguajes se encuentran en constante evolución, pero pueden utilizarse sin que surjan problemas de inconsistencia por renovación de versiones.

## EDITORES DE ONTOLOGÍAS

Cuando no se quiera trabajar la programación de Ontologías directamente en un lenguaje de definición, se puede recurrir a software de autor para el diseño, el cual ofrece interfaces gráficas y amigables.

---

<sup>6</sup> Buscar todas estas en la WEB,C.E.

Algunos de estos productos<sup>7</sup> son Ontolingua y Ontosaurus que fueron las iniciales, y entre las más actuales se encuentran Protégé, WebODE, Ontoedit y Oiled, además de algunas otras como WebOnto como KADS22, Link Factory, OpenknowME y SymOntoX. (Buscar en la WEB,C.E.)

## REPOSITARIOS DE ONTOLOGÍAS

Se han realizado ya muchos desarrollos para el uso de Ontologías en diversos ámbitos y existen iniciativas que concentran esos desarrollos en un repositorio, usualmente de acceso libre, que ponen de forma abierta las Ontologías desarrolladas para ser reutilizadas por aplicaciones con objetivos similares. Incluso hay metodologías que proponen el desarrollo de una nueva Ontología a partir de una existente, y entonces se recurre a su búsqueda en algún repositorio. Dentro de los repositorios se encuentran SIMILE, DAMLontology library, OWL Ontology Library de Protégé, OWL SEC, Schema Web, RDFdata.or y SemWeb Central.

## MOTORES DE BÚSQUEDA

Para la búsqueda de páginas semánticas, documentos RDF y ontologías, existen motores que “escarban” en la red y localizan este tipo de documentos. Algunos Motores de búsqueda son Semantic Web Search, Onto Search y Swoogle. Dentro de este género también se pueden incluir los *crawlers*, robos que indizan las páginas web con tecnologías semánticas, siguiendo los enlaces que en ésta aparecen, algunos ejemplos son DAML crawler, RDF crawler, OCRA y Scutter.

## ONTOLOGÍAS Y CONTENIDOS EDUCATIVOS

Un contenido educativo se encuentra inmerso en un ambiente complejo: usuarios con intereses diferentes, aplicaciones con tecnologías

---

<sup>7</sup> Una revisión detallada de más de cien editores se puede encontrar en el trabajo de Denny (2002), disponible en [http://www.xml.com/2002/11/06/Ontology\\_Editor\\_Survey.html](http://www.xml.com/2002/11/06/Ontology_Editor_Survey.html)

heterogéneas, contextos variables y usos diversos. El desarrollo de Ontologías en este ámbito comienza a ser una necesidad, porque los sistemas tradicionales de clasificación y organización de la información y del conocimiento tienen limitantes en su diseño que les restringen el trabajo de tareas automáticas a los sistemas informáticos.

En general en el ámbito de la educación es fácil identificar que los autores utilizan diferentes terminologías para conceptos similares. En el campo de la enseñanza en línea esta situación es aún más crítica, y al tratar de utilizar términos en diversas plataformas o en objetos de aprendizaje que tienen movilidad entre aplicaciones el problema se hace más evidente. Por ello en los ambientes digitales educativos tiene sentido el uso de las Ontologías, específicamente para distinguir entre tres tipos de conocimiento: sobre contenido, pedagógico y estructural (Antoniou & Harmelen, 2004).

Una Ontología de contenido describe los conceptos básicos del dominio en el que se lleva a cabo el aprendizaje. Esto incluye las relaciones entre los conceptos y algunas propiedades básicas. Las Ontologías de contenido pueden utilizar relaciones para identificar sinónimos y abreviaturas.

Las Ontologías pedagógicas deben describir los recursos pedagógicos como: tutoriales, ejemplos, ejercicios, evaluaciones, etcétera.

Las Ontologías estructurales se proponen definir una estructura lógica entre los contenidos, según las relaciones de jerarquía, secuencia o navegación, como: previo, siguiente, requiere, es parte de, está basado en, etcétera.

Moral y Cernea (2005) dicen que para los Objetos de Aprendizaje, son necesarias Ontologías de dominio, del alumno, de las teorías de aprendizaje y del diseño instruccional, de la estructura y de los mismos metadatos, lo que resumen como: dominio, estructura, pedagogía y metadatos.

## CONCLUSIONES

Es notable la existencia de una constante búsqueda por encontrar mecanismos más precisos para organizar y describir el conocimiento, la información y ahora, más específicamente, los recursos y los conteni-



dos educativos, ya que las aplicaciones requieren descripciones más completas y precisas para poder procesarlos de forma más efectiva.

La complejidad aumenta con la precisión y esto demanda un mayor análisis de parte de los diseñadores y un conocimiento más profundo del dominio que se esté estudiando.

El uso de las Ontologías dentro del ámbito educativo tiene un gran potencial, pero el nivel de desarrollo es incipiente y deben comenzarse a investigar estas líneas de aplicación a fin de dotar a los contenidos educativos de información que los haga más interoperables en sus diferentes perfiles de uso y de sus usuarios. En este trabajo se ha mostrado que ya existen tecnologías, metodologías, herramientas y repositorios para echar a andar aplicaciones específicas, y no debe perderse de vista que para desarrollar Ontologías se requiere de un buen entendimiento con la tecnología y con los conceptos documentales formales.

Para lograr una mejor integración de los contenidos educativos en los ambientes de aprendizaje en línea debe comenzarse a trabajar en el estudio de metodologías específicas para este sector, partiendo de la base de que se deben desarrollar Ontologías de contenido que describan los conceptos del dominio de las diferentes áreas de aprendizaje. Lo que se pretende no es estandarizar el conocimiento, sino las prácticas y las tecnologías que permitan que los contenidos en línea sean más fácilmente manipulados por los sistemas informáticos, y se cuente entonces con aplicaciones más eficientes.

## REFERENCIAS

Antoniou, G. & Harmelen van, F. (2004), *A Semantic Web Primer*, Massachusetts Institute of Technology.

Berners-Lee, T., Hendler, J., Lassila, O., "The semantic web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities", en *Scientific American*, Mayo, 2001.

***Primer Simposio Internacional sobre Organización...***

- Castells, P. & Saiz, F., “Ontologías para la web semántica”, en *La web semántica: tecnologías y aplicaciones*, 2003.
- Connolly, D. *et al.* (2001), “DAML+OIL Reference Description”, disponible en <http://www.w3.org/TR/daml+oil-reference>.
- Fensel, D. *et al.* (2000), OIL in a Nutshell, disponible en <http://www.cs.vu.nl/~ontoknow/oil/download/oilnutshell.pdf>.
- García, E., *Aplicación de las ontologías para la representación del conocimiento*, ISKO, Capítulo español Barcelona, 2005.
- García, A. (2004), “Instrumentos de representación del conocimiento: tesauros versus ontologías”, en *Anales de documentación*, núm. 7.
- García, E., Sicilia, M.A., (s/f) “Una propuesta para la búsqueda semántica de recursos web de nuevas tecnologías aplicadas a la educación”, en línea <http://tecnologiaedu.us.es/ticsxxi/comunic/egb-masu.htm>
- Gruber, T. R. (1993), “Toward Principles for the Design of Ontologies Used for Knowledge Sharing”, en Guarino, N., *Formal Ontology in Conceptual Analysis and Knowledge Representation*, Kluwer Academic Press, Boston.
- Hendler, J. & MacGuinnes, D.L. (2000), “The Darpa Agent Markup Language”, en *IEEE Intelligent Systems*, 16(6).
- IEEE (2001), IEEE Learning Technology Standards Committee (LTSC), IEEE P1484.12, Learning Object Metadata Working Group, [http://ltsc.ieee.org/wg12/s\\_p.html/](http://ltsc.ieee.org/wg12/s_p.html/)
- Lacy, L (2005), *OWL: representing information using the Web Ontology Language*, Trafford. Victoria, Canadá.

- Lee, M., Hua K., Wang, T. (2007), "A practical ontology query expansion algorithm for semantic-aware learning objects retrieval", en *Science Direct, Computers & Education*, article in press, disponible en <http://simile.mit.edu/papers/iswc05.pdf>.
- Moral, M. E. & Cernea D. A. (2005), "Diseñando Objetos de Aprendizaje como facilitadores de la construcción del conocimiento", en SPDECE 2005, Universitat Oberta de Catalunya, <http://www.uoc.edu/symposia/spdece05/pdf/ID16.pdf>, consultado 7 de abril 2007.
- Uschold, M. & Gruninger, M. (1996), "Ontologies: principles, methods and applications", en *Knowledge Engineering Review*, 11(2).



**SECCIÓN V**  
**Representación y recuperación de información**



# La terminología como método para representar y organizar dominios multidimensionales

MARÍA JOSÉ LÓPEZ HUERTAS  
*Universidad de Granada, España*

## 1. INTRODUCCIÓN

El modelo lingüístico ha sido ampliamente utilizado para estudiar y construir herramientas destinadas a la indización y a la recuperación de la información, al menos desde los años sesenta del pasado siglo, como bien pone de manifiesto la larga lista de autores, muchos pertenecientes a la escuela francesa, que abordan el estudio de los lenguajes documentales desde esta perspectiva. Como ejemplo, baste aquí la cita de las obras de Coyaud (1966) y de Hutchins (1975), referentes incuestionables, por ser autores de unas de las contribuciones más completas y tempranas de la aplicación del modelo lingüístico al análisis y construcción de los lenguajes documentales.

Cuando la terminología se configura como especialidad independizada de la Lingüística, se empiezan a ver las grandes concomitancias entre ésta y los lenguajes de indización y recuperación. El objeto de análisis y punto de partida de estudio en ambos casos es el lenguaje especializado. Ésta coincidencia, entre otras, hizo que pronto proliferaran estudios de estas herramientas con el modelo terminológico como referente.

No cabe duda de que la llegada de teorías generales para el estudio e investigación de la Biblioteconomía, la Documentación y las Ciencias de la Información, como es el análisis del dominio propuesto por Hjørland y Albretchsen (1995) que propugnan como uno de los métodos para llevar a cabo dicho análisis, entre otros, (Hjørland, 2002) ha dado un espaldarazo definitivo al modelo terminológico para el estudio de las dinámicas del conocimiento especializado, con la finalidad de ser representado y organizado conceptualmente para la recuperación de la información.

Por otra parte, el desarrollo de Internet ha reactivado en gran medida los retos que ha representado y representa el conocimiento multidimensional<sup>1</sup> para los sistemas de representación y organización del conocimiento.

El estudio de la terminología, para los fines señalados, se muestra aún más interesante y necesario cuando se trata de dominios interdisciplinarios y transdisciplinarios por la indefinición que caracteriza estos espacios, en mayor o menor medida, tanto en lo terminológico como en lo referente a su representación conceptual y por la carencia de modelos que expliquen las dinámicas de integración de espacios disciplinares ajenos dentro de la interdisciplina así como la producción de terminología expresiva de la interdisciplina, exclusivamente. Hay diferencias profundas entre los dominios disciplinares e interdisciplinares, siendo estos últimos más difíciles de representar y organizar por su complejidad. Algunas de sus características son las siguientes:

- Terminología no consensuada e inestable
- Terminología de aluvión (Interdisciplinas de reciente creación)
- Límites conceptuales no siempre bien definidos
- Fronteras epistemológicas confusas

---

1 Se podría llamar conocimiento complejo, pero son varios los autores que coinciden en que el pensamiento complejo aspira al conocimiento multidimensional porque se sabe que el conocimiento complejo es imposible, ya que uno de los axiomas de la complejidad es la imposibilidad de una omnisciencia (Morin, E. Sobre la interdisciplinariedad. <http://www.pensamientocomplejo.com.ar/docs/files/morin%5Fsobre%5Fla%5Finterdisciplinaridad%2Epdf>. (Visitado el 29 de enero de 2007))



- Falta de taxonomías o clasificaciones científicas
- Dinámicas conceptuales sin definir
- Se desconoce la interacción entre las especialidades que constituyen la interdisciplina
- Falta de modelos teóricos para abordar estos espacios

No cabe duda de que el nuevo conocimiento se presenta y se percibe como un problema, en primer lugar, por las dinámicas estructurales y terminológicas que le son propias y que hace que sean espacios calificados por algunos como inestables, impredecibles (Williamson, 1998), tendentes al solapamiento de discursos y presentando una dificultad notable para determinar la interacción entre las disciplinas de origen en el campo disciplinar (McIlwaine, 2000). Varios autores han estudiado la inadecuación entre las herramientas de indización y recuperación al uso del nuevo conocimiento que se traduce en dificultades para la indización y, sobre todo, para la recuperación de la información (Denda, 2005; Williamson, 1998 y McIlwaine, 2000).

Hasta hace poco, las propuestas que se hacían para abordar este asunto eran bastante conservadoras y, en general, no iban enfocadas a la esencia del problema planteado, que no es otro que la necesidad de partir de planteamientos diferentes de los hasta ahora establecidos para la organización del conocimiento, uno de los cuales ha venido siendo el principio de estructuración disciplinar.

La situación actual no es demasiado diferente a la que existía unas décadas atrás porque no hay un modelo más o menos aceptado para la organización de este nuevo conocimiento. No obstante, se observa un interés renovado por estas cuestiones en los últimos años que se han materializado en varias contribuciones de especialistas que proponen alternativas esenciales, no conservadoras, para construir estructuras donde el conocimiento multidimensional pueda representarse de forma adecuada a su dinámica y naturaleza. Una de las propuestas es la sugerida por Claudio Gnoli (2006) que tiene como objetivo sentar la bases para la creación de una clasificación universal no disciplinar facetada. Para ello, se vale de la reutilización de conceptos tales como “fenómeno”, faceta y niveles integradores, de amplia tradición en la investigación de la clasificación, y de la lógica predicativa. La clasificación que propone se basa en “fenómenos del mundo real en vez

de en disciplinas, los fenómenos se pueden ordenar por medio de los niveles integradores a los que pertenecen de los más simples a los más complejos y las marcas de clase se podrían obtener por medio de la combinación de las notaciones constantes de cada elemento que compone el fenómeno” (Gnoli y Poli, 2004). Las facetas estarían referidas a clases de fenómenos en lugar de a clases de disciplinas que se considerarían como clases especiales de fenómenos.

Otras aproximaciones se alejan de planteamientos universales y proponen soluciones a la interdisciplinariedad y transdisciplinariedad a partir del análisis del dominio, es decir a partir de análisis contextuales, entendiéndose que hay que tener en cuenta las dinámicas y circunstancias de los dominios concretos para hacer propuestas ajustadas a la realidad e identidad de cada uno de ellos como paso previo a la construcción de sistemas de recuperación de información. Por tanto, profundizar en el estudio y en la dinámica de estos dominios aparece como una actividad previa a la representación y organización del conocimiento multidimensional que va a garantizar la amigabilidad y eficacia del producto final. Es una aproximación macro inductiva que otorga niveles de calidad indiscutibles a las estructuras resultantes.

Como decía en un principio, al ser reconocido el terminológico como un método de análisis del dominio, junto con la indización de documentos especializados, se observa una reactivación de las investigaciones que abordan esta aproximación para el diseño y construcción de herramientas de representación y organización del conocimiento para la recuperación de la información. En este trabajo se abordarán algunas propuestas recientes que enfocan el problema planteado desde un análisis contextual de los dominios del conocimiento científico.

## 2. LA TERMINOLOGÍA COMO MÉTODO DE ANÁLISIS PARA EL ESTUDIO Y LA EVALUACIÓN DE HERRAMIENTAS DE INDIZACIÓN Y RECUPERACIÓN DE INFORMACIÓN

El estudio de sistemas de indización es, en sí mismo, un método de análisis del dominio (Hjørland, 2002) porque dichos sistemas son representaciones ya realizadas sobre la disciplina o especialidad que se quiere estudiar. Profundizar en su conocimiento es imprescin-

dible como un dato más a tener en cuenta en este proceso de desvelar las características epistemológicas del campo temático. En este caso, se han analizado tesauros especializados en Estudios de Género en uso como ejemplo de dominio multidimensional.

Se utilizaron cuatro tesauros sobre Género: el European Women Thesaurus (International Information Centre and Archives for the Women's Movement, 1998), el Tesauro Mujer (Instituto de la Mujer de Madrid, 2002), Thesaurus d'Historia Social de la Dona (Sebastiá i Salat, Montserrat, 1988), y el Tesauro para Estudios de Género e sobre Mulheres de Brasil (Bruschini, C., D. Ardaillon y S. Unbehaum, 1998) con la finalidad de identificar posibles modelos para representar y organizar el dominio y conocer otras características de este espacio temático.

## 2.1. Metodología

Se usó la terminología utilizada en las fuentes consultadas y las estructuras conceptuales del dominio propuestas por los tesauros así como los nombres empleados para las categorías seleccionadas.

*A) El estudio terminológico* se realizó de acuerdo con siguiente proceso: 1) Recopilación exhaustiva de los descriptores y no descriptores incluidos en los tesauros analizados. Se hace notar que se excluyó del análisis terminológico el tesauro brasileño, por no haber obtenido de sus autoras la versión digital del mismo. 2) Comparación cuantitativa y porcentual, así como la evaluación del número de descriptores y no descriptores aportados al dominio por cada tesauro, estimándose además el cociente de equivalencia. 3) Determinación del nivel de solapamiento.

*B) El análisis estructural del dominio* se realizó como sigue: 1) Identificación de las clases o áreas temáticas principales ofrecidas por los cuatro tesauros, entendiendo por tales sólo las establecidas en el primer nivel de subdivisión. 2) Realización de un cuadro comparativo de las clases o áreas temáticas principales, a efectos de determinar el grado de solapamiento. Se estableció que cada co-ocurrencia de

una clase correspondería a un impacto 3) Identificación de las áreas temáticas con más impacto en la interdisciplina para reconstruir su dinámica estructural (López-Huertas, M.J., Barité, M. Y Torres, I., 2004; López-Huertas, M.J., 2006a).

Los resultados obtenidos evidenciaron que la terminología usada por los tres tesauros tiene un bajo índice de solapamiento ya que sólo 130 términos co-ocurren de un total de 4.057. La tabla 1 muestra este solapamiento por tesauros, por términos totales, descriptores y no descriptores.

Tabla 1. Solapamiento de los términos incluidos en los tesauros

	Co-ocurrencia de términos	% Co-oc..	Descript. Co-oc.	% descript. Co-oc.	No descript. Co-oc.	% No Descript. Co-oc.
Tesaurus de la Dona	130/714	18,20	91/624	14,58	1/90	0,01
Tesaurus Mujer	130/1256	10,35	91/799	11,38	1/457	0,00
European Thesaurus	130/2087	6,22	91/1498	6,07	1/589	0,00
Tesaurus de la Dona	130/714	18,20	91/624	14,58	1/90	0,01
Tesaurus Mujer	130/714	10,35	91/799	11,38	1/457	0,00
European Thesaurus	130/2087	6,22	91/1498	6,07	1/589	0,00

Si tomamos la totalidad del vocabulario teniendo en cuenta el factor limitante del número total de términos que aporta el tesaurus al total de vocabulario y vemos el solapamiento total que se produce tenemos los resultados de muy baja coincidencia terminológica que muestra el gráfico 2

Tabla 2. Porcentajes totales de solapamiento terminológico

	Total ocurrencias	% co-occ. simples	% co-ocurr. con FL
Términos	4057	9,61	18,20
Descriptores	2921	9,34	14,58
No Descriptores	1136	0,00	1,11

Con respecto al análisis estructural, se observó que había 34 diferentes áreas temáticas considerando sólo el primer nivel de la jerarquía lo que revela una diversificación alta. También se detectaron una diversidad notable de los modelos estructurales en los tesauros (4 en T. Dona, 9 en T. Género, 17 en T. Mujer y 20 en Tesauro Europeo) Por otro lado, se analizó la co-ocurrencia de las clases para descubrir su impacto en la interdisciplina, obteniéndose como resultado que ninguna de las 34 eran compartidas por todos los tesauros. Sólo ocho áreas co-ocurren en el primer nivel en tres tesauros (23%), otras ocho áreas co-ocurren en dos tesauros y 18 áreas no tienen co-ocurrencia. Se observó igualmente que había clases dentro del grupo de más impacto, como Filosofía y Medios de Comunicación, que no tienen una relación nuclear con Género y otras temáticas, en cambio, como Familia, Sexualidad y Estudios de Género que están más cercanas conceptualmente hablando están fuera de este grupo de mayor impacto en el dominio interdisciplinar.

Las conclusiones de esta aproximación al análisis del dominio de Género son las siguientes:

1. Una diversidad terminológica excesiva que permite hablar de verdadera dispersión lo que induce a pensar que muchos de esos términos podrían no ser significativos para el dominio. La calificamos como terminología de aluvión, propia de una interdisciplina emergente, que necesariamente confiere inestabilidad a la terminología del dominio. Ésta se expresa en una rápida evolución formal y conceptual que se desarrolla paralelamente a la maduración de la interdisciplina. Por ejemplo, se produce la siguiente secuencia cronológica muy expresiva teniendo en cuenta que es el nombre de la interdisciplina: Estudios de la Mujer, Estudios de las Mujeres, Estudios de las Relaciones de Género, Estudios de Género, Estudios de las Mujeres.
2. Falta de estandarización de los términos del área.
3. Escaso consenso sobre qué terminología pertenece al dominio temático.

4. Numerosos términos procedentes de otras disciplinas, que superan con creces a los generados desde la actividad nuclear interdisciplinaria, que se han incorporado a los tesauros sin aparente reformulación.
5. Problemas derivados de incorporación de préstamos, la mayoría anglicismos, que no tiene equivalentes terminológicas en español: butche-femme, housewifization, co-mothers, etc.
6. Límites epistemológicos del dominio difusos.

### 3. LA TERMINOLOGÍA COMO MÉTODO DE CONOCIMIENTO DE UN DOMINIO MULTIDIMENSIONAL EMERGENTE. LAS CIENCIAS DE LA INFORMACIÓN.

Es difícil identificar la terminología que realmente es representativa de una interdisciplina, sobre todo cuando es un dominio emergente, aunque es muy importante tener este conocimiento para analizar el dominio. Este es el supuesto que se plantean Kobashi, Smit y Tálamo (2002) cuando abordan el estudio de las Ciencias de la Información partiendo de la base de la terminología que utiliza, pretendiendo no solo desvelar sus posibles problemas terminológicos y su estructura conceptual sino determinar su estatus epistemológico a partir de la terminología que utiliza el dominio.

Las fuentes terminológicas utilizadas fueron artículos, diccionarios y enciclopedias especializadas y la terminología de los estándares ISO. La metodología empleada les permitió identificar qué términos eran centrales y cuáles periféricos. Es interesante la distinción que hacen entre términos centrales y periféricos porque es una dinámica que se repite en los espacios interdisciplinarios (López-Huertas, Barité y Torres, 2005). Igualmente interesante sería comparar estos comportamientos terminológicos en interdisciplinas diferentes para poder identificar posibles patrones comunes entre ellas al igual que las razones de sus diferencias, ya que podría llevar a identificar dinámicas interdisciplinarias universales.

Para esos términos se buscaron definiciones que se sintetizaron, aunque no se aclara cómo se procesaron esas definiciones o si sirvie-

ron como fuente textual para hacer un análisis del discurso sobre la terminología seleccionada. Se pudieron así identificar cuatro grupos conceptuales: nociones asentadas, establecidas a lo largo del tiempo, nociones semielaboradas, nociones de experiencia empírica común y nociones confusas que se corresponden con términos exportados desde otras áreas.

El estudio de los términos permitió concluir que los términos centrales provenían en gran medida del área de la Biblioteconomía. El resto de los términos procedentes de otras áreas se han sometido a un proceso de adaptación en el que a veces mantienen su sentido originario, aunque con más frecuencia los términos se adoptaron cuando su estructura conceptual original fue abandonada

A partir de su estudio terminológico, las autoras también han podido recomponer los temas que configuran las Ciencias de la Información, a saber: Biblioteconomía, Lógica, Administración, Lingüística, Ciencias de la Computación, Sociología, Comunicación y Ciencias Cognitivas.

Es una interesante aproximación para conocer el dominio, entre otras cosas porque descubre nuevo conocimiento ante una apariencia que de otra forma hubiera sido muy difícil detectar y ofrece explicaciones fundadas para el aparente caos que, en mayor o menos medida, parece caracterizar a las interdisciplinas. Hubiera sido muy interesante que las autoras dieran cifras para cuantificar sus hallazgos porque se hubiera conseguido una mayor visualización y claridad.

#### 4. LA TERMINOLOGÍA DE INDIZACIÓN DE DOCUMENTOS PRIMARIOS ESPECIALIZADOS COMO MÉTODO DE ANÁLISIS DEL DOMINIO

Se ha elegido los Estudios de Género como campo de pruebas para realizar esta investigación.

El análisis del dominio desde la perspectiva de la indización de los documentos especializados ha resultado ser, al menos en este dominio, muy interesante y muy enriquecedor. Los documentos publicados sobre una especialidad muestran el estado del conocimiento de la misma (Hjørland, 2002), dan noticias de la epistemología de dicha

especialidad, sobre todo si se consideran en su totalidad o lo más cercano a ella y no en forma de muestra representativa del dominio. El estudio del conjunto de la terminología extraída de la indización de los contenidos puede dar noticia de la dinámica terminológica del dominio según se expresa en las fuentes documentales primarias fundamentalmente, de los diferentes discursos implicados en la interdisciplina, de los problemas posibles de deriva terminológica existentes y de la capacidad expresiva del espacio interdisciplinar para generar o no un lenguaje inequívoco y exclusivamente representativo del mismo.

Se han identificado las publicaciones realizadas en dicho campo entre 1990 y 2005. Se analizaron 600 documentos primarios (monografías, publicaciones periódicas, actas de congresos e informes de investigación y de carácter sociopolítico), la totalidad de lo publicado en Uruguay, hasta ahora identificado, para el período indicado. Se extrajo terminología relevante de los contenidos de estos fondos por asignación de descriptores libres. Los términos se seleccionaron de los títulos, resúmenes y encabezamientos de capítulos.

Representación terminológica del dominio. Se observó que una parte no mayoritaria de la terminología nace dentro de la interdisciplina, es decir, no procede de ninguna de las disciplinas de base u origen y surge para denominar objetos o fenómenos que sólo producen o se verifican desde una óptica interdisciplinaria. Otra terminología se origina en una disciplina de base y deriva hacia la interdisciplina. En este proceso de deriva, los términos han sido conservados con la misma forma y significado que tenían en la disciplina de origen la mayoría de las veces constatándose un fenómeno de duplicidad terminológica.

El conjunto terminológico analizado muestra la existencia de dos dinámicas principales que están íntimamente relacionadas con la procedencia disciplinar originaria de los términos. Un primer grupo que representa los términos creados desde la propia interdisciplina que constituirían su núcleo y un segundo grupo que está formado por la terminología incorporada como producto de la interacción de las distintas disciplinas y temáticas que integran el dominio interdisciplinar que sería la más numerosa y presenta una dinámica diferente a la del grupo anterior.

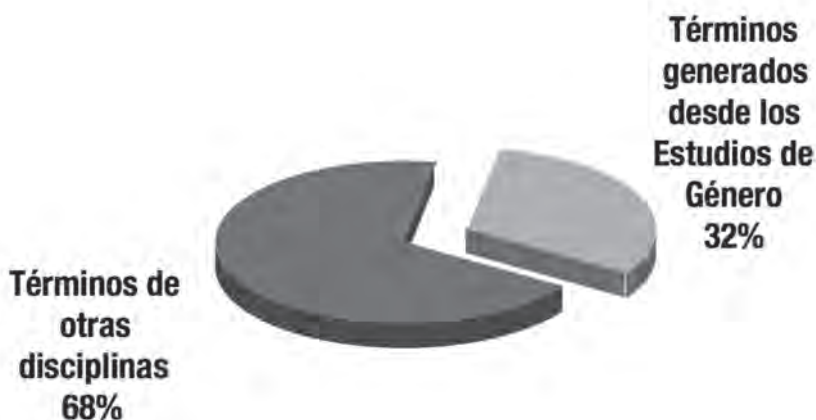


La composición temática de este dominio interdisciplinar afecta a la terminología que lo representa, de forma que cada grupo tiene, a su vez, una dinámica diferente como se explica a continuación.

1) Los términos que pertenecen al primer grupo, a partir de los datos obtenidos hasta ahora y teniendo en cuenta que es una investigación en curso, representan el 32% del total de 460 descriptores y tienen un comportamiento bastante unívoco, de manera que su pertenencia a la interdisciplina es incuestionable, aunque también participa de otras dinámicas generales del dominio como es la inestabilidad, el elevado número de términos en proceso de consolidación conceptual y formal y otras derivas terminológicas que se pueden producir en el caso de que algunos de esos términos interdisciplinares tengan su origen en otra área temática.

2) La terminología perteneciente al 2º grupo es la más numerosa, representando el 68% del total de 460 descriptores. Ha sido incorporada como producto de la interacción de las distintas disciplinas y temáticas con el dominio interdisciplinar, según se muestra en la Figura 1. Los términos pertenecientes a este grupo tienen un comportamiento diferente, distinguiéndose, por ahora, dos tendencias:

Fig. 1 Procedencia de la terminología



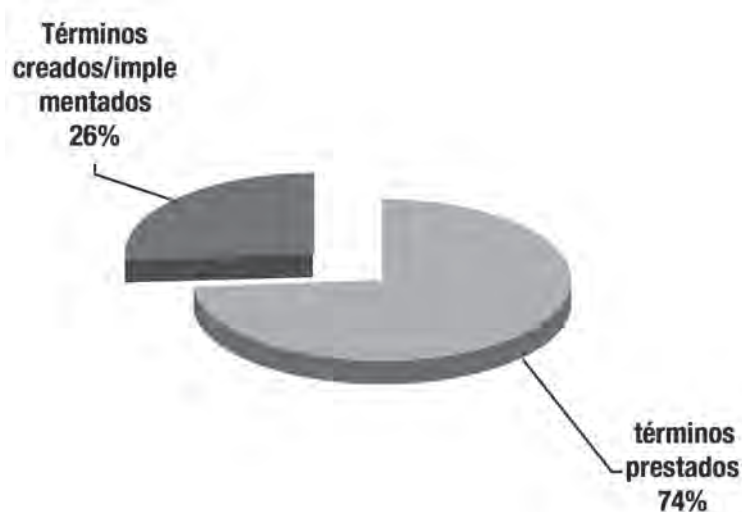
a) Se ha observado que los documentos que se refieren a cualquier tema (higiene, trabajo, educación, etc.) desde la perspectiva de Género muestran una ausencia de lenguaje propio para incorporar esas temáticas a la interdisciplina, haciendo difícil rescatar la marca de Género en las indizaciones y trasladarla posteriormente a los sistemas de recuperación de información. En un elevadísimo porcentaje, los especialistas sortean este problema añadiendo una expresión inequívoca de la interdisciplina, como Género, Mujeres, etc., a la disciplina o temática que tratan (Trabajo y Mujeres), estableciendo así los límites del discurso de Género. Este hecho se aprecia muy bien en los títulos de los documentos donde expresiones como Mujer y política, salarios de las mujeres, Derechos sociales de las mujeres, Salud y Género, etc., son lo habitual. No se genera, por tanto, terminología propia para estos casos y los especialistas aparentemente se sienten satisfechos con la solución mencionada anteriormente.

Esta forma de resolver el problema de ambigüedad que se genera puede ser efectivo a este nivel de comunicación primaria, pero las palabras que se utilizan para marcar el discurso de género en los documentos primarios no sirven en un sistema de indización para estos propósitos porque las expresiones género, mujeres, etc. en este contexto no tienen significado en el dominio de género aisladas de las palabras trabajo, política, economía, etc. De manera que las expresiones de indización resultantes, que serían sólo trabajo, política, economía, etcétera, no expresan distinción alguna de género y aparentemente mantienen la forma y el significado que tenían en las disciplinas de origen. Por tanto, este grupo es un generador potencial de problemas de ambigüedad en la terminología del conjunto del dominio de Género. Representa el 74% del total de los términos que la interdisciplina toma prestados de otros dominios.

b) Se ha identificado un grupo de términos creados por la interacción de la perspectiva de Género con otras disciplinas. Estos representan conceptos nuevos en esa disciplina (trabajo remunerado, techo de cristal, salario del ama de casa, discriminación vertical, etc.) o expresiones que representan conceptos que ya existen pero que ganan importancia por influencia de la perspectiva de género (violencia doméstica, acoso sexual, violencia contra las mujeres, etc.). Este con-

junto representa el 26% del total de términos pertenecientes a otras disciplinas. La composición del grupo 2 se representa en la Figura 2

Fig. 2 Dinámica terminológica del 2º grupo (Interacción de otras disciplinas con Género)



*Estructura temática del dominio.* Si se analiza la terminología que se relaciona con disciplinas externas a los Estudios de Género y se procede a una categorización de la misma, se puede obtener un conjunto disciplinar y temático representativo del dominio que va a poder discriminarse en función del peso de la terminología que lo representa, que viene determinado por la cantidad de términos identificados en los documentos para cada área temática. Siguiendo el método indicado se obtienen los siguientes resultados: Disciplinas y temáticas tales como Derechos/Derecho, Política, Costumbres, Familia/Sociedad y Salud son claramente significativas para la formación del mapa interdisciplinar, mientras que otros temas tales como Psicología, Cultura, Administración, Cuerpo/Imagen, y Otros (Demografía, Religión y Grupos) son menos importantes en la configuración temática de los Estudios de Género. Una representación de los temas y su proporcionalidad se muestra en la Figura 3. En la Figura 4, puede ver-

se la composición interna de cada una de las grandes categorías antes mencionadas en la Figura 3

Para tener una imagen más nítida de la composición temática no sólo de las disciplinas externas sino también de su relación con la terminología generada desde la interdisciplina misma, se han representado los temas con categorías más generales a las se les ha unido Género, tal y como se muestra en la Figura 5.

Fig. 3 Mapa temático proporcional de las disciplinas interactuantes con los Estudios de Género

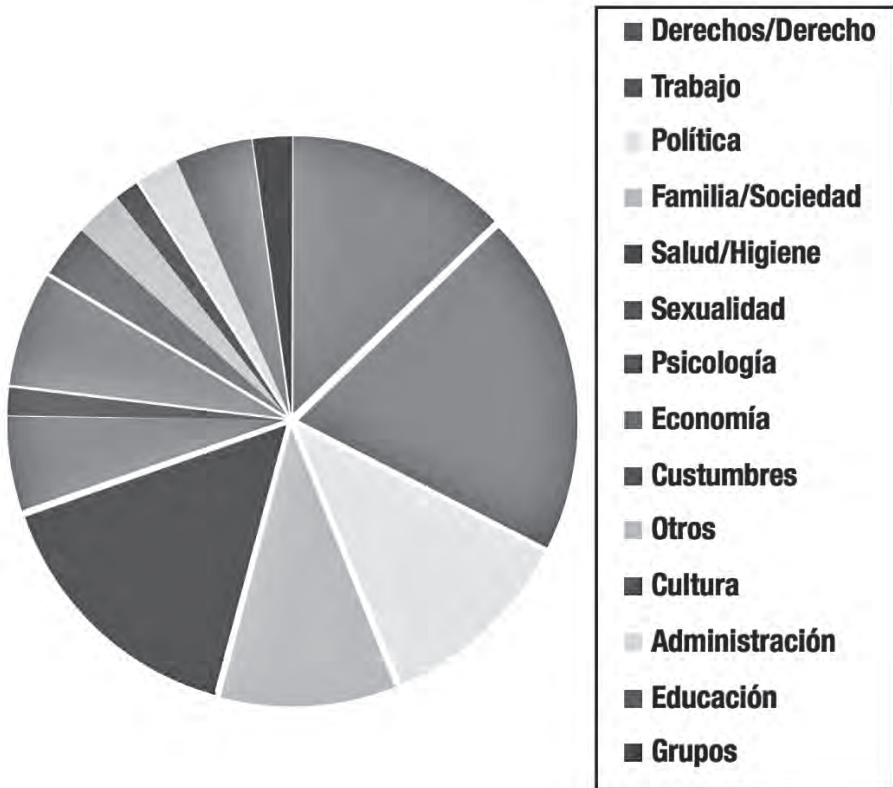


Fig. 4 Composición interna y representatividad de las grandes categorías externas

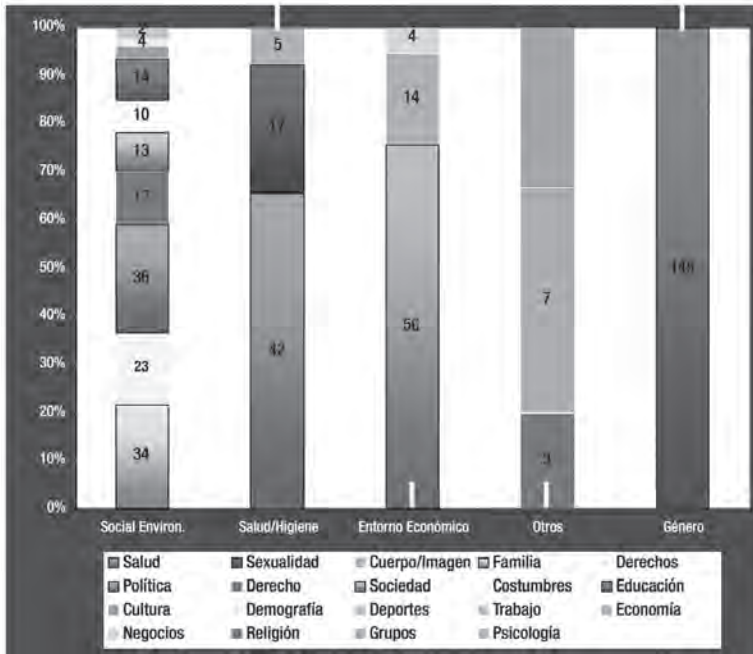
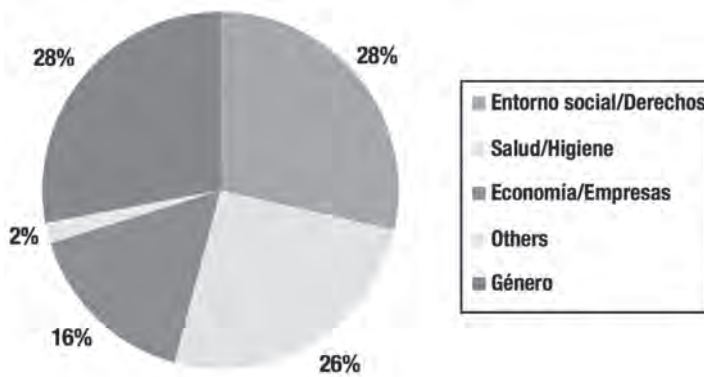


Fig. 5 Composición y representatividad de las principales áreas temáticas en Estudios de Género



Estos resultados tienen el interés de proporcionar una panorámica de las disciplinas que conforman los Estudios de Género basada en datos extraídos del análisis de las publicaciones que, como se dijo anteriormente, reflejan el estado del conocimiento de dicha temática. No sólo se ha podido saber cuál es el mapa temático de los Estudios de Género sino también, y esto es muy relevante, la importancia o representatividad que cada disciplina o temática tiene para el dominio interdisciplinar estudiado. La determinación de la profundidad de tratamiento de cada disciplina es importante para obtener un conjunto temático armónicamente representado de acuerdo con el estado del conocimiento real de la interdisciplina. En algunos sistemas de indexación hasta la fecha, por ejemplo en tesauros, se ha observado una gran anarquía y dispersión al compararlos entre sí, a la hora de representar temáticamente los Estudios de Género. Era evidente una falta de modelo ((López-Huertas, Torres y Barité, 2004; López-Huertas, Barité y Torres, 2005); López-Huertas, 2006b, 2006c). Asimismo, los tesauros (López-Huertas y Barité, 2004) y los buscadores de Internet adolecen de un problema similar (López-Huertas y Barité, 2002; Marcella, 2002). Por todo ello, aproximaciones al problema como las referidas en este apartado, ponen las bases para que se pueda ir diseñando un modelo de representación y organización de las interdisciplinas.

De los resultados anteriormente enumerados, pueden extraerse una serie de conclusiones como son:

1. Es importante identificar y analizar la representación terminológica de los dominios interdisciplinarios porque estos tienen distintas dinámicas entre sí y, por supuesto, muy diferentes dinámicas si los comparamos con los espacios disciplinares.

2. La terminología procesada de la forma descrita en los trabajos citados es una pieza importante en el proceso de representación y organización de los dominios interdisciplinarios.

3. La dinámica terminológica da información importante sobre la situación epistemológica del dominio al tiempo que revela las debilidades y fortalezas que pueda tener

4. El estudio del comportamiento terminológico podría decir si estamos hablando de una inter o una transdisciplina. Los Estudios de Género muestran una conducta transdisciplinar no tanto por su implicación con muchas especialidades sino por la dinámica de estas interacciones, ya que se ha observado que ha penetrado en varias disciplinas alterando sus estructuras de conocimiento mediante la creación de nuevos conceptos y términos, modificando de esta forma su estructura científica tradicional. Estas modificaciones así como la creación de un nuevo lenguaje para describir la realidad se podría identificar con la transdisciplinariedad.

## CONCLUSIONES FINALES

Los resultados de las investigaciones referidas en el presente estudio ponen de manifiesto que la utilización del método terminológico para conocer las dinámicas terminológicas y conceptuales de las disciplinas y especialidades científicas tiene muchas más posibilidades de las reconocidas tradicionalmente, restringidas a estudios principalmente descriptivos. La terminología, tratada de la forma aquí descrita, se muestra como un método adecuado para descubrir nuevo conocimiento, aparentemente oculto, en dominios científicos. Este hecho tiene un gran valor ya que se erige así en una forma de evaluación del estado del conocimiento en el dominio estudiado, de su situación epistemológica e incluso de las dinámicas investigadoras que se generan y de la progresión y evolución científica del campo estudiado.

Cuando el dominio está relacionado con el conocimiento multidimensional, estas características cobran una importancia casi decisiva porque, como se ha venido diciendo, estos espacios científicos se caracterizan por una indefinición en sus dinámicas internas tanto terminológicas como conceptuales y están inmersos en un proceso de evolución muy rápida. En casos así, un método, como el terminológico, que permita esclarecer qué pasa dentro de una inter o transdisciplina a estos niveles es de una gran utilidad y una garantía para la posterior construcción de herramientas y sistemas de indización y recuperación de la información.

## REFERENCIAS

- Bruschini, C., D. Ardaillon T S. Unbehaum (1998). *Tesouro para Estudos de Gênero e sobre Mulheres*. Sao Paulo, Fundação Carlos Chagas.
- Coyaud, M. (1966). Introduction a l'Étude des Langages Documentaires. Paris, Librairie C. Keincksieck, 1966
- Hutchins, W. J. (1975). *Languages of Indexing and Classification. A linguistic study of structures and functions*, Stevenage, Peter Peregrinus, 1975
- Gnoli, C (2006). The meaning of facets in non-disciplinary classifications. En: *Knowledge Organization for a Global Learning Society. Proceedings of the ninth ISKO International Conference*. Edited by Gerhards Budin, Christian Swertz y Konstantin Mitcusch. Würzburg, Ergon, 2006
- Gnoli, C. Y Poli, S. (2004). Levels of reality and levels of representation. *Knowledge Organization*, vol. 1 (3), p. 151-160
- Denda, Kayo (2005). Beyond subject headings. A structured retrieval tool for interdisciplinary fields. *Library Resources and Technical Services*, 49 (4), pp. 266-275
- Hjørland, B. & Albrechtsen, H. (1995). Toward A New Horizon in Information Science: Domain Analysis. *Journal of the American Society for Information Science*, 1995, 46(6), 400-425.
- Hjørland, B. (2002). Domain analysis in information science. Eleven approaches - traditional as well as innovative. *Journal of Documentation*, 58 (4), 422-462
- IIAV (International Information Centre and Archives for the Women's Movement) (1998). *European Women's Thesaurus*. Amsterdam, IIAV.



Instituto De La Mujer (España). Centro de Documentación (2002). *Tesaurus "Mujer"*. Madrid: Ministerio de Asuntos Sociales, Instituto de la Mujer, 1999. 6ª ed. Rev. 2002 en: [www.mtas.es/mujer/t2002.doc](http://www.mtas.es/mujer/t2002.doc). Último ingreso: 10-02-2005.

Kobashi, N.Y., Smit, J.W. Y Tálamo, M. de F. (2002). Constitution of the scientific domain of Information Science. En: *Knowledge representation and organization of Gender Studies on the Internet : towards integration. Proceedings of the 7th International ISKO Conference*. Edited by M.J. López-Huertas. Würzburg: Ergon Verlag, p. 80-92

López-Huertas, María J. , Barité, Mario y Torres, Isabel de (2005). Terminología de Género para la recuperación de la información en documentos relacionados con los Estudios de la Mujeres. En: *Jornadas de Investigación Interdisciplinaria. Democracia, Feminismo y Universidad en el siglo XXI*. Madrid, Universidad Autónoma. Instituto de la Mujer, 2005, pp.321-330

López-Huertas. María J., Barité, Mario y Torres, Isabel de (2004). Terminological Representation of Specialized Areas in Conceptual Structures: the case of Gender Studies. En: *Proceedings of the 8th International ISKO Conference : 13-16 July 2004*, London / Edited by Ia C. McIlwaine. Würzburg: Ergon Verlag. p. 263-268.

López-Huertas, María J. (2006a) Análisis del dominio interdisciplinar para la representación y organización del conocimiento. Conferencia invitada en *Memória, Informação e Organização do Conhecimento: Cruzando as fronteiras da identidade*. Río de Janeiro, UNIRIO-IBICT, 2006, pp. 209-236

López-Huertas, María José (2006b) Thematic map of interdisciplinary domains based on their terminological representation. The Gender Studies. En: *Knowledge Organization for a Global Learning Society. Proceedings of the Ninth*

*Primer Simposio Internacional sobre Organización...*

*International ISKO Conference*. Vienna July 2006.  
Ed. Por Gerhard Budin, Christian Swertz y Konstantin  
Mitgutsch. Würzburg, Ergon, 2006, pp. 331-338.

López-Huertas, María José. (2006c) Aproximación a un  
modelo para la construcción de mapas temáticos  
en dominios interdisciplinares. En: *I International  
Conference on Multidisciplinary Information Sci-  
ences and Technologies*, InSciT2006. Cáceres, Octu-  
bre 2006. Badajoz, Instituto Abierto del Conocimien-  
to, 2006, pp. 548-552

Mcilwaine, Ia. (2000). Interdisciplinarity: a new re-  
trieval problem? En: *Dynamism and stability in  
knowledge organization. Proceedings of the sixth  
International ISKO Conference*. Würzburg, Ergon  
Verlag, pp. 261-267

Sebastiá i Salat, Montserrat (1988). *Thesaurus d'història  
social de la dona*. Barcelona, Generalitat de Catalunya.

Williamson, N. (1998). An interdisciplinary world and dis-  
cipline based classification. En: *Structures and rela-  
tions in knowledge organization*. Ed. by Widad Mus-  
tafá el Hadi, Würzburg, Ergon Verlag, pp. 115-123

# La organización y recuperación de la información de acuerdo con el contexto del usuario

NYDIA ROJAS MORALES

*Escuela de Bibliotecología, Documentación e Información,  
Universidad Nacional de Costa Rica.*

## 1. INTRODUCCIÓN:

La causa de la insatisfacción de los usuarios con los servicios bibliotecarios es una problemática que debe ser objeto de estudio en todas las instituciones informacionales.

Las unidades de información han tratado de resolver el problema incorporando diversos recursos de información con tratamientos más profundos para ofrecer servicios más especializados. No obstante, aunque se mejoran los servicios y se enfatiza en el usuario como el centro de la unidad de información, existe una percepción general que los métodos utilizados y las acciones emprendidas para satisfacer sus necesidades no concuerdan con esa premisa.

Por otra parte se ha promovido desde muchas décadas atrás que el trabajo de las bibliotecas inicia con el conocimiento de las necesidades de información de los usuarios, sin embargo, aún se enseña en muchas de nuestras universidades que el ciclo de trabajo de la unidad de información inicia con la selección y adquisición del material documental.

Se sigue percibiendo una imposición en las bibliotecas. Por ejemplo, cuando se ajusta a estrictos y complicados regímenes internos pa-

ra brindar servicios y atender al usuario o cuando organiza los recursos de información para que sean recuperados como lo establece una regla universal y no como lo buscaría un usuario con características socio-psicológicas, socio-demográficas y socio-económicas particulares.

Por tanto, en este trabajo se intenta establecer y poner en discusión los procesos de las actividades bibliotecarias esenciales para la recuperación de la información, sus bases epistemológicas y los métodos. Se proponen los siguientes objetivos:

- Establecer los procesos involucrados en la recuperación de la información,
- describir las bases epistemológicas en las que se fundamentan estos procesos y
- proponer los métodos para la ejecución de los procesos.

La investigación documental es el método empleado en este trabajo. Se consultaron diversas publicaciones relacionadas con aspectos epistemológicos y metodológicos de la bibliotecología, así como las relacionadas con la descripción de los procesos, para conocer las perspectivas teóricas de diferentes autores sobre la temática. A partir de esta información se elaboró la propuesta de procesos para la discusión.

## **2. LOS PROCESOS DE LA ACTIVIDAD BIBLIOTECARIA:**

Actualmente las unidades de información deben integrar procesos básicos que históricamente han sido relegados. Como lo expresa Cañedo (2004 p.128) “los niveles de exigencia de este nuevo estadio de desarrollo social requieren no solo de la entrega de información útil, sólida, es decir, valorada y significativa, sino que además es imprescindible que las bibliotecas participen activamente en el proceso de percepción, comprensión e introducción en la actividad del usuario...” Es decir, que para cumplir con esta exigencia, es necesario reestructurar la forma de trabajo para ocuparse de actividades fundamentales como: - el estudio y comprensión de las necesidades de información de

los usuarios y – la valoración y selección de la información que eventualmente resolvería esas necesidades.

Desde esta perspectiva, la actividad bibliotecaria se puede resumir en cinco grandes procesos:

1. La comprensión del usuario y sus necesidades de información.
2. La selección y valoración de los recursos de información para presentarlos de manera que su mensaje sea comprendido por el usuario y que le permita crear nuevo conocimiento.
3. La organización de los recursos de información tomando en cuenta los resultados de los procesos anteriores.
4. La suministración de los recursos de información.
5. La evaluación interna y externa.

A continuación se analizan los tres primeros procesos, siendo estos los que tienen más relación con el objeto de discusión de este Simposio.

## 2.1 La comprensión del usuario y sus necesidades de información

Para la ejecución del primer proceso es fundamental la investigación del usuario, es decir, el estudio de sus necesidades de información. Entendiendo que, de acuerdo con la definición que ofrece Núñez (2004), “Las necesidades peculiares de información son necesidades de carácter objetivo, que para cada usuario o lector, en un momento particular están determinadas por: El contenido semántico de la actividad que realiza el usuario o lector, lo que define la temática de la información que se necesita para la realización de dicha actividad —aspecto semántico— la estructura de la actividad (de estudio, de recreación, creadora, etc.), las condiciones objetivas y subjetivas, materiales y sociales para su realización, y las características socio-psicológico-culturales del usuario, lector, categoría (segmento) o comunidad de ellos”.

Tradicionalmente los estudios de necesidades de usuarios se han amparado bajo el paradigma físico, definido como “aquel que centra

su atención exclusivamente por aquello que puede ser observado” (Castillo y Leal, 2006). Por ejemplo, las características sociodemográficas y socioeconómicas del usuario.

También, existen muchas propuestas que consideran fundamental incluir variables psicológicas. Estas propuestas surgen bajo el paradigma cognitivo en el que se reconoce la importancia del incluir el aspecto subjetivo en la investigación.

Sin embargo, actualmente se enfatiza en la variable del entorno social, cultural e histórico del usuario ya que es considerada esencial en el estudio de las necesidades. Este nuevo enfoque es el denominado paradigma socio-cognitivo.

Los estudios de necesidades de usuarios sustentados en el paradigma socio-cognitivo investigan factores individuales y contextuales. Los factores individuales comprenden lo interno de cada persona, sus experiencias, personalidad, estilos de aprendizaje, habilidades, valores, actitudes, factores semánticos, entre otros. Los factores contextuales involucran los aspectos del entorno en el que el usuario se desenvuelve, que influyen en el comportamiento de la búsqueda de información, tales como, los precedentes históricos, la vida social, la influencia de la comunidad y la cultura.

Bajo esta premisa los estudios de usuarios enfocan su atención en entender los procesos mentales de los usuarios para determinar como asimila, comprende e interpreta la información. De este modo se conoce cual es la necesidad de información y “¿qué factores individuales, sociales, económicos y políticos les condiciona la búsqueda de información?” (González, 2005 p.23). Ya que el objetivo del estudio de usuarios, de acuerdo con la misma autora, es “averiguar qué información es la más apropiada para que este usuario resuelva los problemas informativos que le surgen en su entorno laboral, social, económico, político, etc.” (p.25).

La metodología propuesta para los estudios socio-cognitivos de los usuarios se fundamenta en el enfoque hermenéutico. Flores, R. y Tobón, A. (2001 p. 185-186) apoyados en otros autores como Gadamer, Ricoeur y Dewey hacen referencia al significado de la hermenéutica, lo que se sintetiza a continuación: La hermenéutica hace una interpretación global y holística sobre cosas o acontecimientos esencialmente

ligados a su contexto. Se interpreta el texto a una escala superior a la lingüística, como integrante de una realidad más amplia y de mayor complejidad. Focaliza un problema pero sin aislarlo de su contexto. Por su parte, el hermeneuta activa su experiencia previa, su mundo, su subjetividad, su propio horizonte conceptual para poder orientarse hacia el objeto de su indagación.

El estudio socio-cognitivo del usuario es un proceso en el que intervienen dos actores principales: El usuario con sus distintos horizontes de precomprensión compartidos socialmente y el entrevistador u observador con su propio horizonte de precomprensión. El resultado de este proceso es lo que Gadamer denomina una “fusión de horizontes”, que permite la interpretación holística de las necesidades de información del usuario.

El uso de técnicas como la observación participativa y la entrevista en profundidad facilitan el desarrollo de este proceso. El bibliotecólogo se pone de acuerdo con el usuario mediante el diálogo. Como lo expresan Flores, R. y Tobón, A. (2001, p 181) “La conversación es una elaboración del lenguaje común en tanto los interlocutores van entrando en la verdad de la misma cosa que los reúne”.

De este modo, el profesional valora los recursos de información y selecciona los más apropiados según el perfil del usuario. Esto corresponde al segundo de los procesos de la actividad bibliotecaria, antes mencionado, que se describe a continuación.

## 2.2. La valoración y selección de los recursos de información

Igualmente, en la valoración de la información, se parte de la hermenéutica para establecer el enfoque metodológico. El objetivo de aplicar este enfoque es interpretar adecuadamente el mensaje explícito e implícito que el autor del documento ha pretendido transmitir. Como en el proceso anterior, se da aquí una “fusión de horizontes” entre el profesional y el autor del documento. Así, el bibliotecólogo puede comprender y transformar ese mensaje de manera que sea recibido, interiorizado y convertido en conocimiento por un usuario con las características socio-cognitivas que le fueron identificadas en el estudio de necesidades.

En esta etapa, se considera el papel protagonista del profesional de la información como el elemento más importante del proceso. Es un profesional con las habilidades y competencias intelectuales idóneas para valorar la información y transformarla o adecuarla a la necesidad de quien la recibe. Es un trabajo que se desarrolla con el apoyo de otros especialistas quienes facilitan la comprensión del mensaje, su evaluación, la elaboración de comentarios y resúmenes, y otros, “para no quedar sólo en la mera recopilación de datos, sin crear o recrear saber” (Morales, 2005 p.104) y para apoyar fuertemente el proceso de asimilación del mensaje por parte del receptor.

El proceso intelectual realizado por el profesional de la información fundamentado en necesidades de información bien identificadas, y considerado de alto valor agregado, crea las condiciones para la satisfacción de estas necesidades. Como lo expresa Cañedo (2006), “los servicios diseñados a la medida, con alto valor agregado y orientados al conocimiento conforman un precedente básico en la obtención de un alto impacto en la comunidad de usuarios en la sociedad en general. El impacto de los servicios de información en la solución de los problemas de la sociedad, la razón de ser de estos servicios, depende de su adecuación a los requerimientos de quienes los reciben”.

Es así como a través de una interpretación adecuada y pertinente de los mensajes documentales, el profesional puede seleccionar los recursos de información apropiados para resolver las necesidades de información de los usuarios.

A partir de la selección, el trabajo bibliotecario inicia el proceso de organización, para un acceso eficiente y oportuno de estos recursos.

### 2.3 La organización de la información

La organización de la información es un proceso que comprende una gama extensa de actividades tales como el análisis documental (catalogación, indización, resúmenes), la clasificación, la formación de colecciones, la elaboración de bases de datos, entre otros.

Es uno de los procesos más estudiados y practicados en la actividad bibliotecaria. Esto se puede confirmar con revisión de la cantidad de cursos pertenecientes a esta área en los planes de estudio univer-



sitarios y la estimación del tiempo dedicado a estas actividades en las unidades de información. Probablemente superarán las demás áreas.

Esta situación debe ser objeto de análisis. Es común leer o escuchar que las bibliotecas han centrado su trabajo en la gestión de las colecciones y no en las necesidades y prioridades del usuario. Por ejemplo, Fernández Molina, citada por González (2005 p.64), opina que en la actividad bibliotecaria “se dejan al margen los aspectos humanos del proceso de búsqueda de información para centrarse casi exclusivamente en los aspectos técnicos del sistema sin tener en cuenta que este proceso, como en cualquier fenómeno social, intervienen diversos factores de carácter subjetivo”. Por lo tanto, se puede suponer que al dejar de lado estos factores será difícil satisfacer la necesidad de información de un usuario.

Entonces, necesariamente en el proceso de organización es fundamental basarse en el concepto subjetivo de la información, tomando en cuenta que la misma “no trasmite un significado constante y donde los usuarios la interpretan y le dan sentido o significado según la imagen única e individual que cada uno de ellos tiene en el mundo” (González, 2005 p.64) y según el contexto donde se desenvuelva, su entorno social, cultural e histórico. Es implica que el proceso de organización de la información también debe ser una actividad fundamentada en el paradigma socio-cognitivo.

Una actividad fundamental del proceso de organización de la información y que para efectos de esta investigación, es el punto central de análisis en este apartado, es la indización. La indización es la operación que busca representar tanto a los documentos como a las preguntas de los usuarios mediante los conceptos más relevantes de su contenido (Moreira, 2004 p.4). En otras palabras, el fin de esta actividad es guiar al usuario, a través de la comunicación de los contenidos que conforman el discurso científico, social y artístico del documento, hasta la ubicación exacta de la información que necesita.

Para lograr este fin es indispensable la realización de una indización profunda y exhaustiva. El enfoque hermenéutico se posiciona, también en la indización, como el punto de partida de este proceso, recomendada por las posturas más recientes como el paradigma socio-cognitivo.

No obstante, al utilizar este enfoque, es necesario el análisis del uso de los lenguajes documentales que hasta la fecha han sido, para esta tarea, las herramientas de trabajo por excelencia. Se conocen los lenguajes controlados y los codificados que tienen entre sus objetivos que la representación del contenido esté normalizada. Por otra parte están los lenguajes libres cuyas palabras o términos se escogen libremente desde los criterios del analista.

Surge entonces un gran cuestionamiento: ¿cuál es la herramienta más apropiada para el trabajo de indización bajo este enfoque? Capurro (2003) dice que una consecuencia práctica al asumir el paradigma socio-cognitivo, al que él denomina paradigma social, “es el abandonar la búsqueda de un lenguaje ideal para representar el conocimiento, o de un algoritmo ideal para modelar la recuperación de la información a lo aspiran el paradigma físico y el cognitivo”. Ya que la selección de cada término utilizado en la indización depende de la comprensión de su significado desde el contexto de donde se origina el mensaje y de la relación de este significado con la necesidad peculiar de información identificada en el usuario. Como lo expone Rendón (2004 p.180), “el código del sistema de información documental no solamente es recuperar o no información, sino recuperarla en función de un sujeto que se acerca a ese sistema para satisfacer necesidades de información”. Es decir, de acuerdo con el mismo autor, que el auténtico código de la bibliotecología es “satisfacer necesidades de información documental o no satisfacer necesidades de información documental”.

Por lo tanto, se puede afirmar que en la bibliotecología, el valor de satisfacer necesidades de información, está por encima de reglas, normas, herramientas y demás instrumentos utilizados para la organización de los recursos de información.

### 3. CONCLUSIONES

Existen dos procesos fundamentales que deben incluirse o reestructurarse en el trabajo bibliotecario, antes de iniciar las actividades de organización de los recursos de información. Estos son: La comprensión del usuario y la identificación de las necesidades de información,

así como la valoración y selección de los recursos de información que permitan resolver estas necesidades.

La teoría socio-cognitiva y la hermenéutica son los principales enfoques en los que se fundamentan los procesos descritos.

Basados en estos enfoques, los métodos aplicados en el desarrollo de los tres procesos analizados son esencialmente cualitativos. La identificación de las necesidades de información se realiza a través de la comprensión de los factores cognitivos y contextuales del usuario, utilizando técnicas como la entrevista en profundidad y la observación participativa.

Con respecto a la valoración y selección de los recursos de información, el profesional realiza una investigación documental con la que pretende comprender el mensaje explícito e implícito y su contexto, para determinar su pertinencia con las necesidades del usuario.

Igualmente en el proceso de organización de la información de los recursos de información, particularmente en la indización, el profesional realiza la investigación documental para comprender el texto, relacionarlo con el perfil del usuario y encontrar los términos propios para recuperar la información necesitada.

Queda para discusión y análisis, la selección de las herramientas idóneas para la indización bajo el enfoque socio-cognitivo.

#### 4. REFERENCIAS BIBLIOGRÁFICAS CITADAS Y COMPLEMENTARIAS

Camaraza, A. (2005). Recuperación de la información: reflexiones epistémicas de una ciencia en su estado embrionario. *ACIMED*, 13(6). Disponible en [http://bvs.sld.cu/revistas/aci/vol13\\_5\\_05/aci01505.htm](http://bvs.sld.cu/revistas/aci/vol13_5_05/aci01505.htm) Consultado 09/11/06.

Cañedo, R. (2006). Algunas aproximaciones teóricas a los servicios de alto valor agregado, diseñados a la medida y orientados al conocimiento. *ACIMED*, 14(1). Disponible en [http://bvs.sld.cu/revistas/aci/vol14\\_1\\_06/aci01106.htm](http://bvs.sld.cu/revistas/aci/vol14_1_06/aci01106.htm) Consultado 09/11/06.

***Primer Simposio Internacional sobre Organización...***

- Capurro, R. (2000). La hermenéutica y el fenómeno de la información. En: Carl Mitcham, ed.: *Mataphysics, Epistemology and Technology. Research in Philosophy and Technology*. JAI/Elsevier. 19: 79-85 (Traducción de la versión inglesa actualizada) Disponible en: <http://www.capurro.de> Consultado: 18/01/07
- Capurro R. (2003). Epistemología y Ciencia de la Información. Conferencia presentada en al V Encontro Nacional de Pesquisa em Ciencia da Informacão, Belo Horizonte (Brasil). Disponible en: <http://www.capurro.de> Consultado: 18/01/07
- Castillo, L. y Leal, O. (2006). Bibliotecología: ¿Ciencia o técnica? Hacia un nuevo debate. *ACIMED*, 14(2). Disponible en [http://bvs.sld.cu/revistas/aci/vol14\\_2\\_06/aci07206.htm](http://bvs.sld.cu/revistas/aci/vol14_2_06/aci07206.htm) Consultado 15/11/06.
- Flores, R.; Tobón, A. (2001). *Investigación educativa y pedagógica*. Bogotá: McGraw Hill. 212 p.
- Fox, V. (2005). *Análisis documental de contenido*. Argentina: Alfagrama. 251 p.
- Gadamer, G. (1973). *Verdad y Método*. Barcelona: Sígame. 2 Vol.
- González, A. (2005). *Los estudios de las necesidades y usos de la información: fundamentos y perspectivas actuales*. Gijón, Asturias: Trea. 181p.
- Hjørland, B.; Albrechtsen, H. (1995). Toward a new horizon in Information Science: Domain-Analysis. *Journal of the American Society for Information Science*. 46(6):400-425
- Morales, V. (2005). ¿Una o varias ciencias encargadas del estudio de la información registrada y organizada? *Documentación de las Ciencias de la Información*. 28: 197-210

- Morales, V. (2005). *Metodología en la bibliotecología*. Buenos Aires: Alfagrama. 128 p.
- Moreira, J.A. (2004). *El contenido de los documentos textuales: su análisis y representación mediante el lenguaje natural*. España: Trea. 291 p.
- Núñez, I. (2004). Las necesidades de información y formación: perspectivas socio-psicológica e informacional. ACIMED 12(5) Disponible en: [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S1024-94352004000500004&Ing=es&nrm=iso&tIng=es](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352004000500004&Ing=es&nrm=iso&tIng=es) Consultado 09/03/06.
- Peña, V. (2005). Vinculación del constructivismo con el enfoque socio-psicológico para el estudio de usuarios de unidades de información: un primer acercamiento teórico. *Investigación Bibliotecológica*, 19(39): 48-72
- Rendón, M. (2004). Axiología y ciencia bibliotecológica: los valores en el mundo de la información documental. *Investigación Bibliotecológica*. 18(36): 170-184
- Rendón, M. (1998). *Bases teóricas y filosóficas de la bibliotecología*. México: UNAM, Centro de Investigaciones Bibliotecológicas. 148 p.



# Comportamiento de búsqueda y uso de la información ambiental por investigadores del área de medio ambiente<sup>1</sup>

FERNANDO BITTENCOURT DOS SANTOS

HELEN DE CASTRO SILVA

*Universidade Estadual Paulista "Júlio de Mesquita Filho", Brasil*

## 1 INTRODUCCIÓN Y JUSTIFICATIVA

La cuestión ambiental ocupa un importante espacio en la sociedad y viene creciendo cada vez más por su alcance, ya que es tema de discusión en diversos contextos, incluso en conferencias en un nivel nacional e internacional. Los problemas ambientales con la consecuente degradación de la calidad de vida pueden ser identificados tanto en las áreas urbanas, por la enorme concentración de industrias, fábricas y por la desordenada ocupación del suelo, como también en las áreas rurales, donde hay la excesiva concentración agraria y la agricultura dirigida al desmatamiento. La polución, los cambios climáticos, la creación de los alimentos transgénicos, la desertificación y la pérdida de la biodiversidad también son temas de gran repercusión y relevancia dentro del contexto ambiental.

En Brasil y en el mundo podemos notar la actuación de ONGs (Organizaciones No Gubernamentales), que participan en el enfrentamiento de los problemas ambientales, actúan en el interés público de

---

1 Pesquisa em desenvolvimento en el Curso de Maestría em Ciencia de la Información de la Universidade Estadual Paulista "Júlio de Mesquita Filho" - Brasil

forma directa, en la prestación de servicios y en el desarrollo de proyectos que tienen como objetivo: proteger el medio ambiente y luchar por el desarrollo sustentable.

Según Caribé (1992)... respecto a la preocupación del gobierno brasileño con los problemas del medio ambiente es insuficiente y tímida. De hecho esa falta de acción del gobierno federal es presenciada en los días de hoy, cuando, por ejemplo, se le da más prioridad al crecimiento económico e industrial, a la política interna y externa, y se deja casi en la inercia la lucha por una vía de desarrollo sustentable.

Mittermier *et al.* (1992) agregan que Brasil es el país de mayor riqueza en cuanto a la biodiversidad, donde se concentra un 10% de todo el acervo de recursos genéticos del planeta, pero persiste en un proceso generalizado de destrucción de su rico patrimonio ecológico. El modelo actual de desarrollo, no sólo en Brasil, pero en muchos países, es el principal responsable por la agravante situación en que el medio ambiente se encuentra. El hombre marca su presencia al consumir y contaminar los recursos naturales casi escasos, sin darse cuenta de la destrucción y de los daños que causa al medio ambiente. La naturaleza es considerada muchas veces como fuente de recursos inagotable y gratuita, la explotación es ilimitada y devastadora, lo que acaba deteriorando la calidad de vida de la sociedad e influyendo negativamente en su desarrollo.

Debido a la importancia del medio ambiente, muchos profesionales trabajan cotidianamente en la defensa, en el estudio, en la búsqueda y utilización de informaciones de esta área. Las necesidades de información ambiental de esos profesionales, conforme Caribé (1992), varían en función del grado de desarrollo y de la política económica adoptada por el país en que estos actúan.

Los investigadores y expertos del área ambiental, por ejemplo, tienen como principal foco de estudios el conjunto de condiciones, leyes, influencias e interacciones de orden física, química y biológica que permite, abriga y rige la vida en todas sus formas en el planeta Tierra.

Caribé (1992, p. 41) enfatiza que el área de medio ambiente presenta facetas de difícil control por ser multi e interdisciplinar:



[...] el área toma en consideración conceptos científicos, sociales, religiosos y filosóficos, incluye valores políticos y económicos y discute conceptos de las ciencias físicas y biológicas. Los asuntos de medio ambiente están ligados al área científica, médica y de ingeniería, tales como Geología, Geografía, Química, Biología, Hidrología, Ingeniería Química, Ingeniería Ambiental, Ingeniería Sanitaria, Investigación Operacional y otras. Involucran también las ciencias sociales con aspectos económicos, política económica, gerenciamiento y administración, política gubernamental e implicaciones sociales. Para desarrollarse cualquier estudio sistemático en el área, son necesarios parámetros y conceptos pertenecientes a varias ramas de la ciencia y tecnología.

Debido a la complejidad del área y de la información ambiental, estudios que permitan conocer mejor este contexto son de extrema relevancia. La presente investigación se propone a estudiar el comportamiento de búsqueda y el uso de la información, bien como las necesidades informacionales de investigadores de seis instituciones de investigación que actúan en el área del medio ambiente, a saber: Núcleo de Estudios e Investigaciones Ambientales (NEPAM) y Instituto de Biología (IB) de la Universidad Estadual de Campinas (UNICAMP), el Departamento de Ciencias Biológicas de Bauru y de Assis, pertenecientes a la Universidad Estadual Paulista (DCB - UNESP/BAURU) y (DCB- UNESP/ASSIS) y el Instituto Oceanográfico (IO) y el Instituto de Astronomía, Geofísica y Ciencias Atmosféricas (IAG) de la Universidad de São Paulo (USP).

Estos son ejemplos de instituciones de referencia en investigación en el área de medio ambiente, donde la información ambiental es buscada y utilizada por los investigadores que la producen para la generación de nuevos conocimientos. El estudio del comportamiento de búsqueda y uso de la información de los investigadores de estas instituciones posibilitará un panorama bastante apurado y representativo sobre el comportamiento informacional de investigadores brasileños de esta área.

Estudios sobre el comportamiento de búsqueda y uso informacional de diferentes grupos de investigadores son muy importantes, pues cada investigador de un área posee comportamientos informacionales que le son propios (Mueller, 2005). Así, conforme apunta Calva Gonzalez (2004, p. 52) “(..)los resultados obtenidos de las investigaciones realizadas sobre las necesidades, comportamiento y satisfacción permiten tener elementos para el mejoramiento continuo de las unidades de información que atienden a comunidades específicas de usuarios”.

## 2 EL MEDIO AMBIENTE EN BRASIL

Ante la coyuntura actual y si se considera el modelo de desarrollo económico adoptado por muchos países, la cuestión ambiental ocupa un importante espacio en la esfera global.

Albagli (1995, p. 2) afirma que:

El tema ambiental ha dejado de ser visto como problema restringido al medio técnico-científico, y ocupa un lugar prominente en la agenda política tanto de los países de manera individual, como de las negociaciones que ellos llevan a cabo en el área internacional. Los temas relativos al medio ambiente han pasado a ser objeto de preocupación y de debate por parte de los medios de comunicación, de las entidades representativas de la sociedad civil organizada (particularmente las Organizaciones No Gubernamentales - ONG's) y de la opinión pública en general, de las empresas, de los gobiernos, de las instituciones y de las organizaciones internacionales.

Hace mucho tiempo que la cuestión ambiental dejó de ser solamente un movimiento por la defensa de animales amenazados de extinción y de las florestas brasileñas contra el desmatamiento.

La cuestión ambiental ha crecido en todo el mundo, tanto en repercusión como en adeptos, como demostraron las dos Conferencias

Mundiales sobre Medio Ambiente, la Conferencia sobre Medio Ambiente Humano en Estocolmo, en el año de 1972, y la Conferencia de las Naciones Unidas para el Medio Ambiente y el Desarrollo, en Río de Janeiro en 1992, de modo que a partir de esta última se cambió el término “ecodesarrollo” para “desarrollo sustentable”, que es el desarrollo económico, social, científico y cultural de las sociedades y que garantiza más salud, comodidad y conocimiento, pero sin causar muchos daños a los recursos naturales del planeta Tierra. Para eso, todas las formas de relación del hombre con la naturaleza deben ocurrir con el menor daño posible al medio ambiente. Cabe resaltar también la creación de la Agenda 21, por el gobierno federal brasileño en 1997, un importante documento que contiene más de 900 páginas, divididas en 40 capítulos. Algunos de los principales puntos discutidos en este documento son: miseria, cambios en los patrones de consumo, población, salud humana, políticas para el desarrollo sustentable, protección de la atmósfera, residuos peligrosos, salvaguardia de los recursos oceánicos y promoción de la conciencia ambiental. Este documento fue adoptado por consenso entre los países participantes.

Brasil es conocido por sus dimensiones continentales, por su enorme variedad climática, su agigantado patrimonio ambiental y la mayor diversidad biológica del planeta.

Para que la utilización de los recursos naturales sea adecuada a los presupuestos fundamentales del desarrollo sustentable, es crucial el fomento del Gobierno brasileño de la educación e investigación ambiental, a fin de buscar un medio sustentable de desarrollo. La conservación y la búsqueda por mejoras para el medio ambiente no es sólo responsabilidad de ecologistas y profesionales relacionados al área ambiental (biólogos, geógrafos, oceanógrafos, meteorólogos, ecologistas, entre otros), sino también de todos los individuos de una nación, pues si todos se preocupan por la naturaleza y los diversos recursos ofrecidos por ésta, alcanzaremos la calidad de vida que tanto deseamos.

### 3 FUENTES DE INFORMACIÓN EN MEDIO AMBIENTE

Información puede ser considerada cualquier elemento referencial contenido en un documento, sea éste impreso, electrónico o audiovi-

sual. Su valor se define en función del contexto y del interés del receptor y de su grado de aptitud y dominio sobre un determinado asunto.

Las fuentes de información son cualesquiera de los recursos que respondan a una demanda de información por parte de los usuarios, lo que incluye productos y servicios de información, personas o red de personas, etcétera.

Con el objetivo de orientar a los usuarios y/o investigadores en la búsqueda de la información, las fuentes de información se constituyen en un vehículo de significativa importancia en el desarrollo de la investigación científica y en la diseminación del conocimiento.

Las fuentes de información en materia de medio ambiente son elementos imprescindibles para los investigadores en la planeación y el desarrollo de investigaciones con la temática ambiental, especialmente en los campos científicos y tecnológicos.

Existen varias fuentes en los niveles primario, secundario y terciario, que incluyen grandes y pequeñas proporciones de información ambiental. Entre ellas, Harvey (1973) *apud* Caribe (1988, p. 94) resalta la importancia de las Relatorías de Evaluación de Impacto Ambiental como fuente de información no sólo para la toma de decisiones, sino también para el técnico, el investigador y para el público, transformándose en una herramienta imprescindible para la toma de decisión en todos los niveles. En estas relatorías están incluidas todas las informaciones relacionadas con la actividad que se realizará: metas, procesos para ser utilizados, descripción del área y medio ambiente que serán afectados, recursos naturales, entre otros.

La realización de esta investigación tiene como objetivos:

- Caracterizar la información ambiental, lo que incluye aspectos históricos y actuales, con énfasis en el contexto de los países del “Tercer Mundo” y particularmente de Brasil, en cuanto a características de organización y uso;
- Analizar el comportamiento de búsqueda y uso de la información ambiental por parte de los investigadores del NEPAM y del IB (Unicamp), DCB/Bauru e DCB/Assis (Unesp) e IO e IAG (USP), a fin de conocer los hábitos y las necesidades de información que estos presentan;

- Determinar la tipología de la información utilizada por los investigadores del área ambiental, bien como formas y mecanismos de obtención de esta información.

#### 4 REVISIÓN PRELIMINAR DE LITERATURA

La información es, en el mundo contemporáneo, uno de los factores de mayor importancia para el fortalecimiento de las relaciones entre los seres humanos, que pasa por todas las actividades personales, intelectuales y comerciales, además de presentar una diversidad de soportes y formatos.

Actualmente, el desarrollo de los países está significativamente ligado a la producción, organización y al dominio de la información que puede provocar cambios que afecten a la economía global. Entre sus tipos está la información ecológica o información ambiental, resultado de los desasosiegos crecientes en cuanto a la preservación del medio ambiente.

En la concepción de Targino (1994), la información ambiental puede ser considerada como

[...] datos, informaciones, metodologías y procesos de representación, reflexión y transformación de la realidad, que facilitan la visión holística del mundo y, además, contribuyen para la comprensión, análisis e interacción armónica de los elementos naturales, humanos y sociales.

Según Vieira (1986, p. 202), la información ambiental está compuesta por dos tipos básicos de información:

[...] la información tecnológica, económica y social para la orientación de acciones, tanto en la esfera gubernamental como en el ámbito empresarial, y la información para concienciación de la población en cuanto a sus derechos y deberes con el medio ambiente.

Tavares y Freire (2003) afirman que:

La información ambiental es uno de los tipos de información científica y tecnológica. Ella es consecuencia de la preocupación de la sociedad con los efectos e impactos de la producción y del consumo sobre el medio ambiente.

La información ambiental que constituirá el objeto de estudio de esta investigación será la información que refleja el aspecto tecnológico, económico y social, pues ésta se presenta como soporte para los investigadores del área ambiental en el desarrollo de trabajos de carácter científico.

Las definiciones presentadas por los autores conceptúan y caracterizan este tipo de información, que representa un importante aporte para los investigadores en el desarrollo de sus investigaciones y tecnologías y para la concienciación de la sociedad civil sobre los problemas y soluciones posibles sobre la cuestión. Hay un gran volumen de informaciones que están siendo producidas sobre esta temática, lo que dificulta el control y el almacenamiento de la documentación pertinente producida sobre los asuntos relacionados al medio ambiente. El gran obstáculo entonces es la falta de conocimiento así como la dificultad de búsqueda y recuperación de esta información.

Muller (1992, p. 14) afirma que la producción de informaciones ambientales pretende:

[...] proveer subsidios para el enfoque apropiado de los impactos de fenómenos naturales y de las actividades humanas sobre el medio ambiente y sobre la calidad de vida del ser humano en el sentido de proveer informaciones y análisis relevantes al planeamiento y la formulación de políticas sociales, económicas y ambientales integradas.

Por eso, la información ambiental puede desempeñar un papel fundamental en las decisiones político-económicas, con el objetivo de

conseguir acciones que mejoren las condiciones de vida, como el subsidio para el desarrollo de tecnologías apropiadas y en la educación, con vistas al uso racional de los recursos naturales (Almeida, Cordeiro e Caribé, 1987).

Albagli (1995) afirma que la información ambiental constituye un elemento clave en el modelo de desarrollo sustentable, de modo que los principios que rigen ese modelo son: el principio de la eficacia en el uso de recursos, ya que la información es requisito básico en la estructuración de un proceso productivo menos consumidor de materias primas y energía; el principio de la diversidad, considerándose que la información es fundamental en el reconocimiento y potencialización de la diversidad de mercados, recursos y capital humano locales; el principio de la descentralización, a medida que la información es esencial para instrumentalizar los diferentes actores para actuar en conjunto, con responsabilidades y competencias definidas.

Del punto de vista administrativo, las responsabilidades, atribuciones y decisiones sobre los aspectos ambientales se encuentran diluidas entre diversas instituciones que actúan en el área. La información ambiental, consecuentemente también se encuentra dispersa entre varios órganos (Caribé, 1992). Por lo demás, los documentos publicados sobre esos diversos asuntos están dispersos entre una gran variedad de fuentes, y, como resultado, la información no está organizada de forma que pueda ser recuperada con eficacia y rapidez.

La autora aún afirma que en numerosos asuntos en el área ambiental, hay mucho trabajo para realizar a fin de volver los datos disponibles. Los propios organismos internacionales reconocen dificultades para la constitución de un sistema de información en esta área, principalmente en función del carácter reciente de las ciencias ambientales, del enfoque multidisciplinar de las estadísticas ambientales, de los recursos exigidos para la búsqueda y obtención de los datos.

La Conferencia de las Naciones Unidas sobre Medio Ambiente y Desarrollo (1992, p. 575), señala que:

[...] aunque haya una cantidad considerable de datos sobre medio ambiente, es preciso reunir más y diferentes tipos de datos, en los planos provincial, nacio-

nal e internacional, que indiquen los estados y tendencias de las variables socio-económicas, de polución, de recursos naturales y del ecosistema del planeta. En razón de eso, viene aumentando la diferencia en términos de disponibilidad de los datos entre el mundo desarrollado, lo que perjudica seriamente la capacidad de los países de tomar decisiones informadas en lo que concierne al medio ambiente y desarrollo.

La Conferencia de las Naciones Unidas sobre Medio Ambiente y Desarrollo (*op. cit.*, p. 575) resalta aún que es preciso mejorar la coordinación entre las actividades de información y los datos ambientales, demográficos, sociales y de desarrollo.

Con relación a la localización de esta información, Almeida; Cordeiro; Caribé (1987) afirman que la información ambiental se encuentra fragmentada y dispersa entre las más variadas instituciones que proveen informaciones y actúan en el área del medio ambiente, y cita como ejemplos los centros de investigaciones, las bibliotecas, los centros de documentación, etc. Caribé (1992) complementa esa información y afirma que la información ambiental tiene un importante papel para informar a los individuos sobre los problemas y soluciones posibles sobre la cuestión, además de controlar y almacenar la documentación pertinente producida sobre los asuntos relacionados al medio ambiente.

Aún de acuerdo a Somerville (1976) *apud* Caribé (1992, p. 41) existen otras características de la información ambiental, tales como:

β los datos ambientales por lo regular no son publicados, y el conocimiento de su existencia sólo será posible tras larga y/o exhaustiva experiencia en el área;

β gran cantidad de información y datos relevantes aparecen en publicaciones que no son indizadas o que lo son inadecuadamente;

β gran cantidad de informaciones producidas y/o recolectadas por empresas privadas o por el gobierno no están disponibles al público,

β los datos disponibles son dudosos en lo que se refiere a la calidad y duración, ya que los métodos de recolecta no están estandarizados ni son ampliamente conocidos;



β la información está expuesta a la manipulación por presiones políticas, consideraciones emocionales o conservacionistas, o aún presiones de grupos económicos;

β el rápido desuso de datos y de la literatura;

β la insuficiencia de fuentes y de obras de referencia que indiquen informaciones publicadas;

β la urgencia de demanda de la información para la acción del usuario;

β el área donde los colegios invisibles no son estáticos;

Conforme Caribé (1992), las principales dificultades en la producción y diseminación de la información ambiental en Brasil son:

- el crecimiento y la urgencia de la demanda;
- el desconocimiento de fuentes generadoras de datos ambientales;
- las deficientes estandarizaciones de metodologías de recolecta de datos y de indización;
- el acceso restringido a las informaciones producidas por algunos sectores de la sociedad, especialmente aquellos considerados estratégicos;
- las manipulaciones provocadas por presiones políticas, económicas o, incluso, conservacionistas;
- el rápido desuso y constante riesgo de interrupción de series históricas.

Con relación al enfoque del tema en el área de Ciencia de Información, Tavares y Freire (2003, p. 209) afirman que, “estudios que enfocan la información ambiental en el campo de la Ciencia de la Información han aumentado en los últimos años, debido a la crisis ambiental que se vive actualmente, en el país y en el mundo”.

El enfoque del tema “información ambiental” está creciendo mucho en el medio científico y los profesionales de la información están cada vez más interesados por la temática ambiental y desarrollan investigaciones .

## 5 PROCEDIMIENTOS

Para auxiliar la interpretación de las informaciones y fundamentar la investigación será realizado el análisis de la literatura nacional e internacional disponible, con el fin de caracterizar el área de estudio bajo los aspectos: histórico, de desarrollo económico y de la situación de los órganos ambientales del país, así como en las áreas de investigación y enseñanza relacionadas con el segmento ambiental y del escopo del área. Serán enfocados también los temas de comportamiento de búsqueda y uso de la información por grupos específicos, en particular de expertos y en contexto de producción de investigaciones.

Con objeto de caracterizar el perfil de los investigadores brasileños como usuarios de la información ambiental e identificar el comportamiento de búsqueda y uso de la información por los investigadores del NEPAM e IB pertenecientes a la Universidad Estadual de Campinas; DCB (Bauru), DCB (Assis) pertenecientes a la Universidad Estadual Paulista y el IO e IAG, ambos de la Universidad de São Paulo, se pretende en esta investigación emplear una metodología exploratoria/comparativa, que consiste en la aplicación de un cuestionario (que contiene preguntas abiertas y cerradas) con todos los investigadores (alumnos de Postgrado - Maestría y Doctorado; profesores y alumnos de iniciación científica) de las seis instituciones seleccionadas.

La elaboración del cuestionario estará basada en la literatura sobre el comportamiento de búsqueda de la información (Albrechtsen, H.; Hjørland, B., 1997; Johnson, J. D. E. et. al., 2006; Foster, A. E. 2004).

Antes de la aplicación del cuestionario será realizado un pre-test con veinte investigadores del área ambiental de las referidas instituciones para la realización de posibles ajustes en el instrumento de recolecta de datos. La aplicación del cuestionario será realizada por correo electrónico.

Las instituciones citadas figuran como importantes centros de investigaciones en el área de medio ambiente y estos ofrecen infraestructura adecuada para el desarrollo de diversos proyectos de investigación, y tiene como objetivos la formación, científica, tecnológica y cultural de investigadores y profesionales con alto nivel. Presentan

como diferenciales en el área de investigación, la capacidad de desarrollar proyectos multidisciplinares, debido a diversas líneas de investigación que poseen y a la calificación de sus investigadores en cuanto a la temática ambiental.

Por la búsqueda y utilización de la información ambiental en las instituciones citadas anteriormente, el trabajo será desarrollado teniendo como base estas instituciones, pues las mismas constituyen importantes referenciales para la temática enfocada en esta investigación.

Estas últimas podrán conocer tal opinión sobre la información ambiental y buscarán, si es necesario, adecuarse a las necesidades de esos investigadores con relación a la información, de manera que optimice y dinamice sus servicios de búsqueda y referencia concierne al área de medio ambiente. Así las informaciones podrán ser organizadas y, por consiguiente, diseminadas, garantizando la calidad y confiabilidad en la recuperación del contenido informacional, bien como de sus productos documentarios, permitiendo un fácil acceso a ellos por investigadores que necesiten de la información con calidad y confiable.

## REFERENCIAS

- Albagli, Sarita (1995). Informação e Desenvolvimento Sustentável: Novas Questões Para O Século Xxi. *Ciência da Informação*. 24: 1 (1995).
- Albrechtsen, H.; Hjörland, B. (1997) Information Seeking And Knowledge Organization. *Knowledge Organization*. 24:3 (1997) 136-144.
- Almeida, Iêda Muniz De; Cordeiro, Lia Prado Arrivabene; Caribé, Rita de Cássia do Vale (1987). Estudo da Necessidade de criação de uma Base de Dados sobre poluição. *Revista De Biblioteconomia De Brasília*. 15: 2 (Jul/Dez. 1987) 343-353.

**Primer Simposio Internacional sobre Organización...**

Calva Gonzales, J. J. (2004) *Investigacion Bibliotecológica*, 37:18  
(2004) 52 -

Caribé, Rita De Cássia Do Vale. *Informação Ambiental No Brasil: Subsídios Para Um Sistema De Informação*. Brasília: Universidade De Brasília, 1988, 218p. (Dissertação De Mestrado)

Caribé, Rita De Cássia Do Vale (1992). Subsídios para um Sistema de Informação Ambiental no Brasil. *Ciência Da Informação Brasileira*. 21 :1 (Jan./Abr. 1992) 40-45,.

Conferência Das Nações Unidas Sobre Meio Ambiente e Desenvolvimento (1992). Agenda. Brasília: Senado Federal, Subsecretaria De Edições Técnicas, 1996, 585p.

Foster, A. E. (2004) A Non-Linear Model Of Information Seeking Behaviour. *Journal Of The American Society For Information Science And Technology*. 55:3 (2004) 228-237.

Johnson, J. D. E. et. al. (2006) Fiels And Pathways: Contrasting or complementary views of Information Seeking. *Information Processing & Management*. 42 (2006) 569-582.

Mittermeier, R. A et al (1992). O País Da Megadiversidade. *Ciência Hoje*. 14: 81 (Maio/Jun. 1992) 20-27.

Mueller, Suzana Pinheiro Machado (2005). A Publicação Da Ciência: Áreas Científicas E Seus Canais Preferenciais. *Datagramazero - Revista De Ciência Da Informação*, V.6 N.1 Fev. 2005. Online. Disponível Em: <[Http://Www.Datagramazero.Org.Br/fev05/F\\_I\\_Art.Htm](http://www.datagramazero.org.br/fev05/F_I_Art.htm)>

Muller, C. C (1992). Situação Atual da Produção De Informações Sistemáticas Sobre O Meio Ambiente. *Ciência da Informação*. 21: 1 (Jan./Abr. 1992) 14-22.

Targino, Maria Das Graças (1994). Informação Ambiental - Uma Prioridade Nacional? *Informação e sociedade: estudos*. 4: 1 (1994) 1-27.

## *Representación y recuperación de información*

Tavares, Carla; Freire, Isa Maria (2003). “Lugar Do Lixo É No Lixo”: Estudo de assimilação da Informação”. *Ciência da Informação*. 32: 2 (Maio/Ago. 2003) 125-135.

Tavares, Carla; Freire, Isa Maria (2003). Informação Ambiental No Brasil: Para que e para quem. *Perspectivas em Ciência da Informação*. 8: 2 (Jul./Dez. 2003) 208-215.

Vieira, Anna Da Soledade (1986). Pra Não Dizer Que Não Falei De Flores: Uma Proposta Ecológica Para A Biblioteconomia. *Rev. Esc. De Bibliotecon.* UFMG. 15:2 (set. 1986) 202-209.



## Procedimientos terminológicos y documentales utilizados en la construcción, implementación y gestión del Vocabulario Controlado USP – descriptores en portugués para la representación y recuperación de la información.

VÂNIA MARA ALVES LIMA, NAIR YUMIKO KOBASHI, VERA REGINA  
CASARI BOCCATO, MARIA ÂNGELA DE TOLEDO LEME,  
MARIZA LEAL DE MEIRELLES DO COUTTO  
*Universidade de São Paulo,  
Universidade Federal de São Carlos, Brasil*

**E**ste trabajo describe los procedimientos terminológicos y documentales utilizados en la construcción, implementación y gestión del Vocabulario Controlado de SIBi/USP (Sistema Integrado de Bibliotecas de la Universidade de São Paulo). Este vocabulario se utiliza para la representación y recuperación de informaciones del Banco de Datos Bibliográficos de la Universidade de São Paulo - Brasil - *DEDALUS*. Del proyecto de construcción participaron los bibliotecarios del Sistema y los docentes de la universidad. Estos últimos, colaboraron con la selección de términos, estructuración de los sistemas conceptuales y adecuación terminológica de las áreas. La elaboración del Vocabulario se fundamenta en principios terminológicos y de documentación y presenta relaciones lógico-semánticas explícitas entre los términos, en sus diferentes niveles. Está disponible en dos versiones, Intranet e Internet, siendo permanentemente actualizado. Para esto, se utiliza un sistema de sugerencias *online*, dispositivo en red, que permite coleccionar e implementar términos nuevos, desechar términos no utilizados y consolidarlos, incorporando así la dinámica de la creación terminológica de los diferentes campos del conocimiento. Las reglas de utilización del Vocabulario están consolidadas en un manual de Indexación. Es el único lenguaje documental, en lengua

portuguesa, que abarca las más diversas áreas del conocimiento. Por estar disponible en la Web, el Vocabulario Controlado de *SIBi/USP* es utilizado por diferentes instituciones de enseñanza superior del país.

## 1 INTRODUCCIÓN

La utilidad de las clasificaciones bibliográficas, o lenguajes documentales, para organizar acervos y facilitar la localización de documentos, es incuestionable. La discusión sustantiva sobre el tema se refiere, hoy, a las potencialidades de cada tipo de lenguaje para cumplir las funciones para la cual fue creada.

Con la creciente ampliación de las posibilidades de almacenar información en repositorios electrónicos y accederlos remotamente, urge retomar las reflexiones sobre las teorías y métodos para construir los lenguajes documentales, visto que esos instrumentos, hoy, ultrapasan el espacio físico, porque integran conjuntos de catálogos informatizados, interligados en red, con base a protocolos y lenguajes documentales comunes (POMBO, 2002).

La intensificación del uso de las tecnologías de la comunicación y de la información (TICs), en el campo del trabajo de la documentación, permite transferir para el ordenador muchas tareas rutinarias, tales como gestión de acervos, préstamos, consulta al fondo documental, inventario, emisión de informes, entre otras. Las facilidades creadas, por otro lado, muestran la necesidad de instrumentos cada vez más sofisticados de codificación de la información para recuperación. La creación de estos instrumentos presenta inúmeros desafíos teóricos y metodológicos.

La experiencia de la construcción del Vocabulario Controlado del Sistema Integrado de Bibliotecas de USP (SIBi/USP), para indexación y recuperación de la información, en el Banco de Datos Bibliográficos de USP - DEDALUS, puede ser útil para llamar la atención sobre los aspectos pragmáticos de la construcción de lenguajes documentales para tratar la información de acervos de universidades. El camino recorrido será discutido aquí críticamente. Para eso, será presentado, inicialmente, una breve reseña histórica del proceso de construcción del vocabulario, los procedimientos documentales y terminológicos



utilizados, su implementación y, finalmente, la administración de las actividades de actualización del vocabulario.

## 2 CONSTRUCCIÓN DEL VOCABULARIO CONTROLADO DE SIBI/USP: BREVE RESEÑA HISTÓRICA

La Universidade de São Paulo cuenta, actualmente, con 40 bibliotecas, con acervos especializados, subordinadas administrativamente a las unidades específicas de enseñanza e investigación. En 1981, fue creado el Sistema de Bibliotecas - SIBi/USP, con el objetivo de administrar globalmente los recursos informacionales de la universidad. En 1985, fue iniciada la construcción del catálogo global del acervo de las bibliotecas de USP, destinado al acceso online, denominado Banco de Datos Bibliográficos de USP - DEDALUS.

En este proceso, además de la padronización de los registros, fue elaborada una “*Lista de Asuntos USP*”, para ser utilizada en el proceso de indexación para la Base Bibliográfica DEDALUS. Hasta ese momento, cada biblioteca utilizaba el lenguaje documental más adaptado a la especialidad de la unidad de enseñanza e investigación a la que estaba vinculada. La “*Lista de Asuntos USP*” era una lista de títulos de temas muy genéricos y, por ese motivo, presentaba poca compatibilidad con los lenguajes documentales utilizados en cada biblioteca, teniendo, además, diversas limitaciones, tanto estructurales como terminológicas.

Pensando en resolver este problema, en 1993, fue constituido un equipo específico para mejorar la “*Lista de Asuntos USP*”. Este equipo, formado por los bibliotecarios indexadores de cada biblioteca del Sistema, participaron del curso “Compatibilización de lenguajes documentales” para actualizar y homogenizar las competencias para tal tarea. Al final del curso, los participantes elaboraron el “Proyecto para el mejoramiento de la Lista de Asuntos USP” el cual tuvo inicio a fines de 1994.

Como estrategia para el desarrollo del proyecto se elaboró el software “*Programa Meta 1*”, en lenguaje Clipper, para la colecta de los temas de las bibliotecas. Después de la colecta y el análisis de los datos, el equipo optó no por el mejoramiento de la “*Lista de Asuntos*

**Primer Simposio Internacional sobre Organización...**

USP”, y sí por la creación de un vocabulario unificado, compatible con los lenguajes documentales utilizados por las bibliotecas del Sistema. Para dar inicio al trabajo los bibliotecarios representantes de las bibliotecas de SIBi/USP fueron divididos en nueve subgrupos, por áreas temáticas, según lo presentado en el Cuadro

**Cuadro 1 – Bibliotecas de SIBi/USP Organizadas por Subgrupos**

Sub-grupo	Áreas Temáticas	Bibliotecas Participantes
1	Filosofía, Ciencia Política, Antropología, Sociología, Historia, Historia de Brasil, Museología, Geografía, Letras y Lingüística.	<ul style="list-style-type: none"> <li>• Facultad de Filosofía, Letras y Ciencias Humanas</li> <li>• Instituto de Estudios Brasileños</li> <li>• Museo Paulista</li> </ul>
2	Economía, Derecho, Administración, Contabilidad	<ul style="list-style-type: none"> <li>• Facultad de Derecho</li> <li>• Facultad de Economía y Administración</li> </ul>
3	Educación, Psicología	<ul style="list-style-type: none"> <li>• Facultad de Educación</li> <li>• Instituto de Psicología</li> </ul>
4	Artes, Arquitectura y Urbanismo, Comunicaciones	<ul style="list-style-type: none"> <li>• Escuela de Comunicaciones y Artes</li> <li>• Facultad de Arquitectura y Urbanismo</li> <li>• Museo de Arte Contemporánea</li> </ul>
5	Geología, Química, Bioquímica, Farmacia, Matemática, Estadística, Astronomía, Geofísica y Física	<ul style="list-style-type: none"> <li>• Instituto Astronómico y Geofísico</li> <li>• Instituto de Física</li> <li>• Instituto de Física de São Carlos</li> <li>• Instituto de Geociencias</li> <li>• Instituto de Matemática y Estadística</li> <li>• Instituto de Química y Facultad de Ciencias Farmacéuticas (Conjunto de las Químicas)</li> <li>• Instituto de Química de São Carlos</li> </ul>
6	Medicina, Salud Pública, Nutrición, Enfermería, Educación Física y Deportes	<ul style="list-style-type: none"> <li>• Escuela de Educación Física y Deportes</li> <li>• Escuela de Enfermería</li> <li>• Facultad de Medicina</li> <li>• Facultad de Salud Pública</li> <li>• Unidades del Campus de Ribeirão Preto (Biblioteca Central)</li> </ul>
7	Biología, Botánica, Oceanografía, Zoología, Medicina Veterinaria, Zootecnia y Ciencias Agrarias	<ul style="list-style-type: none"> <li>• Escuela Superior de Agricultura “Luiz de Queiroz”</li> <li>• Facultad de Medicina Veterinaria y Zootecnia</li> <li>• Instituto de Biociencias</li> <li>• Instituto de Ciencias Biomédicas</li> <li>• Instituto Oceanográfico</li> <li>• Museo de Zoología</li> </ul>
8	Ingeniería	<ul style="list-style-type: none"> <li>• Escuela de Ingeniería de São Carlos</li> <li>• Escola Politécnica</li> </ul>
9	Odontología	<ul style="list-style-type: none"> <li>• Facultad de Odontología</li> </ul>

El proyecto fue finalizado en 2001 y los aspectos gerenciales de este proceso fueron documentados y diseminados en diversos artículos (LIMA,1994;1996;1998;2002;2004).

## 2.1 CONSTRUCCIÓN DEL VOCABULARIO CONTROLADO DE SIBI/USP: ASPECTOS METODOLÓGICOS

La elaboración de un Vocabulario controlado requiere apoyo teórico, metodológico y técnico específicos. Un instrumento consistente y dinámico, que pueda ser actualizado de forma criteriosa, deve presentar una estructura de relaciones lógico-semánticas explícitas entre las áreas, subáreas y la terminología propiamente dicha. Requiere, también, reglas de utilización compartidas. Es extensamente aceptada la idea de que la Terminología fundamenta las operaciones de selección y control de las unidades de un vocabulario controlado.

De acuerdo con diversos especialistas, tales como Dubuc (1999), Cabré (1993) y Sager (1993), la Terminología es una ciencia interdisciplinaria que importa conceptos y elementos de otras disciplinas para objetivar su campo de estudio. Para Dubuc (1999), la Terminología permite identificar y analizar el vocabulario de una determinada especialidad y, si es necesario, crear y normalizar términos en situaciones concretas de funcionamiento.

Así, la Terminología estudia, teóricamente, los términos y sus respectivos conceptos, los sistemas de conceptos y su representación, teniendo, por lo tanto, tres funciones básicas (BARROS, 2004):

- *conceptual o cognitiva*: relacionada al análisis y descripción de la Terminología y armonización y organización de los dominios, por medio de la sistematización de los términos;
- *comunicativa*: relacionada a la transferencia del conocimiento, por medio de la actualización de las terminologías disponibles en los discursos técnico-científicos;
- *de identidad lingüística*: relacionada al rescate de lenguas en proceso de extinción, resaltando la identidad específica de individuos nacionalmente o regionalmente.

En este sentido, independientemente de sus características funcionales, el enfoque descriptivo y prescriptivo o normativo se hacen presentes. El primero, el descriptivo, se preocupa con la colecta de datos y la descripción de los términos y sus definiciones para la construcción de diccionarios y glosarios. El carácter prescriptivo o normativo, a su vez, promueve la normalización de los términos, fundamentada por la Teoría General de la Terminología (TGT) de Eugen Wuster (BARRROS, 2004).

TGT tiene por objetivo la normalización del lenguaje de especialidad, eliminando la imprecisión, la diversificación y la polisemia de los términos. Su aspecto prescriptivo o normativo se basa en la naturaleza de los conceptos, que mantiene relaciones de tipos diversos para la formación de una estructura conceptual de un área de conocimiento.

Tálamo, Lara y Kobashi (1992, p. 1999) afirman que:

“[...] cabe a la terminología, de este modo, operar el nivel sintáctico-semántico, produciendo terminologías específicas de acuerdo con el estado del arte de cada campo considerado. Tales repertorios o listas de términos especializados de un dominio particular son acompañados de definiciones que remiten el término a su referente [...]”.

En TGT, la relación término-concepto es fundamental, siendo el término considerado como entidad autónoma, constituyente de un lenguaje construido, perteneciente a una única área del conocimiento.

El carácter descriptivo y comunicativo de la Terminología, a su vez, es presentado por la Teoría Comunicativa de la Terminología (TCT), de Maria Teresa Cabré. En esta teoría, los términos son unidades léxicas que representan conceptos técnicos y científicos de la lengua general (CABRÉ, 2005). Así, una unidad léxica pasa a constituirse como término, de acuerdo con su uso, en contexto de comunicación específica.

La interdisciplinaridad y la multidimensionalidad son considerados por la TCT como elementos de suma importancia, teniendo en vista la

diversidad de visiones que los especialistas pueden tener de las unidades léxicas. De este modo, un concepto puede integrar la estructura conceptual de diversas disciplinas.

Un lenguaje documental puede, fácilmente, transformarse en nomenclatura. Como afirma Lara, la ausencia de definiciones puede llevar al predominio de la equivalencia léxica, proceso que transforma un lenguaje documental en una nomenclatura (Lara, 1993, p.68). Se puede afirmar, así, que la información documental representada por unidades de una nomenclatura se torna ambigua porque, al ser alterados los vínculos de significación, se abren, simultáneamente, infinitas posibilidades de enunciación y de interpretación y, por lo tanto, de reconocimiento de la información.

El Vocabulario Controlado de SIBi/USP fue elaborado con base en las concepciones de TGT, de Wuster, presentándose como un vocabulario prescriptivo. Se buscó, con esa medida, garantizar la univocidad, la no-ambigüedad de los términos, la homogeneidad del conocimiento especializado. Para eso, los términos ambiguos fueron definidos, utilizándose una ficha terminológica (*figura 1*) en la cual son registrados los aspectos semánticos y las fuentes de los términos.

Figura 1– Ficha terminológica utilizada para la definición de términos ambiguos del Vocabulario Controlado de SIBi/USP.

<b>FICHA TERMINOLÓGICA</b>	
TÉRMINO:	
DEFINICIÓN:	
SINÓNIMOS:	
FUENTE BIBLIOGRÁFICA:	
FECHA:	

Del punto de vista estructural, el Vocabulario USP es un lenguaje documental monohierárquico. Su Macroestructura se presenta en el siguiente gráfico.

**Figura 2 – Macroestructura del Vocabulario Controlado de SIBi/USP**

<b>Grandes Áreas</b>	<b>Áreas</b>
CA100 CIENCIAS AGRARIAS	CA110 AGRONOMÍA CA120 INGENIERÍA DE PESCA
CB200 BIOCIENCIAS	CB210 BIOLOGÍA CB220 BOTÁNICA CB230 INMUNOLOGÍA CB240 MICROBIOLOGÍA CB250 ZOOLOGÍA
CB300 SALUD	CB310 EDUCACIÓN FÍSICA Y DEPORTE CB320 ENFERMERÍA CB330 MEDICINA CB340 NUTRICIÓN CB350 ODONTOLOGÍA CB360 PSICOLOGÍA CB370 SALUD PÚBLICA CB380 FARMACIA Y COSMETOLOGÍA CB390 FONOAUDIOLOGÍA
CB400 SALUD ANIMAL	CB410 MEDICINA VETERINARIA CB420 ZOOTECNIA
CE500 CIENCIAS EXACTAS	CE510 ASTRONOMÍA CE520 FÍSICA CE530 GEOCIENCIAS CE540 GEOFÍSICA CE550 MATEMÁTICA CE560 QUÍMICA
CE600 CIENCIAS EXACTAS APLICADAS	CE610 CIENCIA DE LA COMPUTACIÓN CE620 INGENIERÍAS CE630 ESTADÍSTICA Y PROBABILIDAD CE640 METEOROLOGÍA
CH700 CIENCIAS HUMANAS	CH710 ADMINISTRACIÓN, ECONOMÍA, ECONOMÍA DOMÉSTICA Y CONTABILIDAD CH720 ARQUEOLOGÍA, MITOLOGÍA Y PREHISTORIA CH730 ARQUITECTURA, PLANEAMIENTO TERRITORIAL URBANO Y HABITACIÓN CH740 ARTES Y COMUNICACIÓN CH750 CIENCIA DE LA INFORMACIÓN Y MUSEOLOGÍA CH760 DERECHO, FILOSOFÍA, RELIGIÓN, CIENCIAS SOCIALES Y CIENCIA MILITAR CH770 EDUCACIÓN, ÓCIO, RECREACIÓN CH780 HISTORIA, HISTORIA DE BRASIL Y GEOGRAFÍA CH790 LINGÜÍSTICA, LENGUAS, LITERATURA Y TEORÍA LITERARIA

Este Vocabulario abarca las áreas del conocimiento inherentes a las actividades de enseñanza, investigación y extensión de la Universidad de São Paulo, y está constituido por descriptores autorizados para indexación, “no-autorizados”, que operan como remisivas, y los eslabones “falsos”, los cuales apenas agrupan términos más específicos.

### 3 IMPLEMENTACIÓN Y GESTIÓN DEL VOCABULARIO CONTROLADO DE SIBI/USP

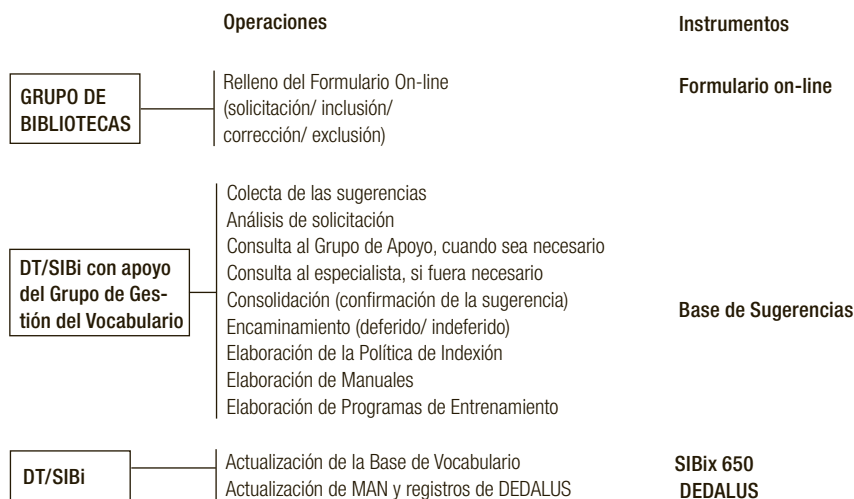
El Vocabulario Controlado de SIBi/USP comenzó a ser utilizado en el Banco de Datos Bibliográficos de USP - DEDALUS-, en 2001. Inicialmente vehiculado en CD-ROM, su formato actual es de base de datos, lo que permite su consulta, por la Internet, por el usuario final, a partir de la dirección *web* <http://www.usp.br/sibi> . Para fines de indexación, por la Intranet, el Vocabulario esta dotado de interface específica, el software SIBix 650 versión 1.6.1, que presenta las funcionalidades necesarias para indexar. Está constituido por una Tabla Principal, en orden alfabética y jerárquica, y de cuatro Tablas Auxiliares: Calificadores, Geográfica, Género/Forma y Profesiones, en orden alfabética.

Su implementación en DEDALUS fue antecedida del entrenamiento de los bibliotecarios indexadores y de los bibliotecarios de referencia, actividad fundamental para asegurar la consistencia de la indexación y de la recuperación de la información. Después de tres años de uso, fue creada una instancia de gestión del vocabulario. En efecto, un instrumento de indexación debe ser permanentemente actualizado para que no se cristalice, aspecto que contribuye para que él pierda la capacidad de tratar y recuperar información. Un manual de indexación “Manual de indexación con el uso del Vocabulario Controlado USP” orienta el proceso de indexación y la elaboración de estrategias de búsqueda. Así, los indexadores son también instrumentalizados para evaluar, periódicamente, los resultados de su trabajo de constitución de la base DEDALUS.

Actualmente, este Vocabulario es mantenido y administrado por el Grupo para Gestión del Vocabulario Controlado de SIBi/USP, organizado de la siguiente forma: 03 bibliotecarios representantes de las

bibliotecas pertenecientes al área de Ciencias Biológicas; 03 bibliotecarios representantes de las bibliotecas pertenecientes al área de Ciencias Exactas y 03 bibliotecarios representantes de las bibliotecas pertenecientes al área de Ciencias Humanas, 03 bibliotecarios y 01 analista de sistemas del Departamento Técnico del SIBi (DT/SIBi). El equipo cuenta con la asesoría académica de un docente del Departamento de Biblioteconomía y Documentación de la Escola de Comunicações e Artes de la Universidade de São Paulo, según lo esquematizado en la *Figura 3*.

Figura 3 – Mantenimiento y Gestión del Vocabulario Controlado de SIBi/USP: flujo de trabajo



La actualización es hecha a través de la base de sugerencias on-line “Sistema de Sugerencias del Vocabulario Controlado” la cual dispone formularios para sugerencias de inclusión, alteración y exclusión de términos del Vocabulario. Esas sugerencias son colectadas y analizadas por el Grupo de Gestión y cuando son pertinentes implementadas en la base de datos del Vocabulario Controlado de SIBi/USP. De esta manera, se asegura la comunicación permanente entre los responsables por la actualización del vocabulario.



La estructura y responsabilidades atribuidas a cada segmento ha garantizado la dinámica del vocabulario. La dimensión del vocabulario, en lo que se refiere a la implantación (33.528 términos) y a su dimensión actual (42.945 términos), certifican el trabajo permanente de actualización del lenguaje .

#### 4 CONSIDERACIONES FINALES

La construcción, implementación y gestión de vocabularios controlados son tareas arduas que requieren un largo tiempo y personal especializado, así como recursos financieros considerables. Así que, se espera, con esta comunicación, crear condiciones para compartir la experiencia. Se puede, de esta manera, contribuir para el mejoramiento de los sistemas de recuperación de la información, en general.

El uso de un vocabulario controlado común amplía, igualmente, las posibilidades de compatibilizar bases de datos de diferentes instituciones, alcanzándose con esto, el mejor aprovechamiento de los repositorios públicos de información científica y tecnológica. Pueden ser superados, de ese modo, los problemas de interoperabilidad, en nivel de recuperación temática de la información, que caracteriza a la gran mayoría de las bibliotecas digitales.

La validación de un vocabulario controlado integra el proceso de gestión de un sistema de información. En este sentido, en el ámbito de USP, son hechas evaluaciones institucionales periódicas, de consistencia de la indexación y de los resultados de búsquedas. Estas actividades son realizadas bajo la forma de talleres, de los cuales participen los indexadores y los bibliotecarios de referencia. Un modelo metodológico de validación fue elaborado especialmente para evaluar la representación y la recuperación de la información a través del uso del Vocabulario por los indexadores y por los bibliotecarios de referencia de SIBi/USP (LIMA, 2007).

Los primeros resultados obtenidos han suministrado indicadores importantes de los problemas que persisten en el sistema. Ellos servirán para fundamentar las formas de intervenir, de modo cada vez más seguro, en los procesos de mejoramiento de los dispositivos de información de USP. Los repositorios de información consistentes son, sin

ninguna duda, instrumentos importantes del ciclo de producción del conocimiento. La creación del Vocabulario Controlado de SIBI/USP es parte integrante, por lo tanto, del esfuerzo para colocar información al servicio de la actividad científica y para formar personal especializado para la sociedad.

## REFERENCIAS

- Barros, L. DE A. *Curso básico de terminología*. São Paulo: EDUSP, 2004. 285 p.
- Cabré, M. T. *La terminología: representación y comunicación: elementos para una teoría de base comunicativa y otros artículos*. Barcelona: Institut Uiversitari de Lingüística aplica [de la] Universitat Pompeu Fabra, 2005. 369 p.
- Cintra, A. M. M. et al. *Para entender as linguagens documentárias*. 2. ed. rev. e ampl. São Paulo: Polis, 2002. 92 p.
- Dubuc, R. *¿Qué es la terminología? Manual de terminología*. Traducción de Ileana Cabrera. 3. e. ed. cor. e actual. Providencia: Ril Ed., 1999.
- Lara, M. L. Linguagen documentária e terminologia. *Transformação*, Campinas, v. 16, n.3, p.231-240, set./dez. 2004.
- Lima, Vânia Mara Alves; Carvalho, Márcia Maria de; Pilnik, Márcia; Yamashita, Marina M; Prati, Suely Cafazzi; Boccato, Vera Regina Casari. Aprimoramento da Lista de Assuntos USP. In: Seminário Nacional de Bibliotecas Universitárias, 8, 1994, Campinas. *Anais*. Campinas: Biblioteca Central- Unicamp, 1994. p. 283.
- Lima, Vânia Mara Alves; Carvalho, Márcia Maria de. ; Yamashita, Marina M ; Pilnik, Márcia ; Prati, S. C. ; Boccato, Vera Regina Casari. Atualização da lista de assuntos USP:

- compatibilização de linguagens documentárias. *Ciência da Informação*, Brasília, v. 25, n. 2, p. 177-181, maio/ago. 1996.
- Lima, Vânia Mara Alves; Kobashi, Nair Yumiko. Metodologia de desenvolvimento, implantação e gerenciamento do Vocabulário Controlado USP. In: XIII SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, *Anais*. 2004, Natal. 2004.
- Lima, V.M.A. *Proposta para a construção e avaliação de linguagens documentárias: uma aplicação ao Vocabulário Controlado do SIBi/USP*. Projeto de Pesquisa. Departamento de Biblioteconomia e Documentação, ECA/USP, 2007.
- Lima, Vânia Mara Alves; Yamashita, Marina M; Pilnik, Márcia; Boccato, Vera Regina Casari; Camargo, Cibele Araujo. Vocabulário USP: um instrumento para integração dos processos de representação e recuperação da informação em bibliotecas universitárias. In: X SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 1998, Fortaleza, CE. *Anais eletrônicos*. Tectreina, 1998. p. 1-5.
- Lima, Vânia Mara Alves; Kobashi, Nair Yumiko; Imperatriz, Inês Maria de M. Vocabulário controlado USP: desenvolvimento, implantação e gerenciamento. In: CONGRESSO INTERNACIONAL DE ARQUIVOS, BIBLIOTECAS, CENTROS DE DOCUMENTAÇÃO E MUSEUS, 2002, São Paulo. *Integrar: 1. Congresso Internacional de Arquivos, Bibliotecas, Centros de Documentação e Museus: textos*. São Paulo: Imprensa Oficial del Estado, 2002. p. 225-235
- Tàlamo, M. de F. G. M.; Lara, M. L. G. de; Kobashi, N. Y. Contribuição da Terminología para a elaboração de tesauros. *Ciência da Informação*, Brasília, v. 21, n. 3, p. 197-200, set./dez. 1992.



## Algunas grietas semánticas en la recuperación de información: una perspectiva deconstructiva para una solución pragmática

RAFAEL IBARRA CONTRERAS

*Dirección General de Bibliotecas*

*UNAM, México*

**A**ctualmente, prevalece la problemática semántica subyacente en la recuperación de información (RI) de bases de datos diseñadas en idioma inglés (BDI) empleadas por nativo-hablantes del idioma español (NHE), lo que representa una inversión lamentablemente considerable en tiempo y en dinero empleados por la falta de resultados rentables.

Seguramente todos los que nos hemos ocupado en buscar y recuperar información en fuentes electrónicas, como las BDI, hemos experimentado cierto grado de agrietamiento, ya sea porque encontramos demasiada, poca, o ninguna información para desarrollar nuestra investigación, saciar nuestra curiosidad o satisfacer nuestras necesidades respecto al quehacer académico o comercial.

En algunas ocasiones, los resultados extremos nos han conducido a prestar poca atención sobre las causas o razones ante la demasía, irrelevancia o de la ausencia de datos; aunque también hemos dedicado algunos momentos a reflexionar el por qué de los resultados obtenidos, ¿qué hice mal?, ¿qué más debo hacer?, preguntas que hacen disparar especulaciones igual de extremas como, “seguramente que soy pionero en el tema”, “necesito cambiar de tema de mi investigación porque ya hay demasiado trabajo realizado”, entre otras.

Esta clase de rutinas produce un efecto atroz, ante la insuficiencia de soluciones razonadas y prácticas por parte de los proveedores de las BDI en sus sistemas de recuperación de información (SRI), o a falta de propuestas de investigación por parte de los profesionales en las ciencias de la información NHE, es común que, tanto usuarios como bibliotecarios, se desarrollen paulatinamente dentro de un estado rústico de “naturaleza profesional” en que ellos se desenvuelven: un ámbito académico autosuficiente y aislado uno del otro, sin importar los defectos o fallas desde los que se trabaja.

Cada uno de los elementos de este escenario nos conduce a la costosa aparición de grietas semánticas, por las que se cuele el sentido, y que exige de la propuesta de una aproximación remedial ahora y, preventiva después, como la que se presenta en este trabajo. Tal tarea resultaría arriesgada si se intentara repetir o mejorar los esfuerzos ya hechos. Este no es el caso.

Desde algunos años antes de iniciar la presente propuesta, se han observado diversos acercamientos para dar respuesta a la entropía y redundancia como producto común de las BDI. La gran mayoría de ellos han sido realizados para resolver el problema desde un punto de vista del nativo hablante del idioma inglés (NHI), cuyos enfoques están centrados desde el punto de vista del SRI. Lo anterior se mostrará con literatura revisada en la BDI Library and Information Science Abstracts<sup>1</sup>

Los esfuerzos en investigación de NHE, documentados en el catálogo de tesis TESIUNAM, ofrecen aspectos descriptivos en cuanto a la modificación del ambiente en el que se trabaja, inducción de aproximación a los servicios que se ofrecen en general, análisis de necesidades, derechos de autor, medida de satisfacción, instrucción bibliográfica, conducta en la biblioteca, hábitos de lectura, personal y usuarios, evaluación de revistas, procesos de clasificación, “desarrollo de búsquedas efectivas”, descripción de bases de datos en español, normas ISO en diseño de bases de datos, diseño y elaboración de bases de datos, estructuración semántico-pragmática en sistemas de recuperación de información enfocado a usuarios NHE.

---

1 Library and Information Science Abstracts. <http://lisa.com>

Es con base en estos argumentos que se decide ofrecer y discutir una perspectiva deconstructiva y una alternativa de solución pragmática que comprende ilustrar las “oposiciones” existentes, inteligencia artificial vs sensibilidad humana, en el ámbito de RI en BDI empleadas por NHE, a través de un esquema que ilustra un breve examen de la interacción del lenguaje empleado por usuarios y bibliotecarios. El objetivo es contribuir a disminuir la vasta brecha de malentendidos desde una postura lingüística crítica, deconstructiva, por medio de la ilustración de la ruta crítica típica infructuosa y la ruta ideal fructífera en la RI. Se describe una serie de factores lingüísticos y humanos opuestos y sus impactos en la RI, iniciando por algunas consideraciones del idioma inglés.

## EL IDIOMA

Hoy en día, como académicos NHE, enfrentamos los impresionantes resultados y confusiones de la tecnología, relacionada con los campos de la bibliotecología y las ciencias de la información, conocidos como BD. También tenemos que manejar la *lingua franca* en que este tipo de tecnología se exige: el idioma inglés. En este momento, los datos académicos son presentados y susceptibles de buscarse y recuperarse por impresionantes grupos de prestigiosas BD, que incluyen más de dos o tres mil títulos de *journals* arbitrados, además de memorias de congresos, coloquios, simposio y libros.

No obstante, si nuestro inglés universitario es suficientemente “bueno” para obtener la información que deseamos, y contamos con bibliotecarios comprometidos que nos facilitan las herramientas y técnicas esenciales y, por su parte, los SRI ofrecidos por los proveedores de las BDI son “efectivos”, así lo dicen, ya que se emplean herramientas diversas y eficientes para la RI, ¿por qué aún enfrentamos altos niveles de frustración al tratar de recuperar la información que necesitamos?

Una de las grietas más conocidas por todos es la del idioma. ¿Por qué representa una grieta? En primer lugar, el idioma inglés no es nuestra lengua materna, pero tenemos que emplearla para obtener información para nuestro quehacer académico y es precisamente la

falta del manejo apropiado, lo que ocasiona desviaciones semánticas. Por ejemplo, en los seis niveles de comprensión de idioma se observan las ponderaciones conducentes al éxito o fracaso de nuestra búsqueda de información. Tomemos como caso que un usuario está interesado en la búsqueda de *bolsa de valores*, ilustrado en un breve análisis lingüístico de seis niveles, iniciado en español y finalizado, inevitablemente en inglés :

Nivel	Variedad de grafemas	Variedad sintáctica
1. Morfológico	BOLSA DE VOLARES BALSA DE VALORES BOLSA DE VALORES	S+Prep+S
2. Léxico	ÍNDICE DE COTIZACIONES	S+Prep+S
3. Sintáctico	ÍNDICE DE COTIZACIONES	S+Prep+S
4. Semántico	FINANZAS	S
5. Discursivo	INDICADOR BURSÁTIL	S + ADJ
6. Pragmático	STOCK EXCHANGE	S

En el nivel morfológico se presentan algunas posibilidades de incurrir en un error semántico por la distribución de los grafemas, considerando el lenguaje natural; en el nivel léxico se presenta, el cambio ortográfico por la naturaleza del vocabulario controlado; el nivel sintáctico presenta el orden de los componentes, un sustantivo, una preposición, un sustantivo; en el nivel semántico, se observa la reducción de tres elementos a sólo uno, y el cambio de referente; en el nivel discursivo se da un cambio en el número de elementos y de referente, compuesto por dos elementos; finalmente, en el nivel pragmático, se toma como base que, si se trata de finanzas y la base de datos recupera principalmente información en inglés, se deberá modificar la sintaxis y el escribir con un vocabulario controlado.

Hay que observar que, a pesar de tratarse de dos términos, ambos funcionan como un sólo sustantivo, pero con la sintaxis indicada, primero *stock*, después *exchange*. En un caso extremo, el usuario podría sugerir, con base en un mal manejo de la lengua inglesa y de un diccionario español/inglés: “bag of values”, porque no se puede negar que *bag* significa bolsa y *values*, valores.



No obstante lo anterior, ¿por qué debemos recuperar información de las bases de datos en inglés? Algunas razones al respecto las desglosa Dungworth (1978:1), “el idioma inglés es el más empleado en el mundo por tres razones 1) porque ocupa la segunda posición en el rango mundial y sólo es superado por los hablantes de chino mandarín, 2) porque es el más empleado en el control de tráfico aéreo, almacenamiento y recuperación de información y 3) está mucho más difundido que cualquier otro y está considerado, según el reporte del Consejo Británico (1974-1975) como el idioma líder para la comunicación internacional”.

Un reporte de 2006 indica que el idioma inglés es la lengua franca de la ciencia y la tecnología, de la experimentación y los descubrimientos y su influencia ha sido tal que:

Los *journals* en muchos países han cambiado de lengua, desde la II Guerra Mundial, de editarse en su lengua vernácula a publicarse en inglés. Gibbs (1995) describe cómo el *journal* médico mexicano *Archivos de Investigación Médica* mudó al inglés: primero se publicaron los resúmenes en inglés, después se ofrecieron traducciones al inglés de todos los artículos, finalmente se contrató a un editor norteamericano, y únicamente se aceptaron artículos en inglés y cambió su nombre a *Archives of Medical Research*. Este cambio de idioma es muy común en otros lugares (Graddol, D. 1997: 9).

Consecuencia de lo anterior es la tendencia de los científicos a buscar y recuperar la información que requieren las bases de datos que nos ocupan. Actualmente, la UNAM ofrece a su comunidad más del 85% de este tipo de bases, de las que en el año 2006, reporta la Biblioteca Digital de la UNAM<sup>2</sup>, se hicieron 55,666 búsquedas de información en las diez bases de datos más consultadas. Nueve de ellas están dispuestas para su consulta en idioma inglés y el total de sus consultas fue de 54,533, que representa el 98%; la última base de datos, dispuesta en español, tuvo 1,133 consultas, lo que representa el 2%.<sup>3</sup> Por su parte, la Internet Society ha reportado sus más recientes hallazgos en

---

2 Cifras tomadas de <http://bidi.unam.mx> , sección Bases de datos. Los más consultados. [accedido 15 mayo, 2007].

3 Ver anexo 1 al final de este trabajo.

una encuesta sobre el idioma de las páginas web que hay en el mundo, siendo el 84.3% en inglés y el 1.2% en español.<sup>4</sup>

Lo anterior nos hace pensar en la necesidad de planeación del idioma para el cambio social, de acuerdo con Cooper (1989:182), “el idioma inglés conduce a reducir la grieta entre las variedades hablada y escrita y para incrementar el acceso a la educación formal, además, la planeación lingüística es una herramienta al servicio de tantas y diferentes metas como la modernización económica, integración nacional, liberación nacional, igualdad social, racial, y el mantenimiento de las elites y su reemplazo por nuevas”. Hasta aquí, las consideraciones del idioma. Pasemos a la revisión de la literatura.

## REVISIÓN DE LA LITERATURA EN TESIUNAM

Se efectuó una revisión de la literatura de tesis en el catálogo TESIUNAM<sup>5</sup> considerando diferentes términos relativos a la *recuperación de información y lingüística* en título; y con los términos *bibliotecología y lingüística* en carrera.

Se seleccionaron aquellas tesis cuyos títulos indicaran una relación estrecha con el propósito de esta investigación y, posteriormente, se revisaron las tablas de contenido de 25 de ellas, y se leyeron los capítulos relacionados con el idioma inglés para extraer la información que comprobara una línea de investigación al respecto. Sólo se obtuvo el siguiente párrafo relativo al interés de esta investigación (Balboa, A. 2005:115) *En este estudio se ve claramente que ya es bastante difícil para bibliotecarios y usuarios el entenderse en español; luego entonces, una barrera de idioma, dificulta mucho más la comunicación y es más frustrante*<sup>6</sup>. Por cuestiones de espacio, sólo se hace referencia a la tesis de la que se tomó el párrafo referido.

El resto de los resultados fue el siguiente:

---

<sup>4</sup> Ver anexo 2 al final de este trabajo

<sup>5</sup> <http://dgb.unam.mx> TESIUNAM [accedido 20 mayo, 2007].

<sup>6</sup> Ana C. Balboa Magaña, La comunicación interpersonal entre el bibliotecario y el usuario en la entrevista de referencia : el proceso de mediación. 2005. (Maestría en Bibliotecología y Estudios de la Información)-UNAM, Facultad de Filosofía y Letras.

## *Representación y recuperación de información*

Palabras / campos:	# Registros
Recuperación(en Tit). 1 Lingüística (en carrera)	SEIS
Lingüísticas (en Tit). 2 bibliotecología (en carrera)	CERO
Lingüística (en Tit). 3 bibliotecología (en carrera)	UNO
Búsqueda recuperación 4 información (en Tit).	CUATRO
CD ROM (en Tit). 5 bibliotecología (en carrera)	DOS
Bases de datos (en Tit). 6 bibliotecología (en carrera)	OCHO
Usuarios (en Tit). 7 bibliotecología (en carrera)	TREINTA Y UNO
Usuario (en Tit). 8 bibliotecología (en carrera)	DOS
9 Habilidades informativas (en Ti)	UNO

Por otro lado, también se realizó una búsqueda en la bases de datos LISA, en línea con el siguiente resultado. Cabe mencionar que, al igual que en las tesis, sólo se hará referencia a los registros de los que se tomó una referencia sólida:

### LISA: Library and Information Science Abstracts<sup>7</sup>

Terminos empleados	Número de registros	Notas
information retrieval	22362	Se restringió la búsqueda agregando "Spanish"
(information retrieval) and Spanish	240	Se seleccionaron 3 registros de 22 <sup>8</sup> , que abordan el idioma español con propósitos adecuados este estudio (5, 10 y 12)
(information retrieval) and spanish and linguistics	1	Este registro representa un ejemplo de irrelevancia <sup>9</sup> .
(information retrieval) and spanish and deconstruction	0	
TI=(information retrieval spanish)	0	

7 <http://www-ca1.csa.com> [accedido el 6 de julio, 2007]

8 Ver anexo 3 al final de este trabajo.

9 ERIC: multinacional /multidisciplinario. Donal P. Ely. Revista Interamericana de

## LA PERSPECTIVA DECONSTRUCTIVA

La aparente facilidad de recuperar información de un texto (artículo, libro, reporte, etc.) en bases de datos sin ningún otro esfuerzo que ingresar los términos necesarios en el teclado de una computadora en que reside un gran cantidad de datos, contrasta con el nivel de frustración que los usuarios experimentamos al comprobar que no es así. Aparentemente, la intervención humana todavía se requiere de manera física, crítica y reflexiva.

Una perspectiva deconstructiva podría ayudar a la creación de un peldaño que ayude, en la medida de lo posible, a acortar la distancia entre la satisfacción y el infortunio, por medio de la irrupción de deconstruir el sistema de oposiciones conceptuales como son lo literal vs lo metafórico, el habla vs la escritura, lo inteligible vs lo sensible.

Es pertinente la propuesta de un enfoque que permita explorar las tensiones y contradicciones no sólo del (o los) textos contenidos en una BD y su significado, sino de las relaciones lingüísticas que se realizan en la interacción entre los usuarios humanos y la máquina con que se opera la BD y, necesariamente también, su significado.

De manera sencilla, si partimos de la siguiente hipótesis: Si poseo buen manejo del idioma inglés, a pesar de no ser mi lengua materna, si me asiste un bibliotecario profesional y consulto una BD apropiada a mis necesidades, encontraré el (los) texto(s) con la información que necesito.

Más del 50% de usuarios estaría en desacuerdo con la hipótesis anterior porque sencillamente han experimentado lo anterior con resultados desafortunados. En este punto es pertinente comentar aún para los nativo-hablantes del idioma inglés, la experiencia de frustración a causa de la ambigüedad en la búsqueda y recuperación de información se hace presente en el momento de revisar los resultados arrojados: (Jaffe, J. 1988b:759) "Los usuarios quedaban confundidos con las citas sin relevancia aparecidas en las búsquedas booleanas, aunque todos los elementos de las mismas estaban representados en las citas". Esta afirmación se hizo después que estudiantes del Sweet Briar Co-

llage, en Virginia, consultaron una base de datos. Lo anterior revela la existencia de una grieta, de un malentendido que se puede interpretar como “ la imposibilidad de aislar un sentido originario principal en el centro de una construcción conceptual o el conjunto de una obra” (Peñalver, P. 1989:15).<sup>10</sup>

En la actualidad, sabemos que las técnicas genéricas empleadas para RI son, a grosso modo, dos: la comparación de las palabras de la búsqueda contra el índice de la BD de que se trate y la de atravesar la BD con la ayuda de vínculos de hipertexto o de hipermedios, por medio de operadores booleanos. Asimismo, el inevitable uso de lenguaje natural y el vocabulario controlado, son aún insuficientes para evitar la ambigüedad que conlleva el lenguaje humano para obtener la información deseada en una BD, en una clara oposición de la inteligencia artificial y la sensibilidad humana.

Este escenario sugiere considerar las limitantes de los sistemas de recuperación para reconocer los significados de las imágenes, idiomas o conceptos en la experiencia *deconstructiva* de Jeff Bezos, ejecutivo en jefe de Amazon.com, quien creó un servicio en-línea que incluye a humanos, denominado Amazon Mechanical Turk, fenómeno que en sus palabras describe de la siguiente manera: “Normalmente, un humano hace una petición a una computadora, y la computadora hace su tarea. Pero las inteligencias artificiales como la Mechanical Turk invierten todo eso. La computadora tiene una misión que es fácil para el humano, pero extraordinariamente difícil para la computadora. De este modo, en vez de solicitar el servicio de computadora para realizar la función, se pide a un humano.” (Pontin, J. 2007: s/p)

Los diversos esfuerzos existentes para obtener la información deseada que van desde realizar procesos sometidos al azar y que son objeto de análisis estadísticos hasta los análisis gramaticales empleados en la minería de datos, hace que los usuarios finales estemos más esperanzados a obtener menos grados de frustración que los actuales. No obstante, dado que el estudio, los descubrimientos y las aplicaciones tecnológicas en los SRI no se detienen para indexar documentos

---

10 En J. Derrida, *La deconstrucción en las fronteras de la filosofía : la retirada de la metáfora*.

académicos con formatos diferentes a los de puro texto, como imagen, video, audio y sus combinaciones, nos reduce la esperanza de obtener información con mayor facilidad.

## LA SOLUCIÓN PRAGMÁTICA

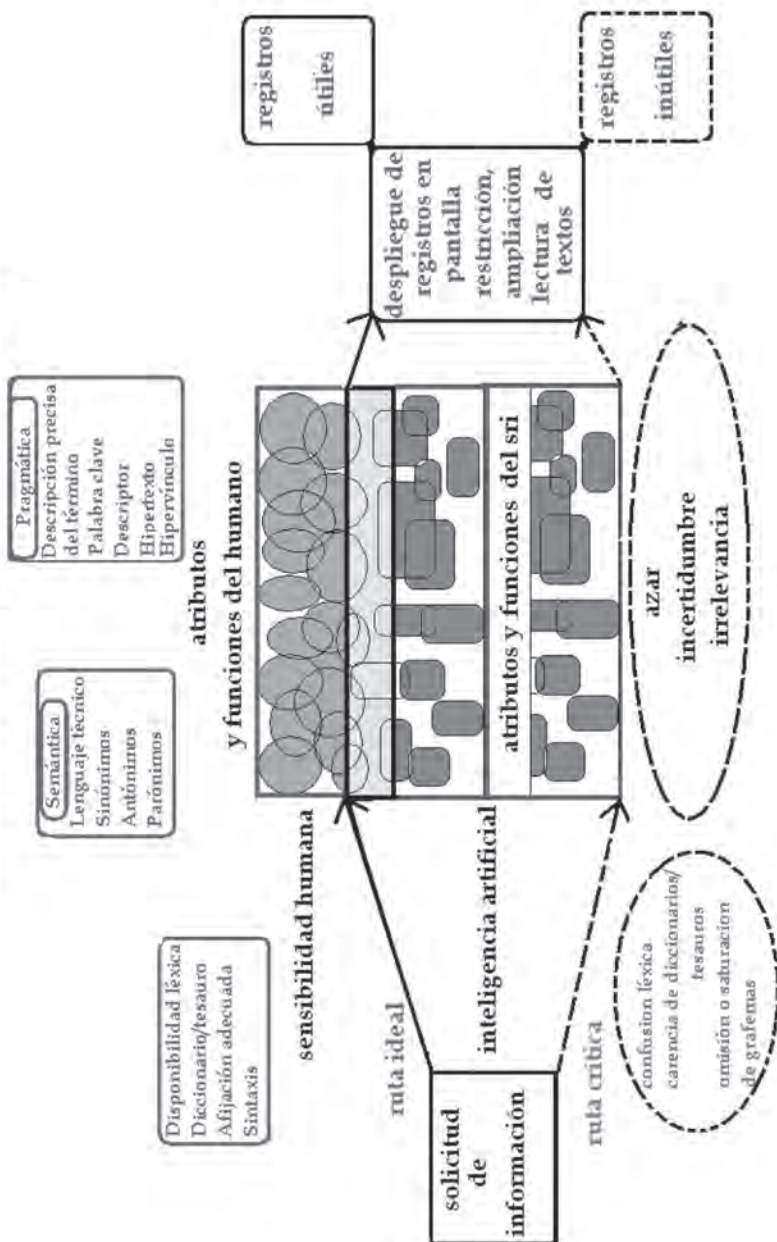
Tomando en cuenta los argumentos anteriores, ya se está en condiciones de sugerir una solución pragmática con base en una serie de operaciones para resolver el problema de la incertidumbre: el algoritmo lingüístico, marcado con los cuadros verdes — en el orden descrito — y que se apoya en el plano de la sensibilidad humana, en oposición a la inteligencia artificial, desde nuestra perspectiva deconstructiva, acusada líneas atrás.

El punto de partida es la solicitud de información con dos rutas a seguir, la ruta crítica es la acostumbrada por los NHI y por los NHE, y depende, en gran medida, de la inteligencia artificial dispuesta en los SRI de las BDI y que difícilmente resuelve los problemas de incertidumbre, a saber (Cfr. Allwood, J. 1995: 29)

- D) Si no se encuentra información...
  - es porque se carece de disponibilidad léxica
  - o porque está mal el orden de las palabras
  - o porque el (los) término(s) empleado(s) no es (son) equivalente(s) en significado
  - o porque el contexto no es el apropiado
  - o porque no existe en la base de datos que se esté empleando
- II) Si la información no es la deseada
  - es porque el lenguaje técnico no es el apropiado
  - o porque la ortografía no es la correcta
  - o porque la palabra clave/descriptor no es tal

En cuanto al espacio de la inteligencia artificial, es de esperar que los SRI de las BDI incluyan puntos de recuperación y campos referentes a imágenes, audio, video y sus posibles combinaciones.

Modelo algorítmico deconstructivo para la Recuperación de Información



## CONCLUSIONES

A lo largo de este trabajo se planteó la posibilidad de resolver la problemática de la RI en BDI, con base en una perspectiva deconstructiva lingüística, y centrada en NHE poniendo en evidencia las variedades existentes en seis niveles para la comprensión del idioma inglés en cuanto a los grafemas y a la sintaxis. Sin este matiz, difícilmente se podría entender la existencia de ángulos diferentes a los ya conocidos, aunque no sean lo únicos.

Se tomaron en cuenta tres aproximaciones a la resolución de la frustración que experimentan los usuarios: las realizadas por los tesis-tas egresados de la UNAM, en sus trabajos recepcionales, registrados durante los últimos años en el catálogo TESIUNAM; una revisión de la literatura que ofrece la base de datos LISA; y algunas consideraciones sobre las herramientas empleadas en los SRI de las BDI con el mismo resultado: insuficientes.

Asimismo, se ilustraron algunas circunstancias históricas y tecnológicas relativas a la influencia de la lengua franca de nuestros tiempos y que sirven para entender las tendencias y necesidades implícitas para los no NHI. Se describieron algunas actitudes de los involucrados en el manejo de las BDI y las interrogantes que surgen durante su que-hacer académico. Se quiera o no, estamos supeditados a acercarnos, invariablemente, al manejo adecuado del idioma inglés.

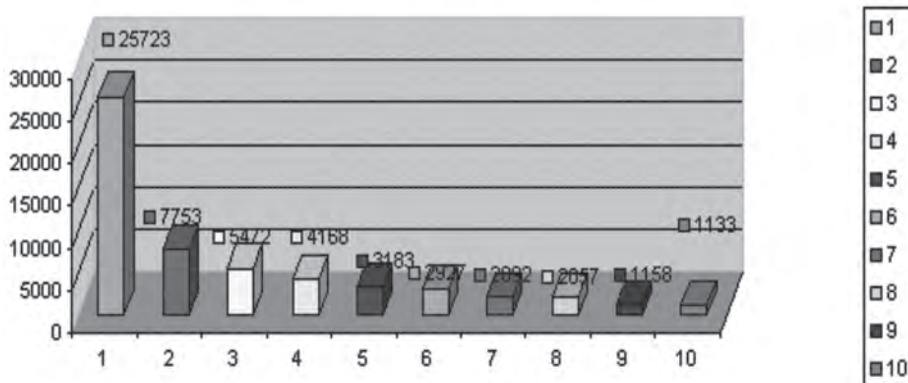
Finalmente, se describió un modelo que incluye un algoritmo lingüístico, y dos oposiciones presentes durante la ruta de búsqueda y recuperación de información, a saber, la sensibilidad humana y la inteligencia artificial, ponderando la ruta crítica, la ruta ideal y contrastando los atributos humanos y los de los SRI de las BDI. Se puede concluir que estimando la ruta ideal y las condiciones de incertidumbre, se podrán lograr mayores satisfacciones si únicamente recurrimos a lo que ofrecen los proveedores de las BDI.



ANEXO 1

Consultas	Bases de Datos más consultadas en 2006 <sup>11</sup>
25723	Academic Search Premier
7753	Elsevier Science
5472	PsycINFO
4168	WorldCat
3183	ProQuest Psychology Journals
2927	Science Citation Index
2092	MEDLINE
2057	Journals@Ovid Full Text
1158	EJS (Ebsco Host Electronic Journal Service)
1133	INFOLATINA

**Bases de datos más consultadas 2006**



11 El presupuesto total ejercido para las BDI fue de 21,184,382.94 (91%). Para las BD en español fue de 2,198,669.48 (9%). Información facilitada por la Secretaria de Servicios Especiales de la Dirección General de Bibliotecas de la UNAM.

ANEXO 2

	Language	Estimated % servers	
1	English	332,778	84.3
2	German	17,971	4.5
3	Japanese	12,348	3.1
4	French	7,213	1.8
5	Spanish	4,646	1.2
6	Swedish	4,279	1.1
7	Italian	3,790	1.0
8	Portuguese	2,567	0.7
9	Dutch	2,445	0.6
10	Norwegian	2,323	0.6

Table 18 Languages of home pages on the Web

12

ANEXO 3<sup>13</sup>

Registro 5 de 22

TI: Título: Methodologies for the development of visual interfaces of information retrieval: analysis and comparison.

AU: Autor: Herrero-Solana\*, Victor; Hassan, Yusef

Registro 10 de 22

TI: Título: Enfoques en torno al modelo cognitivo para la recuperación de información: análisis crítico. Approaches in the cognitive model of information retrieval: a critical analysis

AU: Autor: Vargas-Quesada, B; Anegón, F d M; Lobo, D O

Registro 12 de 22

TI: Título: Evaluación de fuentes de información en Internet: criterios de calidad. Evaluation of information sources on the Internet: criteria of quality

AU: Autor: Tomael, M I; Catarino, M E; Valentim, M L P; de Almeida, O F; da Silva, T E

---

12 (Graddol, D. 1997: 519)

13 [http://www-ca1.csa.com.pbidi.unam.mx:8080/ids70/advanced\\_search.php?SID=392426ee13476b789d3ee31e37387bd9&tab\\_collection\\_id=999](http://www-ca1.csa.com.pbidi.unam.mx:8080/ids70/advanced_search.php?SID=392426ee13476b789d3ee31e37387bd9&tab_collection_id=999)

## BIBLIOGRAFÍA

- Allwood, Jens (1995). *Dialog with a cooperative information system*. Sweden: Dept. of Linguistics. University of Göteborg.
- Derrida, J. 1989. Premisas y contextos de la desconstrucción, en *La desconstrucción en las fronteras de la filosofía : la retirada de la metáfora* / Jacques Derrida ; traducción de Patricio Peñalver.
- Dungworth, D. 1978 *The future of English as a World Language* en *Lebende Sprachen*. 23, 1, pp. 1-3.
- Ely, Donal 1985. P *ERIC: multinational/multidisciplinary* en *Revista Interamericana de Bibliotecología*; 8 (2) July Dec 85, 49-61.
- Graddol, David. 1997. *The Future of English?: A guide to forecasting the popularity of the English language in the 21<sup>st</sup> century*. The British Council. <http://www.britishcouncil.org/learning-elt-future.pdf> [accedido 9 julio, 2007]
- Ibarra Contreras, R. 1999. *Aprovechamiento y optimización de los recursos tecnológicos en la búsqueda y recuperación de información en CD-ROM basados en estrategias lingüísticas*. Tesis de maestría UACPyP - CELE. [http://pbidi.unam.mx/cgi-bin/ezpmysql.cgi?url=http://132.248.9.9:8080/tesdig/Procesados\\_1999/271153/Index.html](http://pbidi.unam.mx/cgi-bin/ezpmysql.cgi?url=http://132.248.9.9:8080/tesdig/Procesados_1999/271153/Index.html)
- Jaffe, J. 1988. *For undergrads: Infotrac MAGAZINE INDEX PLUS or WILSONDISC with Reader's Guide & Humanities Index* en *American Libraries*, Oct. 1988 pp.759-61.
- Liddy, Elizabeth D. 1999. *Whither Come the Words?* Paper presented at the CENDI Subject Analysis and Retrieval

***Primer Simposio Internacional sobre Organización...***

Working Group Conference “Controlled Vocabulary and the Internet,” September 29, <http://cendi.dtic.mil/presentations/liddy.PPT> [accedido 5 julio, 2007].

Pontin, J. 2007. *Artificial Intelligence, With Help From the Humans*. <http://www.nytimes.com/2007/03/25/business/yourmoney/25Stream.html?ex=1184731200&en=8d71530b581d5e00&ei=5070> [accedido marzo 25, 2007].

## Conocimiento de las Artes Plásticas y la Arquitectura mexicanas

ELSA BARBERENA BLÁSQUEZ  
*Facultad de Filosofía y Letras*  
*UNAM, México*

Como bibliotecaria de consulta la palabra conocimiento ha estado presente en mi actividad profesional y docente. Comunicar sentimientos, emociones y conocimientos ha sido siempre un aspecto esencial del comportamiento humano (Castillo Costa p.1).

Existen múltiples definiciones de conocimiento, desde las clásicas, representadas por Aristóteles y Platón, hasta las más recientes y pragmáticas que lo definen, como una experiencia, como un conjunto de valores, como información y “saber hacer” el cual sirve como marco para la incorporación de nuevas experiencias e información, y es útil para la acción ( Davenport y Prusak 1999). Escojo una definición táctica de conocimiento: Conocimiento es la capacidad para convertir datos e información en acciones efectivas.

De los ocho valores que (Gorman p.2) lista en su libro: democracia, mayordomía, servicio, libertad intelectual, privacidad, equidad de acceso, menciona el conocimiento en dos: compromiso a la alfabetización y al aprendizaje y racionalismo.

Propagar datos e información sobre las artes plásticas y la arquitectura mexicanas aumenta el conocimiento de la creación artística, expresión vital de los valores y sentimientos de una sociedad que piensa y que trasciende.

Mi labor ha sido registrar, compilar, organizar, conservar, utilizar y difundir información y conocimientos, por medio de las tecnologías de la información. El papel del bibliotecólogo debe orientarse a clasificar la información disponible en redes (Naumis p.461).

(Fernández de Zamora p.223) menciona la importancia de crear nuevo conocimiento que revalore el patrimonio bibliográfico y documental mexicano al utilizar las tecnologías de preservación y difusión: microfilmación para conservar, digitalización para difundir, páginas web, portales, discos compactos.

Las formas de investigar, y por tanto de generar y transmitir conocimiento, empiezan a modificarse con la presencia de las tecnologías de la información (Ramírez Leyva, p.476).

Mis objetivos se relacionan con la información en las diversas disciplinas y para ello la reconozco cuando se necesita, sé como está organizada, la localizo y la evalúo.

En el reconocimiento de la necesidad de la información están implícitas, según (Lau p.73) las competencias de los usuarios para alcanzar la capacidad de:

- Determinar la naturaleza y nivel de la información requerida.
- Usar y acceder a la información solicitada de manera eficaz y eficiente.
- Evaluar e incorporar la información a su bagaje de conocimiento y valores.

La información como medio transportador del conocimiento, desempeña un papel vital en el autoaprendizaje, por lo tanto las habilidades informativas son necesarias para que el aprendedor las desarrolle durante su estancia en las instituciones educativas. La facilitación del desarrollo de dichas habilidades es una función que le corresponde desempeñar a los expertos en información: los bibliotecarios (Lau p. 78).

La información es un insumo indispensable para los procesos educativos, puesto que los datos que describen un hecho o fenómeno son relevantes para los estudiantes y los profesores (facilitadores) en sus procesos de construcción del conocimiento (Lau, p. 53).

El conocimiento sobre las artes plásticas y la arquitectura mexicanas es limitado, además de existir una fuerte dependencia de infor-

mación internacional y poca presencia de información nacional. Se destaca el hecho de que, a pesar de vivir con una cultura tecnológica, en México existe una muy pobre vinculación entre estas disciplinas por una parte y el diseño de productos de información para el uso o consumo del usuario, por la otra. Razón por la cual se ha elaborado una base de datos sobre arte mexicano, MEXICOARTE, disponible a través de la Biblioteca digital de la Dirección General de Bibliotecas de la UNAM <http://bidi.unam.mx> (Barberena p.36).

Los requisitos para que las publicaciones ingresen a las bases de datos más reconocidas en el mundo son tan onerosos que su presencia resulta muy escasa (Rendón p.9).

Reconocí que se necesitaba organizar la información en español sobre estas disciplinas y elaboré un vocabulario controlado. Existe la necesidad de que los desarrollos lingüísticos se realicen con la participación de bibliotecólogos y especialistas en el ámbito temático del trabajo (Naumis p. 459).

Un vocabulario es un cuerpo de conocimiento representado por el lenguaje. Responde a la pregunta ¿Cómo hablamos o escribimos acerca de un tema específico? (Lanzi p.21-22)

Los términos que se encuentran en este vocabulario en particular:

Nombres de artistas y arquitectos con fechas de nacimiento y fallecimiento:

- Rivera, Diego, 1886-1952
- Barragán Morfín, Luis, 1903-1988

Nombres geográficos:

- Cholula, Pue.
- Palenque, Chis.

Nombres corporativos:

- Museo de San Carlos
- Universidad Nacional Autónoma de México

Nombres de objetos artísticos:

- Esculturas
- Pinturas

Temas:

-Alimentos

Véase: aguacate

-Indumentaria

Movimientos artísticos:

-Constructivismo

-Manierismo

El vocabulario es una base de conocimiento que incluye redes semánticas, que muestra enlaces y caminos entre conceptos y estas relaciones pueden hacer la recuperación más asertiva.

Emprendí la tarea de localizar los documentos, no solamente en material impreso sino electrónico y en bibliotecas especializadas. Un documento es el producto de una actividad humana encaminada a coadyuvar a mantener y cuidar la permanencia de datos e información, a fin de que una comunidad o un individuo puedan valerse de ésta para construir conocimientos y cumplir con los propósitos relativos a la supervivencia y bienestar de una comunidad. (Lafuente p. 478)

Incluí a las bibliotecas, cuya finalidad fundamental es servir de apoyo al desarrollo de planes y políticas para la producción, organización y circulación de documentos que sirvan para apoyar las necesidades de información de diversos tipos de comunidades (Rodríguez Gallardo, p.482) y cuya función puede ser reconocida como una tarea colectiva de construcción de conocimiento (Fernández de Zamora p.222).

(Gorman p. 2) enfatiza que datos, información y conocimiento son conceptos presentes en las bibliotecas, ya que éstas se basan en la idea que impartir, crear y conservar el conocimiento son acciones vitales a una sociedad civilizada.

Las redes de bibliotecas hacen posible compartir conocimientos de manera inmediata y desarrollar proyectos de investigación (Ortega Amieva p. 81).

Evalúe la pertinencia, actualidad y relevancia de la información localizada para incluirla en una página web que agrupa datos, información y conocimiento sobre las artes plásticas y la arquitectura mexicanas en sus épocas prehispánica, colonial, moderna y contemporánea.



Internet ha crecido demasiado y no todo lo que contiene merece la pena (Naumis p.460). Otros autores coinciden en que es una selva informativa, un laberinto, información coloquial con datos no siempre válidos (Lau p.63). No obstante el ciclo casi completo de la generación del conocimiento, puede realizarse a través de los servicios cada vez más integrados que ofrece el Internet.

Por ejemplo, la búsqueda de información, el proceso de investigación, la edición, la publicación, la difusión e incluso la venta o distribución del producto de investigación de un artículo, ponencia o libro, pueden hacerse a través del Internet (Lau p.61).

Existen una explosión de la información y de conocimiento además de un vertiginoso desarrollo de las tecnologías de la información y comunicación (Ortega Amieva p.79-80).

Esta autora se pregunta ¿Estamos mejor informados? ¿Estamos educados para seleccionarla, para analizarla, para producir nueva información y con ella nuevos conocimientos?

La información es un insumo básico del aprendizaje, especialmente en los modelos educativos centrados en el estudiante como son los de tipo constructivista o cognoscitivo. En estos esquemas los procesos se basan en la construcción del conocimiento por parte del alumno (Lau p.78).

A lo anterior (Lau p.53) comenta que los alumnos tendrán mayor éxito en construir el conocimiento que requieren, según sea la calidad y la cantidad de información que tengan a su disposición.

Transmitir la información, en el caso del Instituto Cultural Itaú-ICI, de los pintores de Brasil de los siglos XVIII, XIX y XX, a un número cada vez mayor de personas de forma rápida y eficiente ha sido un desafío presente en la historia de la Sociedad humana (Castillo Costa p. 1).

Quise tomar este desafío, y con el apoyo de CONACYT en octubre del 2000 elaboré la página web Mex@rte (<http://www.mexicoarte.net>) y la base de datos MEXICOARTE (7000 registros) que difunden la información de las artes plásticas y la arquitectura mexicanas con el objetivo de llegar a conocerlas. Es decir dirigí esfuerzos, como lo recomienda (Hernández Salazar, p. 462) a la creación de infraestructuras para que la sociedad científica tenga las herramientas necesarias para poder producir conocimiento.

## BIBLIOGRAFÍA

Barberena Blásquez, Elsa y Gorbea-Portal, Salvador. "La tecnología de la información A favor del usuario de arte mexicano". En: *Memoria de las XXXII Jornadas Mexicanas de Biblioteconomía*. México: Asociación Mexicana de Bibliotecarios, A.C. 2001. p. 36.

Castillo Costa, María Cristina. "Informatica e cultura - uma experiencia concreta já em operacao". *Revista arte e cultura*. Sao Paulo, 2(2) julio, 1991.

*Coloquio Internacional de Investigación Bibliotecológica y de Información. Los grandes problemas*. México: UNAM. Centro Universitario de Investigaciones Bibliotecológicas, 2002. (Colección Sistemas Bibliotecarios de Información y Sociedad).

Lau, Jesús, p. 53,61,63,73, 78-79.

Ortega Amieva, Diana Cecilia, p. 79-81.

*Contribución al desarrollo de la Sociedad del Conocimiento*. México: UNAM. Centro Universitario de Investigaciones Bibliotecológicas, 2000.

Fernández de Zamora, Rosa María, p. 222-223.

Hernández Salazar, Patricia, p. 462.

Lafuente López, Ramiro, p. 478.

Naumis Peña, Catalina, p. 459-461.

Ramírez Leyva, Elsa Margarita, p. 476.

Rendón Rojas, Miguel Ángel, p. 9.

Rodríguez Gallardo, Adolfo, p. 482.

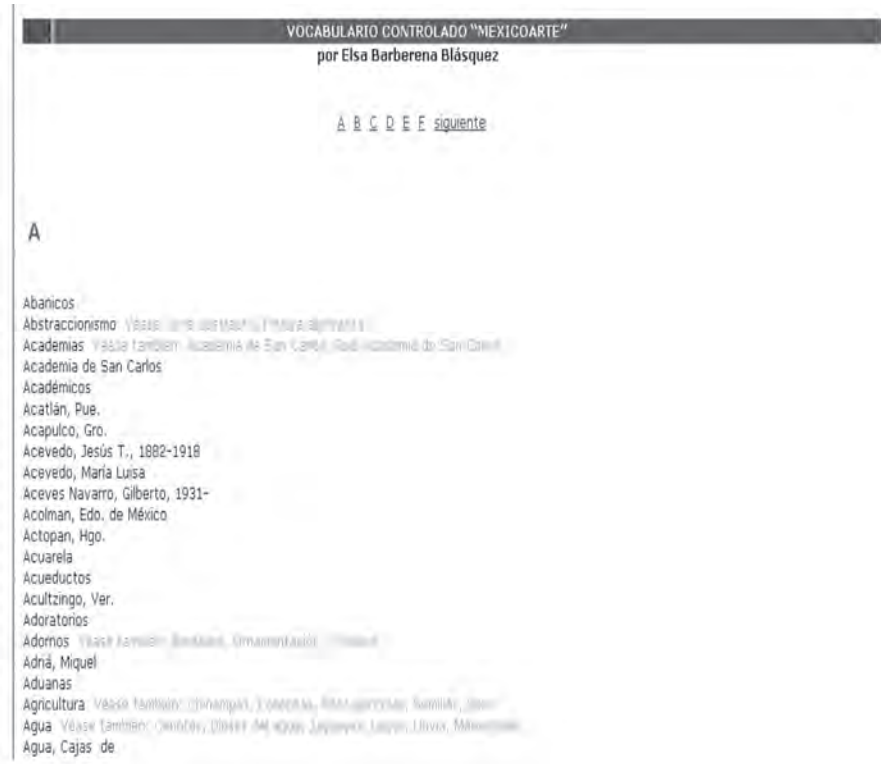
Davenport, T., Prusak, L. (1999).

[http://www.gestiondelconocimiento.com/conceptos\\_diferenciaentredato.htm](http://www.gestiondelconocimiento.com/conceptos_diferenciaentredato.htm) Consultado el 19 de marzo de 2007.

Gorman, Michael. *Our enduring values. Librarianship in the 21st century*. Chicago: American Library Association, 2000. p. 2.

## ANEXOS

### Ejemplos de vocabularios controlados



HISTORIA    FUNCIONAMIENTO    CONDICIONES PARA INTEGRAR LA RED    PRODUCTOS    DIRECTORIO    INVITACION    FORMULARIO  
DESCUOTROS



**VOCABULARIO CONTROLADO DE ARQUITECTURA, ARTE, DISEÑO Y URBANISMO**

**Historia**

En 1994 cuando se forma Vitruvio se detecta la inexistencia de un vocabulario especializado en lengua española en las áreas cubiertas por las bibliotecas de la red que permitiera el control de la terminología usada en el país, y su unificación en un lenguaje. Para superar este vacío terminológico argentino, se forma ese mismo año, la Subcomisión de Lenguaje cuyo objetivo será el desarrollo de una herramienta para las tareas de Indización y de acceso y recuperación de la Información.

En 1995, la Subcomisión de Lenguaje comienza a desarrollar el *Vocabulario controlado* tomando como documento base el *Listado alfabético de términos usados en las bibliotecas cooperantes de la Red Vitruvio* (1995) que compiló alfabéticamente 4400 términos, utilizados en una primera etapa, únicamente por las bibliotecas integrantes de la Subcomisión de Lenguaje en las tareas de clasificación e Indización. En 1996, otras once bibliotecas cooperaron aportando su terminología: Biblioteca Central de la Universidad Nacional de Mar del Plata; Biblioteca de la Asociación Argentina de Arquitectura e Ingeniería Hospitalaria; Biblioteca de la Facultad de Arquitectura y Urbanismo de la Universidad Nacional de La Plata; Biblioteca de la Facultad de Arquitectura y Urbanismo de la Universidad Nacional de Tucumán; Biblioteca de la Facultad de Arquitectura, Urbanismo y Diseño de la Universidad Nacional de Córdoba; Biblioteca de la Facultad de Filosofía y Humanidades de la Universidad Nacional de Córdoba; Biblioteca del Colegio de Arquitectos de la Provincia de Buenos Aires / Distrito 2 (Banfield) y la Biblioteca de la Sociedad Argentina de Autores y Compositores de Música.

También hubo intercambio de información con: Biblioteca de la Facultad de Arquitectura - Universidad de la República Oriental del Uruguay y la Biblioteca de la Facultad de Arquitectura, Diseño y Arte - Universidad Nacional de Ingeniería de Lima (Perú)

El *Listado alfabético de términos usados* -solo accesible en formato impreso- alcanzó una tercera edición con sucesivas actualizaciones hasta el 2003. Con la edición del *Vocabulario controlado* (2004), el *Listado alfabético de términos usados* abandonó su función original para convertirse en una lista de términos propuestos en constante desarrollo. La coexistencia de ambas herramientas es necesaria, pues la terminología en uso del *Listado alfabético* no ha sido incorporada en su totalidad, al *Vocabulario controlado*.

**Alcance**

El *Vocabulario Controlado* es especializado, multidisciplinario y monolingüe. Fue creado para utilizarse en las bases de datos de las bibliotecas integrantes de la Red Vitruvio, que abarcan las áreas de Arquitectura, Arte, Diseño en sus diferentes especialidades y Urbanismo.

La presente edición cuenta con 1943 descriptores de los cuales 431 incluyen notas de alcance, y 279 términos no usados que remiten a sus sinónimos. La Subcomisión de Lenguaje continúa trabajando en la ampliación y actualización de este producto para mejorarlo y enriquecerlo.

**iiisue** UNAM Instituto de Investigaciones sobre la Universidad y la Educación

**iresie** Banco de Datos sobre Educación Iberoamericana

Responsable del íesie  
Ma Angela Torres Verdugo  
resie@coors.unam.mx

**Consulta en línea**

- Información del íesie
- Revista Perfiles Educativos
- Listado de revistas
- Bibliotecas cooperantes
- Obtención de artículos
- Ingreso de revistas a íesie
- Vocabulario controlado
- Créditos
- Tutorial
- Otras bases: IDOCEO, CATMEN, REDMEX

### PRESENTACION

El **VOCABULARIO CONTROLADO (VC)** consiste en una lista normalizada de temas e identificadores. Está integrado por alrededor de 1,500 términos sobre diferentes aspectos del área educativa, que han compilados a lo largo de los veintisiete años de desarrollo del Índice de Revistas de Educación Superior e Investigación Educativa (**IRESE**). Constituye un instrumento de trabajo para la clasificación e indicación de los artículos y documentos que ingresan a las bases de datos **IRESE** y al Índice de Documentos de Educación, (**IDOCED**) que se generan en el Instituto de Investigaciones sobre la Universidad y la Educación (**IISUE**) antes Centro de Estudios sobre la Universidad (**CESU**).

Un gran porcentaje de estos vocablos fueron tomados del Tesouro de la Educación de la UNESCO, OEI en sus diferentes ediciones y del Tesouro del ERIC (Educational Resources Information Center). En ocasiones, algunos vocablos se adaptaron a la terminología utilizada en México. También hay términos que han sido introducidos por los documentalistas del **IRESE** en concordancia con las nuevas subdisciplinas y la evolución de los conceptos en el campo educativo.

Con el fin de facilitar el uso de este vocabulario, tanto al documentalista que analiza el contenido de los documentos y clasifica la información, así como al usuario que consulta las bases de datos mencionadas, a través de su página Web [www.unam.mx/cesu/iresie](http://www.unam.mx/cesu/iresie). A continuación se presentan algunos conceptos básicos y

**iiisue** UNAM Instituto de Investigaciones sobre la Universidad y la Educación

**iresie** Banco de Datos sobre Educación Iberoamericana

Responsable del íesie  
Ma Angela Torres Verdugo  
iresie@coors.unam.mx

**Consulta en línea**

- Información del íesie
- Revista Perfiles Educativos
- Listado de revistas
- Bibliotecas cooperantes
- Obtención de artículos
- Ingreso de revistas a íesie
- Vocabulario controlado
- Créditos
- Tutorial
- Otras bases: IDOCEO, CATMEN, REDMEX

Consulta: EDUCACION\*AND\*ARTE Fichas encontradas: 4

Anterior Fichas: 14

4 / 4

**MUZZACL ANTONELLA**  
"GENERAL EDUCATION AND MUSEUM EDUCATION: BETWEEN SINGULARITY AND PLURALITY"  
REVISTA COMPLUTENSE DE EDUCACION (ESPAÑA), Vol. 17, No. 1, Año 2006, Págs. 65-75.

Temas: MUSEOS, DIDACTICA, ENSEÑANZA, APRENDIZAJE, EDUCACION GENERAL.

Resumen: EL ARTICULO EXAMINA LA RELACION ENTRE EDUCACION GENERAL Y EDUCACION A TRAVES DE LOS MUSEOS, Y ENTRE VARIAS MATERIAS Y LA EDUCACION EN LOS MUSEOS. EL TERMINO "EDUCACION" ES NORMAL EN EL VOCABULARIO GENERAL Y EN SU SIGNIFICADO MAS AMPLIO IMPLICA LA ACTIVIDAD DE ENSEÑAR, PERO DENTRO DE SU SIGNIFICADO ESPECIFICO SE REFIERE AL CONOCIMIENTO PROPIO DE ESTA ACTIVIDAD. ESTO IDENTIFICA UN AREA DE INVESTIGACION Y CONOCIMIENTO DENTRO DE LAS CIENCIAS DE LA EDUCACION DEFINIDO COMO EDUCACION GENERAL EN REALIDAD PUEDE TOMAR MUCHAS FORMAS (POR EJEMPLO EDUCACION ESPECIAL PARA

The Library of Congress >> Go to Library of Congress Authorities

---

**LIBRARY OF CONGRESS ONLINE CATALOG**

[Help](#) | [New Search](#) | [Search History](#) | [Headings List](#) | [Titles List](#) | [Request an Item](#) | [Account Status](#) | [Other Databases](#) | [Start Over](#)

DATABASE: Library of Congress Online Catalog  
 YOU SEARCHED: Subject Browse = Painters and Mexico  
 SEARCH RESULTS: Displaying 1 through 25 of 25

◀ Previous   Next ▶

#	Hits	Headings (Select to View Titles)	Type of Heading
[ 1 ]	23	<a href="#">Painters Mexico</a>	LC subject headings
[ 2 ]	2	<a href="#">Painters Mexico Autographs</a>	LC subject headings
[ 3 ]	85	<a href="#">Painters Mexico Biography</a>	LC subject headings
[ 4 ]	1	<a href="#">Painters Mexico Biography Dictionaries</a>	LC subject headings
[ 5 ]	37	<a href="#">Painters Mexico Biography Juvenile literature</a>	LC subject headings
[ 6 ]	1	<a href="#">Painters Mexico Catalogs</a>	LC subject headings
[ 7 ]	2	<a href="#">Painters Mexico Colima Biography</a>	LC subject headings
[ 8 ]	5	<a href="#">Painters Mexico Correspondence</a>	LC subject headings
[ 9 ]	2	<a href="#">Painters Mexico Dances</a>	LC subject headings
[ 10 ]	1	<a href="#">Painters Mexico Etchings</a>	LC subject headings
[ 11 ]	7	<a href="#">Painters Mexico Interviews</a>	LC subject headings
[ 12 ]	2	<a href="#">Painters Mexico Jalisco</a>	LC subject headings
[ 13 ]	2	<a href="#">Painters Mexico Jalisco Biography</a>	LC subject headings
[ 14 ]	1	<a href="#">Painters Mexico Jalisco Interviews</a>	LC subject headings
[ 15 ]	1	<a href="#">Painters Mexico Maratón (Simola) Biography</a>	LC subject headings
[ 16 ]	1	<a href="#">Painters Mexico Oaxaca Biography</a>	LC subject headings
[ 17 ]	2	<a href="#">Painters Mexico Periodicals</a>	LC subject headings
[ 18 ]	1	<a href="#">Painters Mexico Portraits Catalogs</a>	LC subject headings
[ 19 ]	1	<a href="#">Painters Mexico Psychology</a>	LC subject headings
[ 20 ]	1	<a href="#">Painters Michigan Biography</a>	LC subject headings
[ 21 ]	1	<a href="#">Painters Middle West [from old catalog]</a>	LC subject headings
[ 22 ]	3	<a href="#">Painters Minnesota Biography</a>	LC subject headings
[ 23 ]	2	<a href="#">Painters Miscellanea</a>	LC subject headings
[ 24 ]	3	<a href="#">Painters Mississippi Biography</a>	LC subject headings
[ 25 ]	2	<a href="#">Painters Mississippi Horn Island Biography</a>	LC subject headings

◀ Previous   Next ▶

[Help](#) - [Search](#) - [History](#) - [Headings](#) - [Titles](#) - [Request](#) - [Account](#) - [Databases](#) - [Exit](#)

---




**The Library of Congress**  
 URL: <http://www.loc.gov/>  
 Mailing Address:  
 101 Independence Ave, S E  
 Washington, DC 20540

**Catalog/authority record errors?**  
[Use our Print Report Form](#)  
**Questions about searching?**  
[Ask a Librarian](#)


**Library of Congress Online Catalog**  
 URL: <http://catalog.loc.gov/>  
**Library of Congress Authorities**  
 URL: <http://authorities.loc.gov/>



The screenshot shows the website for The Getty's vocabularies. At the top is the navigation menu with links for Home, Visit, Museum, Research Institute, Conservation Institute, Foundation, and The Trust. Below this is a breadcrumb trail: Research Institute Home > Conducting Research > Learn about the Getty Vocabularies. The main heading is 'Learn About the Getty Vocabularies'. On the left, there is a sidebar with links: Exhibitions, Onlusiting Research, Scholarly Activities, and About the Research Institute. The main content area starts with a paragraph explaining that the Getty vocabulary databases are produced and maintained by the Getty Vocabulary Program, compliant with ISO and NISO standards. It lists three ways the vocabularies can be used: at the data entry stage, by catalogers/indexers, as knowledge bases, and as search assistants. Below this are three sections, each with an image and a description: 1. 'The Art & Architecture Thesaurus (AAT)' with an image of a classical building facade. 2. 'The Union List of Artist Names (ULAN)' with a portrait of a man. 3. 'The Getty Thesaurus of Geographic Names (TGN)' with a starburst map icon. To the right of these sections is a 'Related Sections' box with links to 'Obtain the Getty Vocabularies', 'Contribute to the Getty Vocabularies', 'Frequently Asked Questions', 'Getty Vocabularies Download Center', and 'Editorial Guidelines'. Below this box is an 'Email Vocabulary Program' link and a small image of a map with the caption 'Vicente Maria Coronelli map of Nueva España (detail), late 17th century, hand-colored engraving, PR40001'. At the bottom of the page, there is a footer with 'Useful Links' and a search bar.

 **Research**

Research Home > Conducting Research > Art & Architecture Thesaurus > Search Results

 **Art & Architecture Thesaurus® Online Search Results**

[New Search](#)   [Previous Page](#)   [Help](#)

---


Find Name: **furniture**





Logic:


Note: 58 results

[View Selected Records](#)   [Select All Records](#)   [Clear All](#)   [First](#)   [Previous](#)   [Next](#)   [Last](#)


Page: **1** 2 3

Click the  icon to view the hierarchy.  
Check boxes to view multiple records at once.

-  architectural furniture  
(<furniture by design>, furniture, ... Furnishings and Equipment) [300040029]  
**furniture, architectural**
-  art furniture  
(<furniture by design>, furniture, ... Furnishings and Equipment) [300164268]  
**furniture, art**
-  <bases: furniture components>  
(<furniture components>, <furnishings components>, ... Components) [300190874]
-  bed furniture  
(<coverings and hangings for sleeping and reclining furniture>, <coverings and hangings for furniture>, ... Furnishings and Equipment) [300216511]  
**furniture, bed**

 **Research**

Research Home > Conducting Research > Union List of Artist Names > Search Results

 **Union List of Artist Names® Online Search Results**

[New Search](#)   [Previous Page](#)   [Help](#)

---


Find Name:





Role:

Nationality: **Mexican** 362 results

[View Selected Records](#)   [Select All Records](#)   [Clear All](#)   [First](#)   [Previous](#)   [Next](#)   [Last](#)

Page: **1** 2 3 4 5 6 7 8 9 10

Click the  icon to view the hierarchy.  
Check boxes to view multiple records at once.

-  Aceves Navarro, Gilberto  
(Mexican painter, born 1931) [500118530]  
**Gilberto Aceves Navarro**
-  Aguilar, Manuel  
(Mexican artist, contemporary) [500116024]  
**Manuel Aguilar**
-  Aguirre, Carlos  
(Mexican installation artist, born 1948) [500116030]  
**Carlos Aguirre**
-  Aguirre, Ignacio  
(Mexican artist, born 1902) [500092966]  
**Ignacio Aguirre**



Research

Research Home • Conducting Research • Thesaurus of Geographic Names • Full Record Display

 **Getty Thesaurus of Geographic Names® Online**  
Full Record Display

[New Search](#)   [Previous Page](#)   [Help](#)

[Verbanular Display](#) | [English Display](#)

Click the  icon to view the hierarchy.

**ID:** 1062138      **Record Type:** administrative


 **La Puebla (inhabited place)**

**Coordinates:**  
 Lat: 39 46 00 N *degrees minutes*    Lat: 39.7667 *decimal degrees*  
 Long: 003 01 00 E *degrees minutes*    Long: 3.0167 *decimal degrees*

**Names:**  
**La Puebla** (preferred, C,N,N)  
**So Poble** (C,N,N)

**Hierarchical Position:**

-  World (face)
-  Europe (continent)
-  Spain (nation)
-  .. Illas Balears (autonomous community)
-  .. Illas Balears (province)
-  .. Mallorca (island)
-  .. La Puebla (inhabited place)

 Results: Browse      [Exit Program](#)

**Important Message**

[Advanced Search](#)  
[Basic Search](#)  
[Browse](#)  
[Search History](#)  
[Thesaurus](#)  
[Print Email Save](#)  
[Exporting/Clone](#)  
[Create Alerts](#)

[Help](#)  
[Journal](#)  
[Directory](#)

**Browse For:** art, Mexican  
**In:** Subject(s)  
**Found In:** arttt » Art Full Text

- Art, Mesopotamian - arttt (5)
- ART, Mesopotamian / Collections - arttt (1)
- Art, Mesopotamian / Exhibitions - arttt (4)
- Art, Mesopotamian / Study and teaching - arttt (1)
- Art, Mexican - arttt (6)
- Art, Mexican / 17th century - arttt (1)
- Art, Mexican / 18th century - arttt (1)
- Art, Mexican / 18th century - arttt (1)
- Art, Mexican / 20th century - arttt (26)
- Art, Mexican / 20th century / Exhibitions - arttt (54)
- Art, Mexican / 21st century - arttt (2)
- Art, Mexican / 21st century / Exhibitions - arttt (12)
- Art, Mexican / Collections - arttt (9)
- Art, Mexican / Exhibitions - arttt (52)
- Art, Mexican / Influences - arttt (2)
- Art, Mexican American - arttt (23)
- Art, Mexican American / Collections - arttt (2)
- Art, Mexican American / Collections / Exhibitions - arttt (1)
- Art, Mexican American / Exhibitions - arttt (33)
- Art, Micronesia - arttt (1)

[Revise Search](#)   [Selected Entries](#)   [Clear Selected](#)



## SECCIÓN VI

### La informática en la recuperación de información



# Procesamiento del lenguaje natural: estado de la investigación

ALEXANDER GELBUKH

*Centro de Investigación en Computación,  
Instituto Politécnico Nacional, México*

## RESUMEN

**E**l procesamiento del lenguaje natural es una rama de la ciencia que pertenece a la intersección de la lingüística aplicada y las ciencias de la computación, que estudia los métodos necesarios para que la computadora pueda ejecutar varias tareas relacionadas con el lenguaje humano, como el español, y requiere cierto grado de «entendimiento» de su contenido. Por otro lado esta ciencia desarrolla las herramientas que ayudan al lingüista en su trabajo cotidiano e incluso pueden llevar a descubrimientos lingüísticos nuevos. Este artículo presenta las aplicaciones principales del procesamiento automático de lenguaje natural y discuten los métodos que se emplean para la resolución de problemas. Al final del artículo se da una reseña bibliográfica para el lector interesado en obtener mayor información sobre el tema.

## INTRODUCCIÓN

En los últimos cinco-diez miles de años, la ocupación más importante de la raza humana ha estado enfocada a producir, mejorar y pasar-

les a las siguientes generaciones el tesoro más valioso que tenemos: el conocimiento. El conocimiento se almacena en la forma de lenguaje humano: libros, periódicos y en nuestros días la Internet. Como las computadoras son mucho más capaces que los seres humanos para manejar esta cantidad enorme de conocimiento, el estudio de las técnicas del procesamiento automático se ha vuelto una necesidad prioritaria y urgente.

La tarea final de la ciencia de la lingüística computacional —como la piedra filosofal para la alquímica— es la construcción de una máquina que pueda leer, escribir y conversar en nuestro lenguaje, ya sea español, inglés, chino, etcétera. Es un gran reto porque se trata de modelar en una computadora la actividad mental del ser humano. Aunque para bien o para mal cumplir con esta tarea no se ve como algo factible en el futuro cercano, al ir transcurriendo el camino hacia su tarea final, esta ciencia ya ha producido desarrollos muy útiles (al igual que hizo en su tiempo la alquimia) en tareas limitadas pero de gran importancia práctica.

A pesar de su nombre un poco confuso, la lingüística computacional no es una variante de la lingüística general; mientras ésta última estudia *qué* es el lenguaje, la primera estudia el *cómo* modelarlo computacionalmente con un fin práctico.<sup>1</sup> Se puede comparar esta relación con la que hay entre un ornitólogo y un constructor de aviones: el primero estudia los colores y las costumbres de las aves, el segundo tiene como tarea el *hacer* un ave —y no necesariamente que sea parecida en su estructura a las aves naturales, sino que vuele—.

Sin embargo existe una relación profunda entre las dos ciencias. La lingüística computacional, especialmente en la etapa «simbólica» de su desarrollo (véase más abajo), ha contribuido a la lingüística teórica<sup>2</sup> con los formalismos que ayudan a los lingüistas a expresar sus ideas con exactitud matemática y a comprobarlas experimentalmente sobre grandes corpus de textos. Con las herramientas computacionales los lingüistas pueden evitar el trabajo repetitivo y aburrido, y dejar que haga esto la máquina. Los fenómenos descubiertos durante el

---

1 En este sentido se puede clasificar como una rama de la lingüística aplicada.

2 O ha inspirado su creación.

desarrollo de los sistemas computacionales aplicados han sido una inspiración para los descubrimientos lingüísticos. Por otro lado, la lingüística teórica ha ayudado profundamente, con sus métodos y sus conceptos a la lingüística computacional; las dos ciencias se complementan sin sustituirse.

Desde el punto de vista técnico la tarea del procesamiento automático del lenguaje es (viéndolo por parte del análisis) la definición de la estructura en el texto dado. El mayor problema en este camino es la ambigüedad: no es tan difícil encontrar una estructura en un segmento del texto en la que se tenga que elegir entre más de una estructura posible, al existir diferentes interpretaciones de ese segmento. Por ejemplo, la palabra «sobre» puede ser un verbo, un sustantivo o una preposición («todo lo que sobre se dejará en un sobre sobre la mesa»); la frase «veo al gato con un telescopio» se puede interpretar como «veo al gato, y para eso uso un telescopio» o «veo al gato, el cual tiene un telescopio». Tales ambigüedades se presentan en grandes cantidades en casi cualquier oración o palabra, aunque el lector humano raras veces las nota. Para resolver tales ambigüedades es necesario o bien analizar un contexto más amplio del segmento de texto dado, o bien emplear el conocimiento externo al texto dado: por ejemplo, la ambigüedad con la palabra «sobre» se puede en muchos casos resolver analizando su contexto: «El documento fue entregado en un sobre cerrado».

En este artículo presentamos brevemente las aplicaciones principales del procesamiento automático de texto, discutimos las tareas específicas que técnicamente aparecen en el proceso de análisis, describimos los tipos de métodos principales que se ocupan para resolver los problemas y adjuntamos una bibliografía para el lector interesado en obtener mayor información sobre el tema.

## APLICACIONES

Mencionaremos aquí sólo tres tipos de las aplicaciones principales del procesamiento del lenguaje natural.

**Búsqueda de información.** La forma más tradicional de búsqueda de información es la *recuperación de información*. Se trata del es-

cenario en el cual el usuario especifica los criterios que deben satisfacer los documentos de interés —usualmente unas palabras que éstos deben contener— y el sistema debe proporcionarle estos documentos. El nombre de la aplicación, «recuperación de información», refleja su forma más básica e históricamente primera: el proporcionar, o «recuperar», el conjunto de los documentos que satisfacen el criterio. Sin embargo las técnicas modernas de recuperación de información no implican selección sino ordenamiento de toda la colección de documentos disponible siguiendo el mayor o menor grado de satisfacción del criterio, así que los documentos que mejor satisfacen el criterio aparecen al inicio de la lista.

Dos consideraciones son importantes aquí. Primero, en este escenario el criterio no simplemente se satisface o no, sino que puede satisfacerse en mayor o menor grado, por ejemplo, el documento puede contener mayor o menor número de palabras relacionadas con la petición del usuario, aquí las técnicas del procesamiento del lenguaje natural ayudarían a decidir qué palabras y en qué contextos deberían relacionarse con la petición. Segundo, y más importante, la tarea se entiende no sólo como la comparación entre la petición del usuario con los textos de los documentos, sino como satisfacción de la necesidad informática del usuario. Esto involucra el razonamiento lógico o estadístico sobre la necesidad del usuario y la mejor manera de satisfacer ésta.

Al satisfacer la necesidad del usuario las técnicas modernas van más allá de simplemente proporcionarle una lista ordenada de los documentos. Por el tipo de la necesidad y el uso de la información que se quiere obtener, se distinguen las siguientes corrientes modernas en la búsqueda de la información.

En la *respuesta a preguntas*, la petición del usuario se entiende como una pregunta («¿dónde se organizan los juegos Olímpicos en 2008?») y la salida del sistema se da no en la forma de una lista ordenada de documentos sino en la forma de una simple respuesta a tal pregunta («en China»). Esto le ahorra tiempo al usuario ya que en lugar de tener que analizar los documentos, él recibe la respuesta directa. En este tipo de técnicas el procesamiento del lenguaje natural es necesario para «entender» tanto la pregunta del usuario como la información



presentada en los documentos de la colección en la cual se busca la respuesta. La corriente que activamente se desarrolla en la actualidad es la respuesta a preguntas multilingües: el lenguaje de la pregunta (por ejemplo, español) puede ser diferente del lenguaje de los documentos analizados (por ejemplo, inglés), o incluso éstos pueden estar en diferentes lenguajes; la situación es importante para la Unión Europea y varios otros países.

En la *generación de resúmenes*, el usuario está interesado en obtener una idea general sobre un documento antes de leerlo, o para decidir si lo necesita leer. El sistema entonces le proporciona una representación resumida del documento. Esto es especialmente importante cuando se aplica a una colección de muchos documentos, posiblemente millones, los cuales el usuario simplemente no podría leer. Aún más difícil es la tarea cuando se trata de documentos escritos en diferentes lenguajes, o cuando el lenguaje del resumen requerido no coincide con el lenguaje de los documentos.

En la *minería de texto*, también se busca una especie de representación resumida de una gran cantidad de textos, pero en este caso tal «resumen» no refleja lo que los autores querían expresar en sus textos sino la metainformación sobre estos textos: la información de las tendencias, opiniones prevalecientes, porcentajes de diferentes opiniones, etcétera. Gracias a muchas aplicaciones prácticas en la política, economía, comercio, sociología y otras áreas de la vida social, esta rama del procesamiento del lenguaje natural ha recibido recientemente una gran atención por parte de los investigadores y las empresas.

**Traducción automática.** Con la globalización, la abundancia de la información textual en la Internet, la intensificación de los contactos internacionales y la disponibilidad de Internet por cada vez un mayor número de personas en los países menos desarrollados convirtieron a la traducción automática en una aplicación prioritaria. Debido a que la traducción requiere un procesamiento detallado y profundo del significado del texto, ésta fue una de las motivaciones principales que desarrollaron la ciencia de la lingüística computacional desde la época de las primeras computadoras electrónicas.

El avance de la traducción automática en los últimos años ha sido impresionante: los traductores automáticos pasaron de una herra-

mienta auxiliar o más bien experimental a los sistemas útiles para el uso masivo cotidiano de millones de usuarios no expertos. Uno de los mejores traductores automáticos disponibles en la actualidad es el de *Google*©.

**Interfaces en lenguaje natural.** Esto será probablemente la tarea del futuro, y hasta ahora ha logrado un avance limitado. La idea es que en el futuro cuando las computadoras empiecen a actuar en el mundo real (robots inteligentes) y ayudar en sus quehaceres cotidianos a la gente no experta, será indispensable que puedan hablar con los usuarios en lenguaje natural. Sin embargo tal cosa requiere un alto grado de entendimiento del mensaje, y además involucra áreas de la lingüística computacional que todavía están en pañales: el manejo del diálogo y de ciencia ligüística de la pragmática, entre otras.

Es por eso que el manejo del diálogo atrae la atención de muchos investigadores y es un área de investigación activa. No podríamos decir lo mismo de la pragmática computacional, todavía se conoce muy poco sobre el tratamiento computacional de la pragmática.

## TAREAS

Para lograr el objetivo requerido, los sistemas de procesamiento automático de texto dividen internamente el trabajo de una u otra manera en bloques cuya información de entrada y de salida está bien definida, de tal manera que el sistema completo se puede armar de estos bloques intercambiables. En la mayoría de los casos se trata de la ejecución consecutiva de los bloques, donde la salida del bloque anterior sirve como entrada al siguiente. En muchos casos, cada bloque toma en la entrada una secuencia de elementos más simples y agrupa algunos de ellos produciendo una secuencia de elementos compuestos más «grandes».

Tales bloques permiten el estudio independiente y la especialización de los grupos de investigación, los investigadores, los libros y los recursos computacionales y los léxicos. En muchas tareas se han establecido incluso prestigiosas competencias internacionales donde se definen los mejores sistemas para la tarea en cuestión. A las correspondientes ramas de la lingüística computacional se les refiere como

tareas (*tasks* en inglés). No todas las tareas se emplean en una aplicación dada. Usualmente se emplean las tareas hasta un cierto nivel necesario. Podemos mencionar los siguientes ejemplos de tareas:

**Preprocesamiento.** Es un nombre paraguas para varias tareas que se consideran relativamente simples —más simples que las tareas que le siguen— y que se ejecutan antes de las tareas más complicadas. Diferentes autores pueden considerar algunas de las sub-tareas aquí mencionadas como tareas independientes.

*Limpieza* del texto es la localización y eliminación o el tratamiento adecuado de los segmentos del archivo de texto que no son propiamente texto: imágenes, marcaje del formato, fórmulas, letras mayúsculas y minúsculas, etcétera; es decir, todo aquello que ayuda a convertir el archivo dado en un texto bien formado en el sentido lingüístico.

*Determinación de elementos textuales* (*tokenizing* en inglés) implica convertir el texto de una secuencia de letras a una secuencia de cadenas, aproximadamente correspondientes a palabras. Es importante recalcar que esta tarea en apariencia tan simple está en realidad interconectada con decisiones que se toman en otros niveles. Por ejemplo, en la secuencia de caracteres «etc.», ¿es el punto parte de la palabra o es un elemento textual aparte que indica la finalización de la oración? El programa que efectúa esta tarea se llama en inglés *tokenizer*.

*Reconocimiento de entidades nombradas* (NER, *named entity recognition* en inglés) es la etapa de aglutinar varias palabras en una sola que es el nombre de un concepto u objeto; por ejemplo. «Estados Unidos Mexicanos» se tratará como una sola palabra que refiere al país correspondiente y es sinónima de «México». Se aplican las mismas aclaraciones sobre la no trivialidad de esta tarea. El descubrimiento de tales entidades nombradas y la toma de decisiones sobre la aglutinación, o no, de los elementos textuales en un contexto dado en una entidad nombrada, han sido áreas de activa investigación en los últimos años.

*Separación de las oraciones* (*sentence splitting* en inglés) es la toma de decisión acerca de dónde están las fronteras entre las oraciones. El ejemplo arriba mencionado de la secuencia «etc.» ilustra un ca-

so no trivial de esta tarea. Un programa que efectúe esta tarea se llama en inglés *sentence splitter*.

*Eliminación de las palabras basura (stopwords en inglés)*. Para ciertas aplicaciones, las palabras sin contenido semántico, tales como las palabras funcionales o muy frecuentes, no aportan casi nada al desempeño del sistema y pueden perjudicarlo. La lista de las palabras que deben eliminarse puede variar de una aplicación a otra y en ocasiones está sujeta a decisiones inteligentes automáticas.

**Análisis morfológico.** Esta tarea de índole propiamente lingüística involucra la determinación de la primera forma de la palabra y sus características gramaticales y morfosintácticas. Se puede efectuar sin contexto, es decir, sobre una palabra aislada, y en este caso proporcionará un resultado ambiguo, por ejemplo: «como», «nada», «habla», «hablamos». También se puede efectuar en el contexto, lo que usualmente permite desambiguar la palabra para dar una sola respuesta: «considerarlo como tal», «Juan nada bien», «él habla poco», «ayer hablamos de eso». El programa que efectúa la tarea completa sin contexto —dando usualmente varias variantes de la respuesta— se llama un analizador morfológico, y si lo hace en contexto —produciendo sólo una variante de la respuesta— se llama etiquetador (*tagger*). El programa que sólo determina la primera forma de la palabra dada («comió» «comer»), se llama en inglés *lemmatizer*. Para ciertas aplicaciones, sobre todo en la recuperación de información, es suficiente determinar sólo la base de la palabra sin generar la forma completa («comió» «com-»).

**Análisis sintáctico.** Se trata de agrupar las palabras dentro de una oración, según su relación sintáctica y descubrir así su estructura interna. Existe un número considerable de teorías y metodologías para efectuar esta tarea, las cuales difieren substancialmente en la definición de ella y la definición de la relación sintáctica. Las corrientes más comunes y más distintas entre sí son la corriente de análisis sintáctico de constituyentes, y la de dependencias.

En la primera las palabras se agrupan y se forma un árbol sintáctico cuyos vértices son grupos de palabras relacionadas, y las aristas son las relaciones de anidamiento entre estos conjuntos. En la segunda aproximación las palabras se asignan una como «ayudante»

de la otra, y esta relación forma un árbol cuyos vértices son palabras individuales. Por lo general el análisis de constituyentes tiene un mejor sustento matemático y es más fácil de efectuar, mientras que el análisis de dependencias proporciona información más rica y mejor organizada para el subsecuente análisis semántico. Los programas que efectúan el análisis sintáctico se llaman en inglés *parsers*, aunque este término más comúnmente se aplica al análisis de constituyentes.

**Análisis semántico.** Esta tarea concierne al tratamiento del significado del texto más que a su estructura, y puede involucrar varios aspectos del problema.

La *desambiguación de los sentidos de las palabras* (WSD: *word sense disambiguation* en inglés) es la tarea de elegir una sola acepción de la palabra, del repertorio dado por un diccionario, dependiendo del contexto dado. Por ejemplo las diferentes acepciones que se le deben asignar a la palabra banco en los contextos «Juan tiene cuenta en el banco», «la arena del banco del río Nilo» y «el banco de peces». La tarea es muy importante hasta en las aplicaciones básicas que no requieren del análisis semántico completo, tales como la recuperación de información, y recientemente recibe mucha atención. Se organiza una competencia internacional sobre la tarea, SemEval.<sup>3</sup> En el sitio de SemEval se pueden obtener las colecciones y corpus necesarios para la investigación en esta tarea.

La *resolución de anáfora* es un caso extremo de la desambiguación del sentido, cuando la palabra misma —un pronombre— prácticamente no da ninguna pauta para decidir qué se entiende por esta palabra, sino toda la información viene de su contexto. Por ejemplo: «Juan tomó la torta de la mesa y la comió», ¿qué comió Juan? En ciertos contextos, a veces ni siquiera el pronombre aparece explícitamente en el texto: «Juan compró una casa. La cocina es grande», ¿cuál cocina? [4].

La *detección de correferencia* es una aplicación de tarea relacionada con la resolución de anáfora: es la tarea de detectar las palabras que refieren a la misma entidad o acción en el mundo descrito en el texto, por ejemplo: «María vió a Juan. El hombre se asustó. «Sr. Sánchez, soy yo», le dijo.»

---

3 [www.SENSEVAL.org](http://www.SENSEVAL.org)

La *detección de la implicación lógica (textual entailment* en inglés), es una tarea que ha recibido gran atención en los últimos años, y este interés sigue creciendo rápidamente. Se considera una tarea unificadora de muchas áreas de la lingüística computacional; es decir, una tarea que implica en cierta forma a muchas otras tareas y aplicaciones del procesamiento del lenguaje natural, tales como la desambiguación de los sentidos de las palabras, la resolución de anáfora, la recuperación de información, la respuesta a preguntas. Esta tarea trata de responder «sí» o «no» a una simple pregunta: basándose en dos textos, ¿el primero implica al segundo? Por ejemplo, «Juan vive en Paris», ¿implica «Juan vive en Francia»?<sup>4</sup>; «Juan se murió» ¿implica «Juan vivió»? Es fácil ver que la tarea trivialmente implica varios tipos de desambiguación: «Juan tomó la torta de la mesa y la comió», ¿implica «Juan comió la torta» o «Juan comió la mesa»?; «Juan trabaja en el banco» ¿implica «Juan trabaja en una organización financiera» o «Juan trabaja en una orilla del río»?

Finalmente, el *análisis semántico* propiamente dicho busca la descripción explícita de las relaciones lógicas entre las palabras, el significado y la interpretación de tales palabras. Una parte de esta tarea es la construcción de una red de las relaciones lógicas entre las entidades, las acciones y las propiedades mencionadas en el texto. Las palabras correferentes, corresponden al mismo vértice de tal red, el cual hereda las relaciones en las cuales está involucrada cada una de estas palabras en el texto. A pesar de un esfuerzo significativo destinado a esta tarea, y relativamente poco lo que se ha logrado. Existe un gran número de teorías al respecto, lo que indica que no existe ninguna lo suficientemente buena.

**Análisis pragmático.** Mientras que el análisis semántico todavía trata de estructurar lo que directamente dice el texto, el análisis pragmático proporciona metainformación sobre el texto: para qué se dice lo que se dice, cómo se organiza, qué estrategias emplea el autor para alcanzar el efecto deseado.

4 Estrictamente, la respuesta es «no», porque el topónimo puede referir a una de veinte ciudades en el mundo con este nombre, sin mencionar interpretaciones más exóticas como un virus llamado Juan que vive en Paris Hilton. Sin embargo, en la práctica usualmente se espera una interpretación razonablemente probable.

El *análisis de texto* es una evaluación del texto como una entidad completa, no dividida en oraciones. En esta fase se pretende entender el efecto que quería lograr el autor al comunicarle este texto al lector.

El *análisis de discurso* determina el papel lógico de cada parte del texto —oración, frase u otro fragmento— en la presentación completa de los argumentos del autor. Por ejemplo, en el texto «El libro es muy interesante. Juan lo leyó en una noche» la primera oración es la explicación para la segunda: lo leyó *porque* es interesante.

El *análisis y planeación de diálogo* es necesario cuando no se trata de un texto (monólogo) sino de intercambio de intervenciones de varias personas. Esta situación involucra los elementos que un texto normal no presenta, por ejemplo, la conservación de la información de un turno a otro, duración de los turnos y las estrategias para ceder o solicitar su turno, la estructura de los estímulos y las respuestas entre los participantes, etcétera.

Finalmente, el *análisis pragmático* propiamente dicho es la determinación de las intenciones del autor o hablante: para qué se dice lo que se dice, qué efecto espera lograr el autor con su texto. Por ejemplo, pragmáticamente con la frase «¿me puedes pasar la sal?» se pretende causar al escuchante pasar la sal, mientras que semánticamente se trata de una solicitud de información sobre las habilidades físicas del escuchante de hacerlo.

Las tareas enumeradas en esta sección pueden parecer simples, y cualquier niño las hace sin problemas. Sin embargo, cuando se intenta que las haga un programa de manera totalmente automática y autónoma, se presentan dificultades técnicas tan considerables que a pesar de todo el esfuerzo dedicado a la investigación correspondiente, el avance en muchas de estas tareas es todavía cuestión del futuro cercano o no tan cercano.

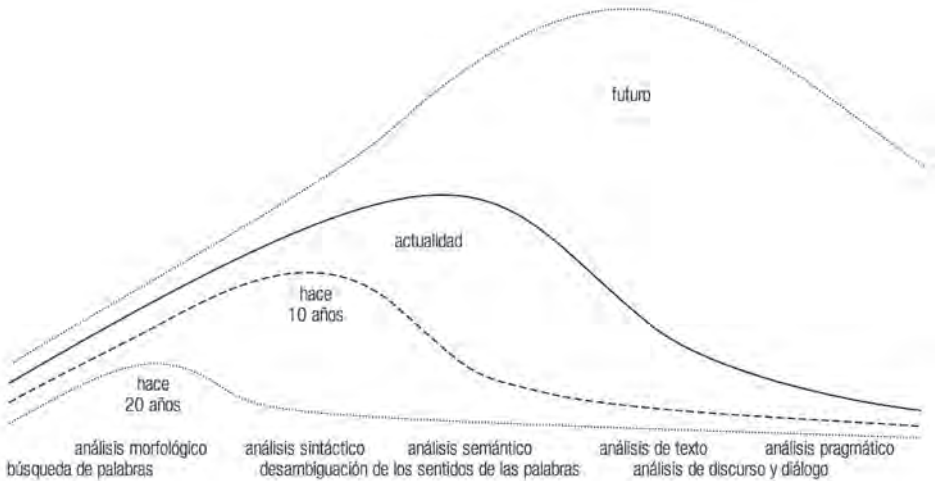
Históricamente en las etapas iniciales del desarrollo de la lingüística computacional, la mayor atención la han recibido las tareas más simples.<sup>5</sup> Con el tiempo estas tareas pasan a ser consideradas como más o menos resueltas; mientras tanto la mayoría de los investigado-

---

5 La tesis de doctorado del autor de este artículo, iniciada hace 20 años, fue en el área del análisis morfológico.

res se enfocan en las tareas más complejas que reflejan un mayor nivel de profundidad de entendimiento del lenguaje; es decir, la estructuración de mayores segmentos del texto o la consideración de un contexto más amplio, véase la Ilustración 1.

Ilustración 1. La evolución histórica aproximada de la inversión del esfuerzo en las tareas de diferentes niveles de profundidad. Obsérvese la evolución del área del mayor esfuerzo invertido, y la mayor actividad de la investigación al ir desde tareas más simples hacia las más complejas.



## MÉTODOS SIMBÓLICOS Y ESTADÍSTICOS

La lingüística computacional nació hace medio siglo como una rama de la lingüística que pretendía dar la descripción de los fenómenos lingüísticos o las recetas de análisis lingüístico del texto con tal claridad que incluso una máquina pudiera aplicar dicha información para realizar las tareas prácticas. Cuando en la escuela se le enseña al niño cómo hacer un análisis formal de la oración, la maestra da ejemplos de tal análisis o explicaciones poco específicas, poco detalladas y poco claras, y la interpretación de tal explicación se le deja a la intuición lingüística natural del niño. La práctica de la presentación del material



en los libros científicos en las humanidades tradicionalmente es similar: se le deja mucho a la intuición y a la interpretación del lector.

Pero para la computadora tal estilo de descripción no es aceptable —ella necesita reglas claras, con todas las posibles opciones explícitamente mencionadas, y con la definición de los conceptos usados en tales reglas. De hecho, esto es sano también para la misma ciencia: lo que es claro para la computadora también será claro tanto para el mismo autor como para los lectores. Además, varios fenómenos nuevos de gran importancia fueron descubiertos sólo tras una introspección en la búsqueda de la claridad completa: los fenómenos de los que los científicos no se daban cuenta porque nunca se hicieron preguntas tales como «¿cómo decido cuál palabra elegir?» o «¿qué específicamente significa el concepto que parece claro cuando veo los ejemplos?»

La lingüística computacional nació así como una ciencia de las reglas claras formuladas por los expertos humanos.<sup>6</sup> Al enfoque basado en las condiciones lógicas que o bien se cumplen o bien no (no se pueden cumplir parcialmente), y en los conceptos discretos que o bien son iguales o bien diferentes (no pueden ser «parecidos»), se les conoce como el enfoque *simbólico*. Con este enfoque resultó posible efectuar el análisis morfológico, sintáctico, semántico (hasta cierto grado), desarrollar los sistemas de recuperación de información e incluso los sistemas de traducción automática de alta calidad.

Sin embargo este enfoque tiene algunas deficiencias. Primero, el costo del desarrollo de un sistema de reglas lo suficientemente complejo se vuelve prohibitivamente alto. Segundo, el número de variantes que generan tales reglas en el análisis de un fenómeno ambiguo rápidamente se vuelve inmanejable.<sup>7</sup> El problema es que con las reglas de tipo lógico es difícil distinguir entre las variantes probables y las menos probables: un humano difícilmente puede razonar en términos de probabilidad, ni tampoco puede estimar correctamente las probabilidades de diferentes fenómenos. Los sistemas basados en el

---

6 Fue en los tiempos del florecimiento de los sistemas expertos en la inteligencia artificial, los cuales también constituyen una formulación clara del conocimiento del experto en el área.

7 Véase más arriba el ejemplo de un virus llamado Juan que vive en Paris.

enfoque simbólico resultaron ser demasiado rígidos y poco robustos: el sistema o bien da millones de variantes o bien —con más restricciones— ninguno.

Con la disponibilidad de enormes cantidades de textos en Internet, hace aproximadamente una década,<sup>8</sup> se descubrió la posibilidad de extraer la información sobre las palabras y los fenómenos lingüísticos a partir de los textos. Esto resultó una mina de oro para la investigación en la lingüística computacional y las técnicas de procesamiento del lenguaje natural avanzaron mucho. Los métodos estadísticos no necesitan el costoso esfuerzo por parte de los expertos líderes para construir los recursos léxicos y las gramáticas.

Más aún, los recursos construidos tienden a ser no cualitativos, como en el caso del enfoque simbólico, sino cuantitativos. No sólo se sabe que existe una palabra sino cuántas veces aparece; no sólo se sabe que existe una regla gramatical sino cuántas veces se usa; no sólo se sabe que dos palabras son sinónimas sino que se puede medir cuantitativamente el grado de su similitud. Esta información habilita el razonamiento estadístico no disponible para los recursos con la información cualitativa.

Como un ejemplo de tal razonamiento, consideremos el siguiente razonamiento falaz:  $P \Rightarrow Q$ ,  $Q$ , entonces  $P$ .<sup>9</sup> Se convierte en razonamiento válido cuando se trata de probabilidades:  $P$  es posible y  $Q$  es poco probable; dado  $P$  es muy probable que  $Q$ ,  $Q$ , entonces muy probable que  $P$ ,<sup>10</sup> lo que se sustenta en la regla de Bayes:

$$P(P|Q) = P(Q|P) \frac{P(P)}{P(Q)},$$

donde  $P(P)$  es la probabilidad de  $P$  y  $P(P|Q)$  es la probabilidad de  $P$  dado  $Q$ .

---

8 Obviamente había trabajos correspondientes desde hace dos o tres décadas, pero no eran la corriente prevalente.

9 Si Juan mató a Pedro, entraría en la casa de Pedro. Lo vieron entrar en su casa. Entonces, él lo mató.

10 Juan tenía enemistad con Pedro y nunca antes venía a su casa. Aquella noche lo vieron entrar en la casa. Entonces, es muy probable que él lo haya matado.

Este tipo de razonamiento les permite a los enfoques estadísticos resolver las ambigüedades que los enfoques simbólicos no pueden resolver. Por ejemplo, si se puede afirmar con alto grado de certeza que si Juan está en París, entonces está en Francia, a menos que el contexto presente evidencias de que se trata de otro lugar llamado París<sup>7</sup> —lo que es  $P(Q)$ . Hay que aclarar que el problema de los enfoques simbólicos no es que no puedan aplicar el razonamiento estadístico, sino que por lo general no disponen de la información necesaria para ello: los expertos humanos no son capaces de proporcionar tal información con un grado razonable de precisión.

Es por eso que los métodos estadísticos constituyen el tema de la gran mayoría de los trabajos de investigación en nuestros días, mientras que los métodos simbólicos casi han caído en el olvido durante la última década. Así la lingüística computacional contemporánea ha dejado de ser una rama de la lingüística, por lo menos de la lingüística en su forma tradicional, y se ha convertido en un alto grado en una rama del aprendizaje automático, un área de la ciencia de la Inteligencia Artificial.<sup>11</sup>

En la ciencia del aprendizaje automático se distinguen dos tipos de aprendizaje: el aprendizaje supervisado —basado en ejemplos— y el aprendizaje no supervisado. El primer tipo de aprendizaje requiere de un corpus de textos donde la tarea ya está resuelta: por ejemplo, para la tarea del etiquetado de las categorías gramaticales (*tagging*) la categoría gramatical de cada palabra ya esté marcada manualmente. El sistema entonces infiere las reglas usadas (quizá inconscientemente) por el individuo que hizo tal etiquetado, y luego puede continuar con la tarea sobre los textos nuevos, que no tienen el etiquetado.

La ventaja es que el anotador no tiene por qué ser experto en lingüística ni entender las razones subyacentes de sus acciones, sino que puede ser cualquier hablante nativo educado del lenguaje dado, al menos cuando se trata de tareas tan simples como el etiquetado de las categorías gramaticales. Las desventajas principales son dos: primero, todavía se necesita una cantidad enorme de trabajo manual, aunque

---

<sup>11</sup> véase [www.MICAL.org](http://www.MICAL.org), [www.SMIA.org.mx](http://www.SMIA.org.mx)

no de un experto; segundo, lo que se descubre en este proceso no es la información sobre el lenguaje sino la información sobre las opiniones subjetivas del anotador; se estudia la cabeza del anotador y no el lenguaje.

Estos problemas son superados en el otro tipo de aprendizaje —el aprendizaje no supervisado— que son las técnicas que no requieren preparación alguna de los textos sino que toman como entrada los textos planos tal cual se encuentran en, digamos, Internet. El resultado del aprendizaje no supervisado se parece a lo que hemos visto en las películas de ciencia ficción, cuando un marciano sale de su plato volador, graba con un dispositivo las palabras que le dirigen los terrícolas, y en un rato el dispositivo empieza a traducirlas al marciano y al revés. Sorprendentemente, las técnicas del aprendizaje estadístico, sobre todo el no supervisado, hacen tal escena cada vez más parecida a ciencia que a ficción [9].<sup>12</sup>

Sin embargo, la marcha triunfal de los métodos estadísticos no implica que no tengan sus deficiencias. Primero, con estos métodos no se aprovecha el conocimiento de los expertos en lingüística y la estructuración tradicional del objeto de nuestros estudios. Segundo, los métodos estadísticos son menos eficientes en el descubrimiento de estructuras muy complejas; como son, por ejemplo, las estructuras semánticas. Se puede esperar que esta mina de oro podría empobrecer pronto, lo que haría atractivo para los investigadores el hecho de combinar los métodos simbólicos, basados en la experiencia e ingenio humanos, con los métodos estadísticos, que pueden proveerles las características probabilísticas y cuantitativas.

## UNA RESEÑA BIBLIOGRÁFICA

Mayor información sobre los temas presentados en este capítulo se puede encontrar en varios libros, revistas, y memorias de congresos.

A los lectores que prefieren leer en español, como una ilustración de la investigación en el campo se les puede recomendar los libros

---

<sup>12</sup> véase también [www.reviews.com/review/review\\_review.cfm?review\\_id=129833](http://www.reviews.com/review/review_review.cfm?review_id=129833)

[3, 5], disponibles para su descarga desde las direcciones indicadas. El libro [3] describe una aproximación prevalentemente simbólica del análisis sintáctico del español, con intentos de combinarlo con aproximaciones estadísticas. El libro [5] incluye varios trabajos de índole prevalentemente estadística, de análisis computacional del corpus y de extracción automática de los recursos léxicos de los textos disponibles. Está en preparación la traducción al español del libro [3], la cual aparecerá en la dirección indicada cuando esté terminado; se puede obtener un borrador directamente de los autores. Artículos de investigación de los mismos autores se pueden encontrar en el servidor del Laboratorio de Lenguaje Natural y Procesamiento de Texto del CIC, IPN.<sup>13</sup>

Otra fuente de información son las memorias y las ponencias de varios congresos nacionales organizados en México sobre el tema. Uno de éstos es el Taller Nacional de Tecnologías del Lenguaje Humano, organizado en distintas ciudades del país por el Laboratorio de Tecnologías del Lenguaje del INAOE, Puebla. Otro es el Coloquio de Lingüística Computacional de la UNAM, organizado por el grupo de Ingeniería Lingüística. En ambos eventos se puede conocer en persona a los expertos y los estudiantes nacionales que trabajan en el tema, y familiarizarse con sus líneas de trabajo y sus proyectos actuales. En México también está en proceso de consolidación la Asociación Mexicana de Procesamiento del Lenguaje Natural, AMPLN.<sup>14</sup>

La revista *Procesamiento del Lenguaje Natural* editada por la Sociedad Española para el Procesamiento del Lenguaje Natural, SEPLN,<sup>15</sup> publica artículos originales de investigación tanto en español como en inglés, orientados hacia el lector preparado o experto. Sin embargo, un lector que recién se inicia en el tema puede encontrar información interesante sobre la actualidad de investigación y sobre los métodos empleados, así como ubicar a las personas y los grupos españoles y latinoamericanos relevantes. Los artículos de la revista están disponibles desde el servidor de la Asociación.

---

13 NLP.CIC.IPN.mx

14 www.AMPLN.org

15 www.SEPLN.org

Los lectores interesados en leer en inglés tienen mucho mayores posibilidades para obtener la información. Una introducción relativamente concisa al tema es el [2]. Al lector avanzado se le puede recomendar un libro de texto muy detallado el [6], el cual probablemente no sea necesario leer completo sino sólo las partes relevantes para el trabajo propio del lector. En cuanto a la aproximación puramente estadística al procesamiento del lenguaje, el libro clásico y también muy detallado es el [8]. A los lectores interesados específicamente en la recuperación de información, se les puede recomendar el clásico libro [1] y el más moderno [7], que presentan tanto el material básico como, en sus últimos capítulos, el de vanguardia.

Existen varias revistas de alta calidad dedicadas al procesamiento del lenguaje natural, entre las cuales destacan *Computational Linguistics*, *Research in Language and Computation*, *Language Resources and Evaluation*, *ACM Transactions on Speech and Language Processing*, *Natural Language Engineering*, *Journal of Quantitative Linguistics*, por mencionar algunas. Estas revistas publican artículos de investigación originales, orientados hacia los expertos.

Un panorama más amplio se puede ver en los mejores congresos, los cuales discuten los temas de investigación de frontera. La estructura de las memorias de un buen congreso da una idea clara de las direcciones de investigación que en la actualidad son las más activas y son una muestra representativa sobre las preguntas de investigación y los métodos que se aplican. Se le puede recomendar al lector que revise las memorias del congreso internacional CICLing,<sup>16</sup> organizado en México —y en ocasiones en otros países del mundo— por el Laboratorio de Lenguaje Natural y Procesamiento de Texto del CIC, IPN<sup>13</sup>. Desde el sitio Internet del congreso están disponibles las tablas de contenido de sus memorias de los 10 años de su historia, así como los resúmenes de los artículos. La lectura de los artículos de un congreso de esta naturaleza es probablemente la mejor manera para familiarizarse con el estado actual del área.

Otros excelentes congresos internacionales en el área son los organizados por la *Association for Computational Linguistics*

---

<sup>16</sup> [www.CICLing.org](http://www.CICLing.org)

(ACL):<sup>17</sup>ACL, NAACL, EACL, así como COLING,<sup>18</sup> RANLP, TSD<sup>19</sup> y NLDB.<sup>20</sup> Las memorias de estos congresos también están disponibles, ya sea en texto completo o sólo los resúmenes.

Finalmente, los postgrados en México y en el extranjero ofrecen los programas tanto del nivel de Licenciatura (UNAM<sup>21</sup>) como del nivel Maestría y Doctorado (CIC-IPN,<sup>13</sup> INAOE,<sup>22</sup> UNAM<sup>21,23</sup>).

## CONCLUSIONES

En este artículo hemos presentado las aplicaciones principales que en la actualidad trabaja el área del procesamiento automático de lenguaje natural, y hemos discutido también los métodos e ideas principales que se emplean para resolver los problemas en esta ciencia.

El procesamiento del lenguaje natural, llamado también la lingüística computacional, es una rama de la ciencia que se encuentra en la intersección entre la lingüística aplicada y la ciencia de la computación. La lingüística computacional y la lingüística teórica son dos ciencias que mutuamente se complementan y cada una aporta valor para el desarrollo de la otra.

En la etapa pasada la lingüística computacional se ocupaba principalmente de codificar claramente el conocimiento lingüístico, y en este sentido fue, como lo indica su nombre, una rama de la lingüística. En la etapa contemporánea se ha convertido en una rama de la ciencia de la inteligencia artificial; a saber, el aprendizaje automático. Un reto muy interesante es el aprendizaje no supervisado (de los textos disponibles no marcados manualmente) del conocimiento lingüístico y de los datos (diccionarios y gramáticas) necesarios para desarrollar los sistemas prácticos.

---

17 [www.ACLweb.org](http://www.ACLweb.org)

18 [nlp.shef.ac.uk/iccl](http://nlp.shef.ac.uk/iccl)

19 [www.TSDconference.org](http://www.TSDconference.org)

20 [www.NLDB.org](http://www.NLDB.org)

21 [www.IINGEN.UNAM.mx](http://www.IINGEN.UNAM.mx)

22 [ccc.inaoep.mx/labtl](http://ccc.inaoep.mx/labtl)

23 [turing.IIMAS.UNAM.mx](http://turing.IIMAS.UNAM.mx)

De acuerdo con esto existen dos corrientes en la lingüística computacional: la corriente simbólica y la estadística. La primera involucra el conocimiento humano y opera con los datos cualitativos. En esta corriente, las ambigüedades del análisis del texto se resuelven a través del análisis de un contexto más amplio y en los niveles más profundos del análisis: por ejemplo, la ambigüedad de la categoría gramatical de la palabra se resuelve durante el análisis sintáctico de la oración completa; la ambigüedad sintáctica de la oración se resuelve en la fase del análisis semántico del texto completo.

El enfoque estadístico tiende apoyarse en el aprendizaje automático de los datos necesarios y a operar con los datos cuantitativos utilizando los cálculos probabilísticos para resolver las ambigüedades. Este enfoque es el prevaleciente en la etapa contemporánea y ha producido resultados excelentes. Sin embargo tiene sus limitaciones, aunque se espera que en el futuro los dos enfoques habrán de combinarse.

Al lector interesado en obtener más información se le invita a consultar los libros de texto mencionados arriba, o las memorias de los mejores congresos, tales como el ACL o el CICLing, o bien integrarse en uno de los grupos de investigación o programas de posgrado existentes en el país.

## REFERENCIAS

- R. Baeza-Yates, B. Ribeiro-Neto. *Modern Information Retrieval*. Addison-Wesley, 1999.
- I. A. Bolshakov, A. Gelbukh. *Computational Linguistics: Models, Resources, Applications*. IPN-UNAM-FCE, 2004, 186 pp.; [www.gelbukh.com/clbook](http://www.gelbukh.com/clbook).
- S. N. Galicia Haro, A. Gelbukh. *Investigaciones en análisis sintáctico para el español*. IPN, 2007, 324 pp.; [www.gelbukh.com/libro-investigaciones](http://www.gelbukh.com/libro-investigaciones).



- A. Gelbukh, G. Sidorov. *On Indirect Anaphora Resolution*. Proc. PACLING-99, Canada, 1999, pp. 181-190; [www.gelbukh.com/CV/Publications/1999/PACLING-1999-Anaphora.htm](http://www.gelbukh.com/CV/Publications/1999/PACLING-1999-Anaphora.htm).
- A. Gelbukh, G. Sidorov. *Procesamiento automático del español con enfoque en recursos léxicos grandes*. IPN, 2006, 240 pp.; [www.gelbukh.com/libro-procesamiento](http://www.gelbukh.com/libro-procesamiento).
- D. Jurafsky, J. H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Published by Pearson Prentice Hall, 2008, 1024 pp.
- C. D. Manning, P. Raghavan, H. Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.
- C. Manning, H. Schütze. *Foundations of Statistical Natural Language Processing*. MIT Press. Cambridge, MA, 1999.
- D. W. Oard. The surprise language exercises. *ACM Transactions on Asian Language Information Processing (TALIP)* 2, 2, 2003, 79-84.



# Actualización del concurso simultáneo en el uso del lenguaje libre y del controlado: folksonomías y taxonomías.

JOSÉ ANTONIO MOREIRO GONZÁLEZ, JORGE MORATO LARA, SONIA SÁNCHEZ-CUADRADO, VICENTE PALACIOS.  
*Universidad Carlos III de Madrid, España*

## INTRODUCCIÓN

Tradicionalmente los sistemas de información han venido empleando diversos lenguajes que posibilitan la mejora en la recuperación de los documentos. Los soportes actuales no han podido decidir cual de ellos es el más idóneo debido a que todos presentan ventajas e inconvenientes de uso. Estos lenguajes ofrecen variaciones en su grado de estructuración, desde los lenguajes libres a los más controlados y formalizados, como se muestra en el siguiente esquema:

- Palabras-clave independientes que se usan en indización libre tanto por extracción como por asignación, uno de cuyos tipos son las folksonomías.
- Listas de palabras, como los glosarios, listas de nombres, diccionarios, donde ahora se incluyen también los anillos semánticos (*hypernymy tree* en *wordnet*) (Gómez, 2004).
- Facetas, categorizaciones y clasificaciones: lenguajes propiamente taxonómicos o con esquemas categorizadores (Zeng, 2004) (Z 39.19-2005).
- Grupos de relaciones, basados en relaciones entre términos y

conceptos, de estructura más compleja, entre los que se sitúan los tesauros, los *Topic maps* y las Ontologías.

En las próximas secciones se revisarán algunos de estos conceptos, su utilidad, así como su relevancia actual para las búsquedas en la Web.

## 1. Las palabras claves en la Web. Las folksonomías

Se denominan folksonomías a los conjuntos de palabras clave incorporadas y asignadas por cualquier internauta para colaborar en la indización de todo tipo de contenidos en un espacio compartido y abierto. La propuesta de este neologismo se atribuye a Thomas van der Wal quien fusionó las palabras *folk* (gente, popular) y taxonomía (Gestión —*taxis*— de la clasificación —*nomos*—), de forma que folksonomía viene a ser etimológicamente “Clasificación (mejor, indización) gestionada popularmente”. La asignación de estas etiquetas públicas se realiza sin ánimo de lucro y sin la supervisión de un organismo centralizador, de manera que una de las características de este lenguaje libre es la ausencia de estructuración entre los términos, salvo la formada por el conjunto que describe determinado objeto o concepto, si bien es cierto que cada término tiene sentido de forma individual.

Las folksonomías tienen un gran interés para mejorar la navegación y recuperación de todo tipo de materiales. Ejemplos de folksonomías se pueden ver en las etiquetas para *blogs* en *Technorati*, *Del.*

Tabla 1. Clasificación de las folksonomías según Hammond (2005)

CREADOR DE LAS ETIQUETAS	AJENO	Technorati HTML MetaTags	(Wikipedia)
	PROPIO	Flickr	CiteULike Del.icio.us Furl Frassie
		PROPIO CREADOR DEL CONTENIDO	AJENO

*icio.us social bookmarks*, para etiquetar sitios Web, o Flickr para fotografías.

Una clasificación popular es la que divide los tipos de folksonomías según la asignación y autor de las palabras-clave y del contenido, como se puede ver en Hammond (2005):

### 1.1 Beneficios de las folksonomías:

Existen numerosos motivos que explican la popularidad que experimentan estos recursos. Entre los cuales destacan su:

- **Simplicidad en la utilización:** Se trata de una solución simple para usuarios noveles en tareas de indización de contenidos, que no requieren del aprendizaje de un elevado conjunto de reglas para utilizarlas. En los lenguajes facetados, con amplio número de términos y de asociaciones, se complican las decisiones que deben tomarse para indizar un documento, lo que supone un notable coste cognitivo. En la mayoría de los casos esta inversión es muy elevada, por lo que prefieren describir sus documentos con palabras-clave libres (Wal, 2005).
- **Economía:** El carácter social y cooperativo de las folksonomías tiene gran rentabilidad debido a su bajo coste. Los internautas no persiguen lucrarse, sino beneficiarse de mejores búsquedas y navegación, cuantos más usuarios cooperen mayores ventajas se obtienen (Tonkin, 2006).
- **Adecuación al entorno Web:** es el único enfoque posible para indizar los enormes volúmenes de información existentes en la Web (Shirky, 2005), sobre todo cuando la información a indizar no es textual, haciéndose imprescindible la indización manual, es el caso de la indización de videos, fotos, etc.
- **Ejecución de consultas:** las búsquedas pueden ser más específicas, pues los términos asignados por los usuarios tienden a ser concretos. Si bien, uno de los rasgos principales de las taxonomías surge de su cualidad de asociar las verdaderas necesidades de los usuarios con la lengua, no de buscar la precisión (Quintarelli, 2005). Además permite recuperar, como ya se ha comentado,

material multimedia (solucionando en parte el problema de la Internet Invisible).

- **Simplicidad en la gestión:** a diferencia de los lenguajes controlados, los lenguajes libres son más sencillos y económicos por su escaso mantenimiento. Su evolución para incorporar nuevos términos es instantánea al carecer de una autoridad de control, por lo que están siempre actualizadas.
- **Flexibilidad:** la asignación de etiquetas (palabras-clave) a los recursos es flexible, ya que no se trata de un lenguaje precoordina-do y no cuenta con un vocabulario definido a priori.

## 1.2 Desventajas:

Las folksonomías, por el contrario, presentan asimismo notables desventajas, lo cual las relega a ser un instrumento útil en materiales no excesivamente críticos. Los principales inconvenientes se derivan de (Al-Khalifa, 2007; Smith, 2004):

- Ser etiquetas imprecisas, inexactas y ambiguas. Así, la asignación se realiza con criterios subjetivos. Frecuentemente se observan palabras clave que identifican a personas del entorno del usuario, por ejemplo, “Elena” o “Juan”.
- Muchas folksonomías solo permiten el uso de unitérminos (p.e. *Del.icios.us*)
- Existen problemas de sinonimia y homonimia que producen imprecisión en las búsquedas debido a un *recall* bajo.

## 1.3 Otros usos de las folksonomías

Las folksonomías tienen gran interés para estudios sociolingüísticos, pues ayudan a determinar las variaciones que se registran en determinado grupo de hablantes, por lo que permiten realizar estudios de la terminología empleada en un dominio concreto o por una comunidad específica. Una forma de realizar estos estudios consiste en observar la frecuencia de los términos más empleados en determinado contexto y, a su vez, destacar los términos no dominantes (*meta-noise*) que añaden comprensión semántica (Folksonomie, 2004). Además permi-

ten realizar estudios diacrónicos de utilización de determinado término. Un enfoque complementario es analizar los intereses de grupos reducidos de usuarios (Porter, 2004).

Otros estudios interesantes son la denominadas folksonologías, que tratan de crear ontologías a partir de folksonomías (Damme, 2007). Es evidente el valor de muchas de las palabras propuestas a la hora de servir como candidatas a formar parte de algún posible vocabulario controlado, por tratarse de conceptos empleados comúnmente por los usuarios de Internet. Aportan, por tanto, elementos de extracción de palabras libres que, tras su normalización, acabarían conformando los términos de algún tesoro. Para obtener el signo que mejor describe determinado concepto se utiliza la frecuencia de los términos en las folksonomía y las relaciones presentes debidas a la agregación de términos o por relaciones entre sus URL.

No todas las folksonomías son adecuadas para realizar estudios sobre la frecuencia de uso. Así las ontologías genéricas, aquellas en las que muchos usuarios indizan un mismo objeto, permiten un estudio más interesante, un ejemplo de estas ontologías es del.icios.us. Las ontologías específicas (Wal, 2004) permiten solamente a uno o a pocos usuarios clasificar los objetos, este es el caso de Flickr, por lo que su utilidad para medir el consenso de uso de un término es menor. Quizás el mayor inconveniente de esta aproximación es que el consenso tiende a darse más en términos genéricos que específicos, como puede comprobarse empíricamente en la aplicación Image Labeler de Google©.

## 2. SISTEMATIZACIÓN JERÁRQUICA DE CONOCIMIENTOS: FACETAS, CATEGORIZACIONES Y CLASIFICACIONES

La necesidad de definir unas categorías generales y jerarquizarlas para así mejorar la organización de la información es una constante a lo largo de los siglos. A continuación se muestra una breve sinopsis que analiza y muestra como ha sido esta evolución.

1) Clasificaciones Clásicas: Según Aristóteles, se pueden definir unas categorías generales en que agrupar todos los objetos, estas categorías son (Aristóteles, 1995): Sustancia, Cantidad, Cualidad, Relación, Lugar, Tiempo, Situación, Posesión, Acción, y Pasión. Posteriormente

Porfirio fue quien propuso disponerlas en forma de árbol, en razón de su género, subtipo y diferencia (Ferrater, 1999). El orden introducido en las categorías aristotélicas por el árbol de Porfirio supuso el primer mapa conceptual, pues representaba gráficamente las relaciones existentes entre los conceptos (Sowa, 2000), (Moreiro, 2006):

Tabla 2. Clasificación de Porfirio

<b>Género Supremo:</b> <b>Dif. Genérica:</b>	Sustancia material / inmaterial
<b>Gen. Subalterno:</b> <b>Dif. Genérica:</b>	Viviente sensitivo / insensible
<b>Gen. Próximo:</b> <b>Dif. Específica:</b>	Animal Racional / irracional
<b>Especie:</b>	Hombre Miguel, Isabel, etc.

En este esquema ya se observa un orden jerárquico. De manera, que el género supremo de los universales es la sustancia material o compuesta, descendiendo en la escala jerárquica de los universales de acuerdo con el orden marcado en el árbol para las categorías aristotélicas por *Genus* o Género supremo (*Top Term* o Macrodescriptor en un tesoro) y *Species* (Específicos de diferente nivel en un tesoro), hasta llegar a los individuos: Específicos, pasando por Géneros y especies subordinados (descriptores intermediarios o *Middle Terms*) y la Especie espacialísima: Genéricos.

Ramón Llull (Llull, 1998), propuso posteriormente elementos de relación, que se siguen empleando en la actualidad para asociar términos en los tesoros: *Utrum*; *Quid*; *De quo*; *Quare*; *Quomodo*; *Ubi*; *Quando*; *Quantum*; *Cum quo*; *Quale*, que además de ser el fundamento lejano de las propuestas lógicas de Port-Royal marcan buena parte de las relaciones asociativas existentes entre los conceptos que un tesoro define (Gayà, 1996). Port-Royal coincidía con Porfirio y Llull en considerar cinco predicables y no cuatro, porque incluían también la especie entre las ideas universales: géneros, especies,



diferencias, propiedades y accidentes. La influencia de estas sistematizaciones ha llegado hasta hoy, pudiéndose afirmar que la sistematización luliana del razonamiento sigue estando vigente en disciplinas como la Inteligencia Artificial (Boden, 1994), las redes semánticas, o la representación del conocimiento (Trillas, 1998).

Kant (Ranjan, 2007) también revisó las categorías Aristotélicas dividiéndolas en: Cantidad (unidad, pluralidad y universalidad); Cualidad (realidad, negación, limitación); Relación (sustancialidad y causalidad); y Modalidad (posibilidad, actualidad y necesidad).

2) Las facetas bibliotecarias: En las bibliotecas se han empleado frecuentemente las facetas para dar mayor flexibilidad a las clasificaciones bibliotecarias. Así la clasificación de Vickery, empleada en bibliotecas inglesas define seis facetas: objeto, parte, propiedad, proceso, operación y agente, y cada faceta se subdivide a su vez mediante una taxonomía. De cualquier modo las facetas que se han empleado con más asiduidad son las definidas por Ranganathan esto es tiempo, espacio, energía, materia y personalidad. Es curioso observar la similitud entre algunas de estas facetas y las categorías discutidas anteriormente.

Según se intentan adaptar las tecnologías a los nuevos formatos la necesidad de definir unas facetas generales permanece, así en el artículo “Clasificaciones Facetadas y Metadatos (I): Conceptos Básicos” de Hassan et al. (2003) se define una clasificación para recursos Web, en el que bitácoras, portales y listas de correo se clasifican según usabilidad, utilización de la Web semántica, temática, tipología e idioma.

En definitiva parece evidente que existen categorías generales que son necesarias para organizar el conocimiento, y que la estructuración jerárquica de estas categorías da un valor añadido a la clasificación original. Esto se ha traspasado a la Web actual en forma de:

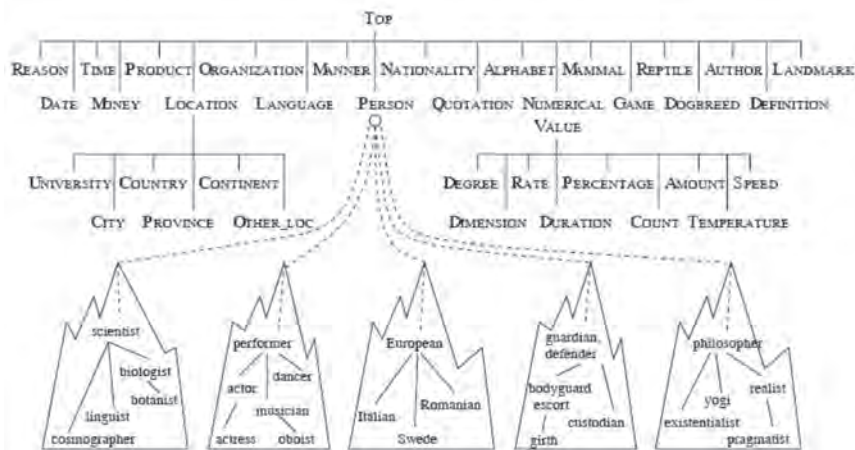
- Vocabularios de Metadatos. Los metadatos son descripciones que facilitan la recuperación, utilización o gestión de recursos de información. Así, los metadatos pueden ser utilizados para organizar recursos electrónicos o facilitar la interoperatividad. Los metadatos suelen estar agrupados en conjuntos de elementos denominados esquemas o vocabularios de metadatos. Cada esquema de metadatos se diseña con una finalidad concreta y se define

dentro de un contexto o espacio de nombres concreto. Dentro de dicho contexto, cada uno de los elementos del esquema posee una definición única. Es habitual el uso de varios espacios de nombres dentro de un mismo esquema, lo que permite el uso de los elementos definidos en otros esquemas, identificados mediante un nombre cualificado (*QName*), esto es: un identificador del esquema, seguido de dos puntos y el nombre del elemento en dicho esquema. Esta característica hace posible la reutilización y difusión de esquemas en la Web. Un ejemplo de estos vocabularios es *Dublin Core*, recurso que establece un conjunto reducido de categorías con las cuales se puede clasificar un recurso, estas son: Elementos (15): Título, Creador, Materia, Descripción, Editor, Colaborador, Fecha, Tipo, Formato, Identificador, Fuente, Lengua, Relación, Cobertura, y Derechos de autor.

- Los directorios de buscadores Web utilizan también unas categorías principales a partir de las cuales palntear las subdivisiones.
- Ontologías de alto nivel o genéricas: un tipo especial de ontología es la denominada de alto nivel. Estas ontologías son útiles para enlazarlas con ontologías de dominio más específicas. De nuevo, un análisis de las mismas muestra fuertes similitudes con las categorías genéricas discutidas en este apartado. Así (Gangemi et al. 2001), las nociones generales proponen agrupar los genéricos en: abstracto, concreto o relación. que la ontología de Cyc, incluida en CycKnowledge Base (Lenat and Guha, 1990), contiene unos 3000 elementos. Estos conceptos han sido agrupados en 43 categorías (*fundamentals, time and dates, spacial relations, etc.*). WordNet por su parte también subdivide su contenido en diferentes clases, entre las raíces están actividad, fauna, artefacto, atributo, cuerpo, conocimiento, comunicación, evento, sentimiento, comida, agrupamiento, lugar, motivo, objeto natural, fenómeno natural, persona, planta, posesión, proceso, cantidad, relación, forma, estado, sustancia y tiempo. De nuevo aparecen algunas constantes como son la sustancia, el tiempo, la forma, la cantidad o el lugar.
- Sistemas de pregunta respuesta: Uno de los desarrollos futuros de la Web consistirá en la evolución de los buscadores desde la

recuperación de documentos a la recuperación de respuestas. Y esto es factible si recuenta con la definición de grandes categorías de consultas. Los sistemas pregunta-respuesta también han podido categorizar a estas alcanzando un conjunto de categorías análogos. Un ejemplo se puede ver en la taxonomía de Pasca (2003)

Figura 1: Taxonomía de Pasca



Los sistemas OLAP (*On-Line Analytical Processing*), tienen también definidas estas categorías. Se trata de almacenes de datos estructurados para facilitar el procesamiento analítico de apoyo a la toma de decisiones estratégicas. De esta manera, los datos se organizan en torno a hechos (por ejemplo, ventas), cada hecho tiene unos atributos/medidas (por ejemplo: importe, cantidad, número de clientes, etc.) con mayor o menor granularidad a la hora de reflejar sus dimensiones (por ejemplo; CUÁNTO se ganó con la venta, CUÁNDO fue venta, DÓNDE se realizó venta, QUÉ se vendió, etc.) (Hernández, et al., 2004). Dicho de otro modo, de nuevo se detectan unos presupuestos universales que son constantes en todos los esquemas: como las facetas: cuándo, qué, dónde, etc, o los atributos, que suelen responder al cuánto.

## 2.1 Taxonomías

El concepto de taxonomía proviene del siglo XIX cuando con ellas se ordenaban, describían y clasificaban los seres vivos para las Ciencias naturales, partiendo de la especie como unidad de clasificación. Algo parecido sucedía con la Terminología científica pues, como consecuencia de la internacionalización progresiva de la ciencia, los científicos se esforzaban en sus reuniones por ordenar las nomenclaturas dentro de estructuras taxonómicas.

En la actualidad, las taxonomías se aplican con otro sentido en el mundo empresarial e institucional para organizar y gestionar los recursos de información digitales que como organizaciones complejas alojan en sus servidores Web, buscando categorizarlos, hojearlos y navegar por ellos. Así, se emplean términos autorizados en cada institución, con definiciones que usa una organización para clasificar sus contenidos (Corcoran, 2002). Los usuarios clasifican las materias dentro de jerarquías para hacer fácil la búsqueda de los recursos de información (Zhongong 2007).

Una taxonomía organiza no sólo los contenidos propios de una organización, sino también los servicios que ofrece, sus productos y cuanto se deriva de la experiencia y datos sobre los recursos humanos. De esta manera, resulta una red semántica de conceptos interrelacionados para cubrir con validez específica las necesidades empresariales y la forma con que los trabajadores se relacionan con la información (Conway, 2002). La aplicación práctica de las taxonomías se consigue cuando las usamos para navegar en la Web. Permiten a los usuarios acceder a ítems de interés específico enlazando recursos a partir de sus correspondientes categorías que posibilitan ir estrechando sus campos de búsqueda. Un ejemplo son los directorios tipo *Yahoo*© o *DMOZ*.

Las taxonomías funcionan dentro de un contexto específico, se basan en razones internas. Son flexibles y fáciles de modificar por funcionar tanto por jerarquía como por facetas. Integran nuevas áreas de interés y modifican fácilmente su estructura de acuerdo con las necesidades de cada momento. Como característica de los términos que las componen, se resalta que son categorías representadas en entradas

etiquetadas y orientadas al usuario. Por todo ello, podemos afirmar que las taxonomías generan sus estructuras jerárquicas de acuerdo con un contexto y unos usuarios determinados.

## 2.2. Las taxonomías en los lenguajes documentales: tesauros y ontologías

En 1876, Dewey (Dewey, 1979) presentó su *Clasificación decimal* que marcó el camino a seguir por los sistemas clasificatorios, precoordinados y de estructura jerárquica, y las *Rules for a dictionary catalog* de Charles Cutter (Cutter, 1962) que se anticipó a las listas de encabezamientos de materia y, de alguna forma, a los lenguajes controlados. Las teorías de Cutter han tenido mayor proyección hacia los lenguajes controlados que las clasificaciones debido a su carácter precoordinado y a su estructura asociativa, así como al control de vocabulario de aplicación específica a los conceptos y a la facilidad de uso para el usuario, frente a la rigidez arbórea de los sistemas clasificatorios, surgidos de la idea de Dewey para analizar el conocimiento humano. Más adelante apareció la Clasificación Decimal Universal (CDU), adaptación de la ideada por Dewey, para relacionar los conceptos del *Repertorio universal* de La Fontaine y Otlet por jerarquía, similitud o diferencia, por lo que, además de mostrar características de taxonomía, tenía en cuenta las asociación.

Fue durante la década de 1960 cuando, ante la aparición de las bases de datos como una de las aplicaciones para enfrentar el creciente número de publicaciones científicas y técnicas, el tesauro pasó a ocupar un lugar protagonista en la recuperación de los documentos. Se requerían soluciones que los sistemas tradicionales eran incapaces de suministrar, pues ahora había que representar los conceptos contenidos en los documentos, así como las relaciones existentes entre los conceptos, en una forma de lenguaje estandarizada, obtenida mediante el control de sinónimos y con una estructura sintáctica más simplificada que la del lenguaje natural, que representara el contenido informativo de los documentos de forma normalizada. Cualquier tesauro parte de una categorización del dominio que cubre y, por lo tanto, de una taxonomía del conocimiento temático que, en este sen-

tido, es una terminología jerarquizada. Las taxonomías están presentes en todos los *Esquemas, Tesoros, Modelos conceptuales y Ontologías*. Entendiendo por Taxonomía la clasificación o categorización de un conjunto de términos (en nuestro caso descriptores) de forma jerárquica, que establece una relación esquemática entre los objetos de generalización-especialización (Daconta, 2003).

### 2.3 Web Semántica y Ontologías

En 1998, Tim Berners-Lee publicó en la página principal del *World Wide Web Consortium* (W3C en adelante) su conocido artículo *Roadmap to the Semantic Web*. En dicho artículo se introducía por primera vez el término Web Semántica (*Semantic Web*) destacándose la necesidad de expresar la información de forma que ésta fuera procesable por máquinas. El artículo presentaba un conjunto de pasos hasta lograr una Web en la que el razonamiento fuera automático, llevado a cabo por máquinas, y distribuido. En 2001, Berners-Lee entregó un nuevo artículo *The Semantic Web* en el que presentó las principales características que tendrá la futura Web convencido de que las máquinas facilitarán nuevas prestaciones al mejorar su capacidad de procesar y comprender la información dispersa por la Web. Como solución a las limitaciones semánticas de la actual Web, propuso hacer procesable de forma automática el contenido de la Web, llegando así a la definición de Web Semántica:

*The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.* De modo que la Web Semántica no sería una nueva Web sino la extensión de la existente, mediante la adición de metadatos que describan la semántica de las páginas de forma procesable por máquinas.

Por ello, en la Web Semántica el usuario final de las ontologías no son las personas sino las aplicaciones informáticas. La forma de hacer que cualquier aplicación pueda leer y procesar los datos de un repositorio estableciendo identificadores de los documentos basados en URLs y un lenguaje con sintaxis estandarizada denominado XML (W3C, 2006). Además es necesario expresar el conocimiento de ma-

nera uniforme, algo que se consigue mediante tripletas expresadas en RDF (W3C, 2005), por último, se necesitan vocabularios controlados que aporten semántica y que denominen los conceptos de forma no ambigua, permitiendo la navegación a través de sus estructuras. Esto es posible gracias a las ontologías y a los vocabularios de metadatos. La semántica basada en vocabularios controlados y ontologías permitiría a las aplicaciones interoperar con cualquier repositorio de datos, convirtiendo la Web en una gran base de datos.

### 2.3.1 Ontologías

Aunque con cierta confusión en su concepto actual, una ontología o conjunto de conocimientos representado formalmente se fundamenta en la conceptualización de los objetos y entidades que existen en un área de interés, así como en las relaciones que existen entre ellos (Genesereth, 1987). Una ontología sirve, pues, para categorizar los conceptos propios de un dominio y sus relaciones (Gruber, 1994), estableciéndose, por tanto, como especificación explícita de una (o una parte de una) conceptualización que incluye un vocabulario organizado de términos.

Tanto tesauros como ontologías, tienen un conjunto de etapas comunes en su construcción entre los que destacan (Sanchez-Cuadrado, 2007):

- 1) Identificación de conceptos relevantes del dominio
- 2) Selección de los términos que mejor representan los conceptos
- 3) Creación de una taxonomía de conceptos
- 4) Identificación de otras relaciones no jerárquicas

Una ontología puede tener además una mayor riqueza semántica, ya que puede contener funciones calculadas, constantes, propiedades, instancias, axiomas o restricciones. La necesidad de esta mayor riqueza semántica se justifica en la capacidad de algunas ontologías de realizar inferencias o razonamientos. Así pues, la construcción de ontologías pasa por un proceso de jerarquización, que se efectúa mediante la subdivisión o agrupación de clases hasta alcanzar una taxonomía bien organizada (Noy, 2000). De forma que la organización taxonó-

mica aparece como condición mínima en la superestructuración de los lenguajes documentales de carácter combinatorio, formando la agrupación primera de términos en los tesauros y en las ontologías, y alcanzando niveles de esencialidad en los de carácter clasificatorio.

### 2.3.2 Implantación Actual de la Web Semántica

Lamentablemente, la Web Semántica, operativa desde 1999, no ha tenido el éxito esperado. Una búsqueda en *Google*® muestra que, en junio de 2007, existían 2,770.000 documentos con extensión RDF, 41,500 con OWL, 2,330 XTM, 4,540,000 RSS y 212,000 ATOM, es decir 7,5 millones de documentos en un conjunto de 10,000 millones de documentos. Con Swoogle<sup>1</sup>, un recuperador especializado en la Web Semántica, los resultados no son mucho mejores: 158,000 documentos semánticos conteniendo el término RDF y 2,323,857 con el término OWL.

Las causas son diversas, pero debemos considerar entre las principales:

- Falta de legibilidad de los lenguajes RDF y OWL, lo cual supone un cuello de botella para que los expertos validen las ontologías (Gómez-Pérez, 2004). En 2005, Mika (Mika, 2005) subrayó la importancia que tienen los usuarios (denominada Dimensión Social en el artículo original) para la aceptación de la Web Semántica. De hecho la Web Semántica tiene diferentes grados de complejidad en la creación de recursos, y esta complejidad es inversamente proporcional a la proximidad al usuario (Fig.2). En la figura 2 se muestra gráficamente este hecho, así existe una tendencia que provoca que cuando se incrementa la complejidad en la representación semántica se produce una disminución en la dimensión de contacto con el usuario. Esta dimensión social engloba diferentes afectos como usabilidad, legibilidad o necesidad de formación previa para su interpretación.

---

1 Swoogle <http://swoogle.umbc.edu/>



- Escasez de herramientas que faciliten la creación de documentos semánticos mediante formularios usables, un ejemplo de un entorno más amigable se puede ver con Protégé o con Tabulator.<sup>2</sup>
- La migración de folksonomías a folkontologías es un tema aún por desarrollar, aunque ya existen estudios (Damme, 2007; Matsuo, 2006). Básicamente, esta migración se hace mediante herramientas estadísticas (Bagelman, 2006) y lingüísticas, o incluso proponiendo una normalización para asignar etiquetas (Xu, 2006).
- La incorporación de técnicas semiautomáticas para la creación de Sistemas de Organización del Conocimiento basadas en PLN y Minería de Datos (Sánchez-Cuadrado, 2007), ya que la carencia de estos recursos, junto con la lentitud en su creación dificulta la implantación de la Web Semántica.
- Presencia de duplicidades en los Vocabularios de Metadatos y Ontologías, lo cual provoca la desconfianza y confusión del usuario que no sabe cuál es el vocabulario idóneo o más generalizado. Como ejemplo están los vocabularios de metadatos para expresar tesauros. Actualmente existen, entre otros, el SKOS-Core<sup>3</sup> del W3C, los PSI<sup>4</sup> de los Topic Maps, Zthes<sup>5</sup> y MADS<sup>6</sup>.

---

2 Tabulator: Async Javascript and Semantic Web. <http://dig.csail.mit.edu/2005/ajar/release/tabulator/0.8/tab.html>

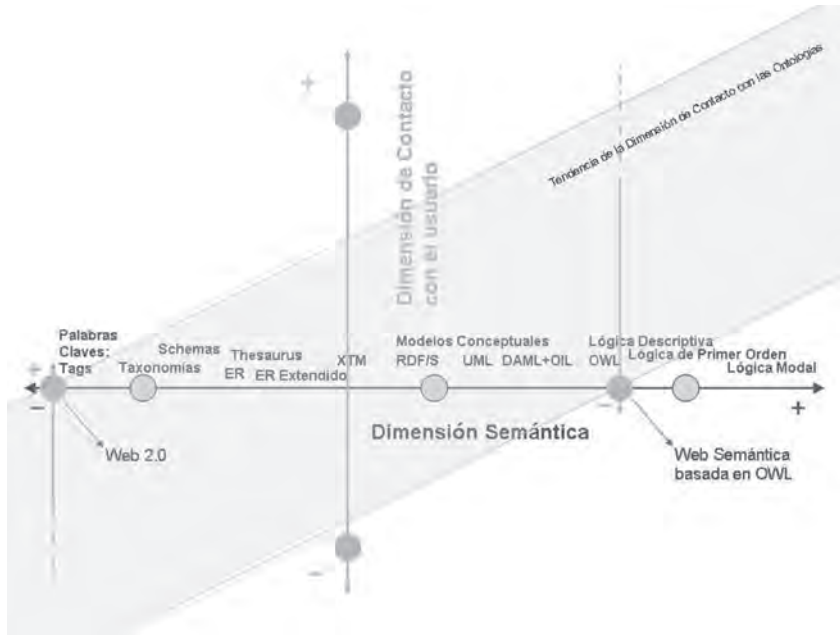
3 SKOS Core Guide <http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20050510/>

4 Published Subject Indicators For Modelling Thesaurii <http://www.techquila.com/psi/thesaurus/>

5 Zthes - specifications for thesaurus representation, access and navigation <http://zthes.z3950.org/>

6 Metadata Authority Description Schema (MADS) <http://www.loc.gov/standards/mads/>

Figura 2. Relación entre la Complejidad de la Web Semántica y la proximidad al usuario



### 3. UNA VISIÓN INTEGRADORA DE LOS VOCABULARIOS PARA LA WEB

Como consecuencia del recorrido hecho en este texto, aparece con claridad que:

- Las folksonomías tienen su propio espacio en la Web, y que si bien sería deseable un control del vocabulario para evitar ambigüedades semánticas e inexactitudes en la asignación de términos, su reemplazo parece complicado debido a la simplicidad de uso que presentan.
- Los vocabularios controlados y su encaje en la Web Semántica necesitan un gran desarrollo, pero la falta de apoyo popular hace que este se ralentice.

Como consecuencia, parece evidente que la única forma de popularizar la Web Semántica y de evitar los problemas derivados de la ambigüedad inherente a las folksonomías sería implicando a los usuarios en el desarrollo y empleo de los vocabularios controlados, pero de una manera simplificada. La mayoría de los esfuerzos en esta línea (Fumero, 2007) tratan de mejorar los interfaces y la usabilidad, y de mostrar a los usuarios los beneficios de estos vocabularios. A continuación se muestran ocho elementos en los que habría que avanzar para conseguir este propósito:

### 3.1 Analizar cuales son en la actualidad los documentos de la Web Semántica Social

Posiblemente, en la actualidad, el enfoque más próximo a la Web Semántica Social son los documentos RSS.<sup>7</sup> Estos documentos, que se cuentan por millones en Internet, aúnan la sindicación de contenidos (un objetivo típico de la Web 2.0) con la expresión de dicha sindicación en RDF (en el caso de RSS 1.0) o XML (en el caso de RSS 2.0). Sin duda, como en el caso del vocabulario Dublín Core (DCMI, 2007), el éxito se ha basado en dos factores: la aparición de lectores usables dirigidos al usuario final y la simplicidad del vocabulario de metadatos subyacente.

Las etiquetas Meta de HTML y los Microformatos<sup>8</sup> son dos soluciones que tienen respaldo popular, si bien son poco ambiciosos para crear la Web Semántica, muestran que la solución para implantar esta Web debe ser simple. La simplificación (Senso, 2007) se puede dar de diferentes formas, por ejemplo, prescindiendo de los lenguajes formalizados para ontologías a favor de expresar el contenido mediante documentos XHTML (una solución denominada microformatos), esta opción dado su carácter de solución local, no consigue uno de los principales objetivos de la Web Semántica, la interoperabilidad. Por el contrario, la interoperabilidad no se ve mermada en determinados documentos semánticos que han tenido un gran éxito gracias a la sim-

---

7 Wikipedia: RSS <http://es.wikipedia.org/wiki/RSS>

8 Wikipedia: Microformatos <http://es.wikipedia.org/wiki/Microformatos>

plicidad del vocabulario subyacente, un ejemplo son la sindicación de contenidos mediante RSS.

### 3.2 Promover la creación de ontologías de dominio por usuarios expertos

La creación de ontologías de forma colaborativa es un tema complejo, ya que las ontologías se basan en el consenso para determinar los conceptos e interrelaciones relevantes entre los mismos, y este consenso se dificulta a medida que aumenta el número de usuarios. Por otra parte, cuanto más específico sea el dominio más complicado es identificar expertos dispuestos a colaborar.

Una posible solución es que los organismos públicos incentiven y financien la creación de estos recursos. La opción de que exista capital privado para crearlos no parece realista, ya que el resultado deberá ser un producto gratuito y reutilizable por terceros.

### 3.3 Establecer mecanismos para centralizar los documentos semánticos en un repositorio común y eliminar duplicidades innecesarias

Actualmente existen al menos cuatro vocabularios de metadatos para expresar un KOS tipo tesoro. El SKOS-Core<sup>9</sup> del W3C, los PSI<sup>10</sup> de los Topic Maps, Zthes<sup>11</sup> y MADS<sup>12</sup>. La duplicidad para crear de forma descontrolada vocabularios controlados provoca la desconfianza y confusión del usuario que no sabe cuál es el vocabulario idóneo o más generalizado.

---

9 SKOS Core Guide <http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20050510/>

10 Published Subject Indicators For Modelling Thesaurii <http://www.techquila.com/psi/thesaurus/>

11 Zthes - specifications for thesaurus representation, access and navigation <http://zthes.z3950.org/>

12 Metadata Authority Description Schema (MADS) <http://www.loc.gov/standards/mads/>

### 3.4 Mejorar la dimensión de contacto en la utilización de documentos RDF-OWL

La dificultad de uso de algunos lenguajes como OWL no anima a los usuarios a colaborar. De hecho, según Gómez-Pérez et al. (2004) la validación de la ontología OWL, por parte de los expertos en el dominio, es uno de los cuellos de botella en la creación de ontologías.

El problema ha sido atajado en anteriores ocasiones con algunas propuestas como es el caso de la Wikipedia en el que las mejoras se realizan mediante diferentes perfiles para crear, editar o validar los nuevos contenidos, pero bajo un interface amigable, o como pueda ser la creación de Wikis de noticias mediante tags que sean codificados en RDF de forma transparente (Fuentes, 2007).

Algunas aplicaciones de Altova (SemanticWorks<sup>13</sup>) y Microsoft (InfoPath<sup>14</sup>) o los ya mencionados formularios para edición de RSS facilitan la incorporación de instancias en RDF

Otro problema relativo a este punto es la mejora de los interfaces para navegar por estos documentos, algo que intenta realizar el proyecto Tabulator.<sup>15</sup>

### 3.5 Desarrollar herramientas para convertir las folksonomías en KOS más complejos

Algunos recursos Web están siendo motivo de análisis para su evolución hacia KOS complejos, un ejemplo son las folkontology (Damme, 2007), que estudian mecanismos de evolución a partir de una folksonomía.

---

13 Altova SemanticWorks - visual RDF and OWL editor that autogenerates RDF/XML and N-Triples [http://www.altova.com/products/semanticworks/semantic\\_web\\_rdf\\_owl\\_editor.html](http://www.altova.com/products/semanticworks/semantic_web_rdf_owl_editor.html)

14 Información de Producto de Infopath 2003 <http://www.microsoft.com/spain/office/products/infopath/default.msp>

15 Tabulator: Async Javascript and Semantic Web <http://dig.csail.mit.edu/2005/ajar/release/tabulator/0.8/tab.html>

Otro ejemplo, sobre esta evolución se da en la Wikipedia<sup>16,17</sup>. Sin duda, la incorporación de las nubes de tags con herramientas lingüísticas y estadísticas puede ayudar en este proceso. Un ejemplo es Piggy Bank,<sup>18</sup> una aplicación que captura localmente las etiquetas de los RDF visitados, para organizarlo en una ontología local, y que puede ser compartida en el Semantic Bank (Senso, 2007). La aplicación permite poner *tags* a las URL seleccionadas para recuperarlas posteriormente. Un proyecto muy similar ha sido propuesto por el W3C bajo la denominación de Annotea,<sup>19</sup> instalable mediante un plugin en Firefox denominado Annozilla. El problema de ambas aproximaciones son las inherentes a las *tags*, es decir, la sinonimia y la polisemia.

### 3.6 Promover una arquitectura de colaboración con la implicación de los usuarios

El usuario debe tener claro que su colaboración para mejorar el recurso no implicará la obligación de pagar un sobreprecio para su futuro uso del recurso mejorado. Además, el usuario debe ser consciente de los beneficios que la Web Semántica le supondrá a corto plazo.

### 3.7 Establecer mecanismos para mejorar la fiabilidad de las búsquedas

El spam semántico será una realidad cuando se comience la transición hacia la Web Semántica. Un gran número de motores Web no utiliza en el posicionamiento las etiquetas Meta de HTML porque inicialmente las técnicas fraudulentas de optimización las utilizaron con asiduidad. Así, parece evidente que se deberán establecer sitios de confianza para evitar la desconfianza del usuario.

---

16 Harvesting Wiki Consensus - Using Wikipedia Entries [citeseer.ist.psu.edu/747700.html](http://citeseer.ist.psu.edu/747700.html)

17 Ontoworld.org [http://ontoworld.org/wiki/Main\\_Page](http://ontoworld.org/wiki/Main_Page)

18 SIMILE- My Piggy Bank [http://simile.mit.edu/wiki/Piggy\\_Bank](http://simile.mit.edu/wiki/Piggy_Bank)

19 W3C: Annotea Project <http://annotest.w3.org/>

### 3.8 Reducir la desconfianza de las empresas

Si los repositorios de todas las empresas de un dominio exportan los datos bajo el mismo modelo y vocabulario de metadatos, se incrementará la interoperabilidad para consultar distintos repositorios, lo cual facilitaría que la comparación de precios y servicios se pueda automatizar en diferentes grados. Esto puede provocar la desconfianza inicial de determinadas empresas, aunque es previsible que esta tendencia se reduzca cuando se alcance determinado peso crítico de la Web Semántica, ya que se correrá el riesgo de quedar relegado fuera del mercado.

## CONCLUSIONES

El enorme número de documentos que pueblan actualmente Internet hace necesario el empleo de diferentes lenguajes que varían en su grado de estructuración. Cada uno de estos lenguajes tiene su sitio dependiendo de la funcionalidad perseguida y de los recursos disponibles. Así, para describir los materiales multimedia no críticos desde el punto de vista de su valor se han impuesto las folksonomías. Las taxonomías y las facetas han ocupado su lugar para organizar portales Web. Las ontologías se hacen necesarias para hacer de la Web una gran base de datos en la que cualquier aplicación pueda comprender la información allí disponible.

Por otro lado, los lenguajes han tenido que ser transformados, creándose relaciones adaptables a cada dominio, así como la ampliación del concepto de tesoro de descriptores mediante la admisión de nuevas categorías gramaticales que han enriquecido con nuevos matices la semántica del mapa conceptual y, desde luego, aumentando con nuevas categorías las relaciones interconceptuales que han alcanzado incluso a los recursos de información, y que han extendido las posibilidades de asociación conceptual aproximándolas a la riqueza casuística del lenguaje natural. Sin embargo, la razón fundamental de las relaciones entre los términos de los lenguajes documentales sigue basándose en la estructuración jerárquica, tal como se establece para la terminología propia de un campo científico. La clasificación ha sido

estudiada por los principales autores filosóficos cuando se han acercado a poner las bases retóricas de los discursos. Esos principios siguen siendo fundamentales a la hora de organizar los lenguajes combinatorios, e incluso en sus evoluciones, pues si la diferencia entre los tipos de lenguaje parte de las posibilidades aumentadas de asociar términos, e incluso del concepto de término que se tenga, lo común sigue siendo la organización jerárquica en taxonomías, que afecta comúnmente a las clasificaciones jerárquicas, a los lenguajes combinatorios y a las propias ontologías.

Otra alternativa es la que ha aparecido con la asignación de palabras libres por los propios usuarios a los contenidos digitales que la red difunde y cuyo contenido sería imposible de analizar de otro modo. Las Folksonomías vienen a cubrir las necesidades de indización de los documentos Web que no son atendidos por los grandes servicios de pago o públicos. En este sentido suponen una solución popular al problema de los legítimos intereses de grupo en los documentos situados fuera de los cauces de circulación controlada o económicamente muy productivas.

Las folksonomías han venido a renovar las formas de indizar, pues han distribuido su responsabilidad entre los usuarios y han impuesto métodos descentralizados, alejados de cualquier jerarquía sistemática. Si bien actualmente se hacen necesarias técnicas que aproximen las folksonomías, propias de la Web 2.0, a unos lenguajes controlados, eliminándose problemas propios de los lenguajes libres, como la sinonimia, la homonimia y la ausencia de niveles de estructuración de los términos entre si. En el documento se exponen algunas recomendaciones para conservar la percepción de simplicidad de las palabras claves pero dentro de un entorno de gestión orientado a la implantación de lenguajes controlados.

## AGRADECIMIENTOS

Los resultados del presente artículo han podido obtenerse gracias a la ayuda financiera de la Comunidad de Madrid a través de su programa 2007/04055/001



## BIBLIOGRAFÍA

- (Al-Khalifa, 2007) Al-Khalifa, H.- Automatic document level semantic metadata annotation using folksonomies and domain ontologies. 2007. [http://eprints.ecs.soton.ac.uk/14181/01/Hend\\_Thesis.pdf](http://eprints.ecs.soton.ac.uk/14181/01/Hend_Thesis.pdf)
- (Aristóteles, 1995) Aristóteles.- *Tratados de Lógica. Organon*. Madrid: Gredos, 1995. v. 1: 45.
- (Begelman, 2006) Begelman, G.; Keller, P. and Smadja, F.- Automated Tag Clustering: Improving search and exploration in the tag space. WWW2006, May 22-26, 2006, Edinburgh, UK. <http://www.raw-sugar.com/www2006/20.pdf>
- (Berners-Lee, 1999) Berners-Lee, T. A roadmap to the Semantic Web. Disponible en: <<http://www.w3.org/DesignIssues/Semantic.html>>. [Consultado en junio de 2007].
- (Berners-Lee, 2001) Berners-Lee, T.;Hendler, J.; Lassila, O.- The Semantic Web. Scientific America, 2005, vol. 284, num. 5, pp. 34-43. ISSN 0036-8733.
- (Boden, 1994) Boden, M. (comp.).- *Filosofía de la Inteligencia Artificial*. México: Fondo de Cultura Económica, 1994.
- (CiteULike) CiteULike.com. A free online service to organise your academic papers, en [www.citeulike.org/](http://www.citeulike.org/). [consulta 18-03-2007].
- (Conway, 2002) Conway, S. y Sligar C.- Building taxonomies, en su *Unlocking knowledge assets*. Redmont: Microsoft Press, 2002: 105-124.
- (Corcoran, 2002) Corcoran, M.- Industry insights: taxonomies, hope or hype?, en *Online*, 2002, 26, nº 5: 76-79.
- (Cutter, 1962) Cutter, Ch.- *Rules for a dictionary catalog*. 4th ed. London: Chaucer House, 1962.

**Primer Simposio Internacional sobre Organización...**

- (Daconta, 2003) Daconta, M.; Obrst, L; y Smith, K.- *The Semantic Web. A guide to the future of XML, Web services, and Knowledge management*. Indianapolis: Wiley Publishing, 2003: 145.
- (Damme, 2007) Damme, C. et al.- *Ontology: An Integrated Approach for Turning Folksonomies into Ontologies*. 2007. <http://www.heppnetz.de/files/vandammeheppsior-paes-folksonology-semnet2007-crc.pdf>
- (DCMI, 2007) DCMI. Dublin Core Metadata Initiative. 2007 <http://es.dublincore.org/>
- (Dewey, 1979) Dewey, M.- *Decimal classification and relative index*. 19th ed. Albany: Forest Press, 1979.
- (Ferrater,1999) Ferrater Mora, J.- *Diccionario de filosofía*. Nueva ed. rev., aum y act. por Josep-Maria Terricabras; supervisión de Priscilla Cohn Ferrater. Barcelona: Ariel, 1999. v. 1: 49.
- (Folksonomie, 2004) Folksonomie. Many 2 many. A group weblog on social software, en [www.corante.com/many/archives/2004/08/25/folksonomy.php](http://www.corante.com/many/archives/2004/08/25/folksonomy.php). 2004 [Consulta 15-03-2007].
- (Fuentes, 2007) Fuentes, D. et al.- *CoolWikNews:More than meets the eye in the XXI century journalism. Emerging technologies form semantic work environments: techniques, methods, and applications*. Idea group, Germany, 2007
- (Fumero, 2007) Fumero, A Roca G y Encinar, J.- *Web 2.0. Colección Fundación Orange*. 2007. [http://www.fundacionauna.com/areas/25\\_publicaciones/indice\\_web2.asp](http://www.fundacionauna.com/areas/25_publicaciones/indice_web2.asp)
- (Gayà, 1996) Gayà, Jordi. El arranque filosófico del Ars lulliana. Constantes y fragmentos del pensamiento lulliano, en Domínguez, F. y Salas, J. de (eds).- *Actas del simposio so-*

*bre Ramon Llull en Trujillo (1994)*. Tübingen: Max Niemeyer, 1996: 1-8.

(Genesereth, 1987) Genesereth & Nilsson.- *Logical Foundation of Artificial Intelligence*. Los Altos (Ca.): Morgan Kaufmann Publishers, 1987.

(Gómez, 2004) Gómez, F.- Grounding the ontology on the semantic interpretation algorithm, en *Proceedings of the Second International WordNet Conference*. Masaryk University, Brno, 2004: 124-129.

(Gómez-Pérez, 2004) Gómez-Pérez, A. et al.- *Ontological Engineering*. Springer, London, 2004.

(Gruber, 1994) Gruber, T.R.- Toward principles for the design of ontologies used for knowledge sharing, en Guarino, N. y Poli, R. (Eds.).- *International Workshop on Formal Ontology, Padova, Italy. Revised August 1993*. Publicado en Guarino, N. y Poli, R. (Eds.).- *International Journal of Human-Computer Studies*. Special issue on Formal Ontology in Conceptual Analysis and Knowledge Representation 1994. Disponible en [http://ksl-eb.stanford.edu/KSL\\_Abstracts/KSL-93-04.html](http://ksl-eb.stanford.edu/KSL_Abstracts/KSL-93-04.html) [consulta 14-03-2003].

(Hammond, 2005) Hammond, T., T. Hannay, B. Lund and J. Scott.- Social Bookmarking Tools (I): A General Review. *D-Lib Magazine*, 2005, 11(4).

(Hassan, 2003) Hassan Montero, Y.; Martín Fernández, F.J. y Martín Rodríguez, O.- Clasificaciones Facetadas y Metadatos (I): Conceptos Básicos. No Solo Usabilidad [en línea], 2. (28 Febrero 2003). 2003. [http://www.nosolousabilidad.com/articulos/clas\\_facetadas1.htm](http://www.nosolousabilidad.com/articulos/clas_facetadas1.htm)

(Hernández, 2004) Hernández Orallo, José.; Ramirez Quintana, Mª José; Ferri Ramirez, César - *Introducción a la Minería de Datos*. Madrid. Pearson Prentice Hall, 2004.

**Primer Simposio Internacional sobre Organización...**

- (Matsuo, 2006) Matsuo, Y. et al.- Spinning multiple social networks for semantic web. In Proc. AAAI-06, 2006.
- (Mika, 2005) Mika, P.- Ontologies are us: A unified model of social networks and semantics. Proceedings of the 4th International Semantic Web Conference (ISWC 2005), LNCS, 3729, Springer-Verlag, 2005. <http://www.cs.vu.nl/~pmika/research/papers/ISWC-folksonomy.pdf>
- (Moreiro, 2006) Moreiro González, José Antonio; Morato Lara, Jorge; Sánchez Cuadrado, Sonia; Rodríguez Barquín, Beatriz A.- Categorización de los conceptos en el análisis de contenido: su señalamiento desde la Retórica clásica hasta los *Topic Maps*, en *Investigación Bibliotecológica: archivonomía, bibliotecología e información*, 2006, 20, nº 40: 13-31.
- (Noy, 2000) Noy, N. y McGuinness, D. L.- *Ontology development 101: a guide to creating Your First Ontology* [Página web]. Disponible en: [http://protege.stanford.edu/publications/ontology\\_development/ontology101-noy-mcguinness.html](http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html) [Consulta 14-03-2007].
- (Porter, 2004) Porter, J.- Controlled vocabularies cut off the long tail, en [http://bokardo.com/archives/controlled\\_vocabularies\\_long\\_tail/](http://bokardo.com/archives/controlled_vocabularies_long_tail/). 2004. [consulta 18-03-2007].
- (Ranjan, 2007) Ranjan Hatua, Sudip. Categories in Classification. [www.geocities.com/sudiphatua/class.html](http://www.geocities.com/sudiphatua/class.html)
- (Sánchez-Cuadrado, 2007) Sánchez-Cuadrado, S.- Definición de una metodología para la construcción automatizada de un sistema de organización del Conocimiento. 2007. Tesis Doctoral Universidad Carlos III.
- (Senso, 2007) Senso José A.- Navegadores semánticos o semantizar el navegador. ThinkEPI. 2007. <http://www.thinkepi.net/repositorio/navegadores-semanticos-o-semantizar-el-navegador/>

- (Shirky, 2005) Shirky, C.- Folksonomies + controlled vocabularies, en [http://www.corante.com/many/archives/2005/01/07/folksonomies\\_controlled\\_vocabularies.php](http://www.corante.com/many/archives/2005/01/07/folksonomies_controlled_vocabularies.php). 2005, [Consulta 14-03-2007].
- (Smith, 2004) Smith, G.- Folksonomy: social classification, [http://atomiq.org/archives/2004/08/folksonomy\\_social\\_classification.html](http://atomiq.org/archives/2004/08/folksonomy_social_classification.html). [Consulta 14-03-2007].
- (Sowa, 2000) Sowa, J. F.- *Knowledge representation: Logical, Philosophical and Computational Foundations*. Pacific Grove: Brooks / Cole Thompson Learning, 2000.
- (Tonkin, 2006) Tonkin, Emma.- Folksonomies: the fall and rise of plain-text tagging, en *Ariadne*, 47. <http://www.ariadne.ac.uk/issue47/tonkin/intro.html> [Consulta 07-07-2007].
- (Trillas, 1998) Trillas, E.- *La inteligencia artificial*. Madrid: Debate, 1998.
- (W3C, 2005) W3C- Primer: Getting into RDF & Semantic Web using N3. 2005. <http://www.w3.org/2000/10/swap/Primer>
- (W3C, 2006) W3C- Extensible Markup Language (XML). 2006 <http://www.w3.org/XML/>
- (Wal, 2004) Wal, Thomas Van der.- Explaining and Showing Broad and Narrow Folksonomies: <http://www.vanderwal.net/random/entrysel.php?blog=1635>. 2004. [consulta 23-07-2005].
- (Wal, 2005) Wal, Thomas Van der.- *Off the Top: Folksonomy Entries*, en <http://www.vanderwal.net/random/category.php?cat=153>, 2 de noviembre de 2005 (Consultado el 11 de noviembre de 2005)
- (Xu, 2006) Xu, Z. et al.- Towards the Semantic Web: Collaborative Tag Suggestions. WWW2006, May 22-26,

***Primer Simposio Internacional sobre Organización...***

2006, Edinburgh, UK. <http://www.rawsugar.com/www2006/13.pdf>

(Z 39.19-2005) Z 39.19-2005: NISO (National Information Standard Organization.- *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*. NISO Press, Bethesda, Maryland, U.S.A., 2005: <http://www.niso.org/standards/resources/Z39-19-2005.pdf>

(Zeng, 2004) Zeng Lei, Marcia, y Chan, L. M.- Trends and issues in establishing interoperability among knowledge organization systems, en *Journal of the American Society for Information Science and Technology*, 2004, 55, nº 5: 377-395.

(Zhongong, 2006) Zhongong, W., Chaudry, A. S., y Khoo, C.- Potential and prospects of taxonomies for content organization, en *Knowledge Organization*, 2006, 33, nº 3: 160-169.

# Importancia de la terminología para localizar y recuperar información científica.

IMELDA VELÁZQUEZ MONTES

*Facultad de Química.*

*Universidad Nacional Autónoma de México.*

## INTRODUCCIÓN

La información digital actualizada es un medio muy importante, para que los investigadores puedan mantener sus trabajos en la frontera del conocimiento; en virtud de la velocidad con la que se publican y difunden los descubrimientos en el ámbito científico, la importancia en el manejo de esta información y la rapidez con la cual se puede tener acceso a ella, es de gran beneficio para los investigadores, profesores y estudiantes; su uso permite incrementar su productividad, colaborando de ese modo al incremento del conocimiento.

En México es posible lograr estos beneficios, en la Universidad Nacional Autónoma de México y específicamente en la Facultad de Química, dentro de la Hemeroteca se ofrecen Servicios de Información Digital (SID), que tiene como objetivo principal ampliar y diversificar los servicios de información digital principalmente a la comunidad universitaria.

En estos servicios, no solo se encuentra la información de divulgación nacional, sino también internacional ya que cuenta con acceso a bases de datos a nivel mundial. Donde el principal objetivo es poder brindar un buen servicio, donde se obtenga de manera ágil y rápida

información actualizada en la mayor cantidad de bases de datos, para un mejor aprovechamiento de las mismas.

La información química que se produce en el mundo es enorme, se calcula en más de las tres cuartas partes del total de la información científica en todas sus modalidades; lo que equivale a montañas de documentos de Química o de cualquiera de las áreas relacionadas con la misma. Sería imposible que una persona pudiera revisar la información, ni siquiera trabajando las 24 horas del día los trescientos sesenta y cinco días del año. Es por eso que las bases de datos que tratan de incluir toda esta información, ya no pueden ser publicadas en formato impreso; un ejemplo es la base de datos más grande a nivel mundial, llamada Chemical Abstracts, que próximamente dejará de imprimirse, ya que por su gran tamaño, no es posible seguirse consultando en forma impresa. Desde hace tiempo el Chemical Abstracts se produce en varios formatos, como son los discos compactos, la versión llamada STN Easy, etcétera, hasta llegar a la producción de softwares especiales para el manejo de esta información, como el programa llamado STNEXPRESS y recientemente se diseñó un software más amigable llamado SciFinder Scholar, diseñado para que el usuario final lo utilice de forma fácil y dinámica.

A pesar de la facilidad que puede brindar el uso de herramientas tecnológicas para la localización y recuperación de información científica tan diversa y abundante, se presentan los siguientes problemas:

- 1 Si el investigador que publica los resultados de sus trabajos no utiliza la terminología apropiada, que los editores llaman "Key words" o palabras clave, para que sirvan como un verdaderos descriptores del conocimiento producido, de manera que permita ser identificado plenamente, para ser buscado y consultado por otros investigadores o lectores, su trabajo corre el riesgo de quedar en el olvido, simplemente porque nadie más podrá localizarlo.
- 2 Si los editores no establecen criterios claros y precisos de indización y clasificación de los artículos publicados, sobre todo en formato digital, de tal forma que se les pueda localizar con acierto, usando para ello la terminología apropiada, este hecho también



puede provocar que los trabajos científicos sean enterrados para siempre.

3 Las llamadas nuevas Tecnologías de Información y Comunicación (TIC) proporcionan herramientas útiles para que los usuarios con cierta facilidad busquen y encuentren en las bases de datos, los artículos que necesitan para realizar sus investigaciones, pero, que pasa al no utilizar la terminología adecuada para ello, se corre el riesgo de no poder acceder a documentos que puedan ser clave para la realización de su trabajo.

4 En el caso de la Química, donde las reglas de nomenclatura son complicadas y difíciles de entender, porque su organización lógica se basa en las propiedades de los compuestos, esto es, en el forma de reaccionar dependiendo de las condiciones en las que se encuentran; por lo tanto, la localización de un compuesto se dificulta demasiado, sobre todo al no utilizar la terminología precisa.

5 Considerando que al pasar el tiempo, existe cada día mayor interacción entre las ciencias y cada día se van perdiendo las barreras entre ellas, de manera que ahora se sintetizan compuestos órgano metálicos que tienen actividad biológica o fisiológica, incidiendo en la vida social y económica de los países; esto genera una situación complicada para el buscador de información, ya que aún cuando se usen bases de datos multidisciplinarias, actualmente no existe la base de datos que agrupe varias especialidades, como Química, Biología, Medicina, economía y sociología; por tanto, si no se usan las bases de datos y los términos adecuados, la consulta fracasará.

La finalidad de este trabajo es plantear las diferentes formas con las que se pretende abordar la problemática existente, de manera que no sea un obstáculo para localizar y recuperar información requerida, en todas sus modalidades.

Partiendo del hecho de que las ciencias, en su mayoría cuentan con un lenguaje especializado, inherente a las mismas, además de que el conocimiento está sistemáticamente organizado en cada una de ellas.

En el caso de la Química existen verdaderos tratados, que intentan organizar en forma lógica y característica a todos los compuestos

existentes; se plantea como una alternativa, el hecho de que los buscadores de información deben ser expertos en la ciencia donde pretendan localizar y recuperar información, así en el caso de la Química se observa que se contratan personas con posgrado en Química, con la particularidad en el manejo de la terminología apropiada y de los sistemas de organización de la información, para especializarse como buscadores de dicha información.

## DESARROLLO

Los servicios de información digital (SID), proporcionan servicios de información competitivos a nivel nacional e internacional que apoyan los programas académicos y de investigación, promoviendo los valores y habilidades de los estudiantes y profesionales, definidas en la misión de la Universidad.

Basándose en catálogos electrónicos y bases de datos que integran los servicios de información bibliográfica digital que adquiere la Dirección General de Bibliotecas (DGB), se ofrecen los nombres de las diferentes bases de datos a la comunidad universitaria, presentándolas en un formato accesible para el usuario.

Algunos sistemas contienen artículos de texto completo y otros se encuentran relacionados con algún servicio de documentación que permite la obtención inmediata del documento localizado en la base de datos. La información que se puede obtener en este servidor es muy variada, ya que cuenta con información de diferentes bases de datos, revistas electrónicas y catálogos globales; ayudándose del uso de nuevas tecnologías para poder realzar una búsqueda más ágil y rápida.

Es indispensable que la(s) persona(s) encargada(s) de brindar los servicios propuestos, tengan la suficiente preparación, tanto en el manejo de la terminología química, como en los sistemas de nomenclatura de los compuestos sintetizados en toda la historia de la humanidad, para tener éxito en dicha empresa.

## RESULTADOS

En los catálogos de las bases de datos se encuentra contemplada la siguiente información:

- Nombres de las bases de datos: Se presentan en la lengua y con las características tipográficas originales.
- Cobertura: Se refiere a los años que cubre la información contenida en cada base de datos.
- Número de registros: Proporciona el tamaño de la base de datos, ayudando al usuario a saber escoger entre los datos que aparecen y la información que necesita.
- Sistema: Ubica los conjuntos de bases de datos que pueden ser consultados con un mismo software de recuperación de información.

Entre todos ellos están el autor del artículo, él o los temas, la fecha de publicación, desde luego, el título de la revista, el título de los artículos y otros.

Como por ejemplo:

Base de datos	Cobertura	No de revistas o registros incluidas	Sistema en que se encuentra	Texto completo
<b>34.Health Source Plus</b>  Esta base de datos contiene información especializada en temas de nutrición, ejercicio y cuidados médicos, fármacos y alcoholismo.	1984 a la fecha	Más de 200 revistas  400 000 artículos	EBSCO	Si

Dentro de las principales bases de datos se encuentra “Elsevier” publicada por uno de los más importantes editores científicos del mundo, así como Chemical Abstracts, siendo esta última para el área científica, la base de datos más completa, que tiene registros de más

### *Primer Simposio Internacional sobre Organización...*

de once mil títulos de revistas internacionales arbitradas, patentes y libros, ésta base, se constituye en la base de datos más consultada tanto para localizar información como para encontrar la totalidad de las citas de trabajos publicados en el ámbito de las ciencias exactas.

Otro ejemplo:

Base de datos	Cobertura	No de revistas o registros incluidas	Sistema en que se encuentra	Texto completo
<b>22.Dissertation Abstracts</b> Contiene todas las tesis doctorales presentadas en los últimos 137 años en instituciones reconocidas en los Estados Unidos; Contiene tesis de maestría y tesis de doctorado en idiomas extranjeros.	1861 a la fecha	1400000 registros	OCLC	Una parte es gratis. Se puede comprar la tesis completa

### TÍTULOS DE REVISTAS ELECTRÓNICAS DE CONSULTA DIRECTA

Los creadores y los editores de recursos de información en red, no utilizan ni formatos, ni visualizadores iguales, tampoco organizan los datos de manera normalizada. Es por ello que, los servicios que incluyen servicios digitales varía considerablemente; mientras que algunos permiten el acceso directo a un título, y la búsqueda de los artículos en cada fascículo, semejando a los formatos impresos, en otros casos la información se organiza en base de datos con múltiples accesos en páginas web que tiene poco parecido con la revista en papel.

Por lo anterior, la Dirección General de Bibliotecas (DGB), agrupa mas de quince mil títulos de revistas electrónicas en un catálogo que facilita el acceso directo a las mismas. En dicho catalogo el lector puede buscar un titulo, un tema, un editor o una palabra del titulo, cuando no recuerda la palabra inicial del mismo.

Entre las revistas que se encuentran en el proceso de instalación están los títulos de EBSCO, uno de los más grandes editores del mundo.

Ejemplo de revistas electrónicas de consulta directa:

- ABSTRACTS-PERIODICAL OF THE AMERICAN PHYSIOLOGICAL SOCIETY
- AMERICAN JOURNAL OF PHYSIOLOGY, CELL PHYSIOLOGY
- BIOCHEMICAL AND MOLECULAR MEDICINE
- JOURNAL OF INORGANIC BIOCHEMISTRY
- JOURNAL OF PHARMACEUTICAL AND BIOMEDICAL ANALYSIS
- SCIENCE
- NUCLEAR PHYSICS B
- WEED RESEARCH
- WATER SCIENCE AND TECHNOLOGY

### TÍTULOS DE REVISTAS ELECTRÓNICAS DE CONSULTA INDIRECTA

La información que se puede obtener en este tipo de servicios es muy variada; entre los puntos de acceso se encuentra el autor del artículo, el o los temas, la fecha de publicación, desde luego el título de la revista, el título del artículo y otros.

Al recuperar información pertinente en la base de datos, el lector podrá encontrar la opción de visualizar el texto completo e imprimirlo, grabarlo o enviarlo a un correo electrónico.

A esta información se le conoce como consulta indirecta, porque no presenta los datos de manera semejante a la que contiene la revista impresa.

Ejemplo de revistas electrónicas de consulta indirecta:

ABA JOURNAL  
ADVOCATE  
AMERICA'S NETWORK  
HUMANIST  
HECATE  
MANUFACTURING CHEMIST  
MEDICAL TEACHER  
SOCIOLOGY  
STUDENT BMJ

Los nombres y direcciones de las páginas llegan continuamente a manos de la Profesora Imelda Velázquez Montes por medio de ligas

mundiales en Internet o comunidades virtuales que se dedican a la divulgación de información científica, donde a través de un servidor se manda la información más actualizada acerca de los temas en los cuales está interesado el usuario, que en todos los casos es información científica, en especial relacionada con la química; con estos datos, se verifica y seleccionan los sitios que contienen información útil y específica, ya que al hacer la selección se valúa el contenido de éstas para saber en qué temas se relacionan para así poder colocarlos en los sitios específicos, para que el usuario los pueda buscar posteriormente.

Es por eso que el trabajo que realiza el departamento de Servicios de Información Digital (SID), ayuda de manera titánica en la búsqueda y recopilación de información, proporcionando los medios para facilitar la obtención de información actualizada y veraz, a todo el usuario que lo solicite.

A partir del presente año SciFinder Scholar 2007 ofrece nuevas y poderosas herramientas:

- **Guardar Respuestas**— Ahora usted puede guardar listas de referencias, sustancias, o reacciones y re-utilizarlos mas tarde en otra sesión de SciFinder Scholar.
- **Combinar conjuntos de Respuestas**—La herramienta de combinar hace posible enfocarse en un tópico de investigación particular de una nueva forma, combinando sustancias, reacciones y referencias en un conjunto de respuestas activo con aquellos que hayan sido guardados previamente. Empleando varias opciones de combinación, intersección, o remoción, usted puede manipular y personalizar la información contenida en esos conjuntos combinados de respuestas.
- **Exportar fichas comerciales desde CHEMCATS® en Excel**— Crear reportes listando sus fuentes comerciales favoritas, incluyendo información de contactos y precios para sustancias de interés. La información del Catálogo se importa a hojas para su clasificación, manipulación y referenciación.
- **Explorar desde Desplegado de sustancias**—Llegue a ser más productivo usando el desplegado de una sustancia, para iniciar una nueva búsqueda o refinar una búsqueda activa.

- **Imprima Estructuras de sustancias en formato de tablas**—Obtenga una vista comparativa única de múltiples estructuras químicas, imprimiéndolas una junto a otra en este útil arreglo en forma de tabla. Compatible con Windows Vista™ and Mac OS X
- **Categorizar Referencias**—Esta herramienta le permite ordenar y evaluar información en su conjunto de respuestas, creando un histograma de fácil análisis. Categorizar separa referencias relacionadas por sus términos asociados por el Índice de CA. Con Categorizar usted puede re-armar un subconjunto de una forma única creando oportunidades para nuevas perspectivas

## DISCUSIÓN

En cuanto a la terminología en el ámbito químico, al utilizar el mencionado Sistema SciFinder Scholar, se pueden usar las herramientas descritas durante la búsqueda de un compuesto específico; incluyendo el sistema de nomenclatura aprobado a nivel mundial, que incluye un número de registro llamado CAS Registry Number para cada sustancia, elemento o compuesto, que además de confiable puede eliminar equivocaciones o errores que se pueden cometer al utilizar la nomenclatura o un idioma diferente al del origen del documento, así como otro tipo de términos científicos incluidos en los documentos y que son clave para su identificación.

La razón por la cual es indispensable que la búsqueda y selección de información la realice un profesional de la química, de preferencia con posgrado es porque, cuando no se cuenta con el número CAS, ni la fórmula exacta, ni la estructura del compuesto; se hace necesario conocer las propiedades, las características químicas o bien su comportamiento en las condiciones de reacción, de manera que se garantice el éxito en la consulta a la base de datos adecuada.

Para el caso específico de localización del tema pigmentos o colorantes, fue posible hacer el siguiente comparativo:

El criterio de selección e los vínculos en buscadores y meta buscadores fue en función del contenido de la página vinculada; éste tenía que poseer al menos una de las siguientes características:

*Primer Simposio Internacional sobre Organización...*

- Información de interés técnico
- Información de interés divulgativo
- Información de interés científico
- Información respecto al contexto cultural de los pigmentos

Excluyendo:

- Publicidad
- Textos en idiomas diferentes al español y al inglés
- Información de pigmentos de interés o aplicación estrictamente en procesos biológicos.

La información disponible en Chemical Abstracts fue seleccionada por su contenido:

- Información de interés técnico
- Información de interés científico

Excluyendo:

- Información en idiomas diferentes al español y al inglés
- Información de pigmentos de interés o aplicación estrictamente en procesos biológicos

Todos los artículos o patentes encontrados a través del Chemical Abstracts fueron de interés científico, tecnológico o ambos; se referían a tecnologías o investigaciones de punta al respecto del tema de pigmentos.

De estos datos, se puede concluir que con fines de recavar información divulgativa e introductoria para los temas se vuelve conveniente la búsqueda en fuentes de acceso libre, siendo la más efectiva Copernic; que si bien no arrojó muchos resultados, si brindó información de mejor calidad que Google.

Mientras que la búsqueda a través del Chemical Abstracts brinda información mucho más específica, actual e interesante al respecto de temas científicos y tecnológicos. Sin embargo, la única desventaja



que se puede encontrar en el Chemical Abstracts es que no fue posible recavar toda la información de interés en texto completo que se encontró debido a que está escrita en idiomas como el ruso, japonés, chino y coreano. Lo que me hace pensar que los adelantos realizados en el área de pigmentos en el mundo fueron principalmente desarrollados en países asiáticos; sin embargo, se localizaron todos los resúmenes o abstracts de los trabajos publicados durante 100 años, utilizando la terminología apropiada.

## BIBLIOGRAFÍA:

*Servicios de información digital para la comunidad universitaria*. 1999 Universidad Nacional Autónoma De México. Dirección General de Bibliotecas. <http://www.unam.mx>

Imelda Velázquez-Montes, "Internet para todo y para todos" Antología 1, Ciclo de Conferencias La Ciencia más allá del Aula. FQUNAM. 137-150, (2006). 3. Portal de Internet [en línea] [lugar de publicación México, D: F.] [31/05/06 14:20 PM]. Disponible en World Wide Web: ([www.mtas.es/ins-ht/ntp\\_276.htm](http://www.mtas.es/ins-ht/ntp_276.htm))

Answer extraction and ranking strategies for definitional question answering using linguistic features and definition terminology. By: Han, Kyoung-Soo; Song, Young-In; Kim, Sang-Bum; Rim, Hae-Chang. *Information Processing & Management*, Mar2007, Vol. 43 Issue 2, p353-364, 12p; DOI: 10.1016/j.ipm.2006.07.010; (AN 22946071)

Archival Terminology; or, What Did You Just Call Me? By: Sandleback, Nancy. *MAC Newsletter*, Jan2007, Vol. 34 Issue 3, p39-39, 2/3p; (AN 23799323)

A Glossary of Archival and Records Terminology. By: Robertson, Anne. *Archives & Manuscripts*, Nov2006, Vol. 34 Issue 2, p203-204, 2p; (AN 23583760)



# La primera banda del contínuum de metadatos, un nivel poco gestionado por los portales bibliotecológicos y las bibliotecas digitales

ALEJANDRO URIBE TIRADO  
*Escuela Interamericana de Bibliotecología  
Universidad de Antioquia, Colombia*

## INTRODUCCIÓN

Este trabajo presenta un análisis realizado a diferentes portales (o sitios web) de escuelas de bibliotecología y ciencias de la información, bibliotecas universitarias y públicas de iberoamérica, y a algunas bibliotecas digitales (o directorios de bibliotecas digitales y proyectos *open source*) representativas en esta área del conocimiento, en el cual se detecta que la gestión de metadatos en su primera banda (*Dempsey y Heery*) no han sido suficientemente gestionada por los diseñadores y administradores de estos recursos, lo cual genera una problemática para la indización y recuperación de información que realizan los buscadores, como los principales servicios web de localización de información que utilizan los usuarios de todo el mundo, además de ser un problema de coherencia en la gestión de información, en la gestión de contenidos, por parte de los profesionales e instituciones que deberían gestionar con más rigor las diferentes bandas de este *contínuum de metadatos* necesarias para una más estructural y semántica búsqueda y recuperación de información digital.

El análisis que se describe, es resultado de diferentes procesos y espacios docentes e investigativos desarrollados los últimos

dos años desde la Escuela Interamericana de Bibliotecología de la Universidad de Antioquia, a partir de sus grupos de investigación de Gestión de Conocimiento, Tecnologías de Información y Terminología, y concretamente, de prácticas docentes a partir del “Seminario-Taller Arquitectura de la información y estructuración de metadatos para unidades de información” (<http://aprendeonline.udea.edu.co/lms/moodle/course/view.php?id=30>), “Seminario de Portales” (<http://aprendeonline.udea.edu.co/lms/moodle/course/view.php?id=87>) y de la asesoría como parte de la investigación-extensión universitaria desde esta escuela de bibliotecología, al proyecto Portal U. de A. (<http://www.udea.edu.co/portal.exe>)

## LOS METADATOS Y LA ORGANIZACIÓN DE LA INFORMACIÓN DIGITAL - EL CONTÍNUUM DE METADATOS

Antes de iniciar la presentación de los resultados de este estudio aplicado, es necesario identificar qué se entiende por metadato, su relación con la organización de la información digital, y cómo esta organización de la información digital que pueden proveer los metadatos (además de las ontologías, los tesauros y otros desarrollos) tiene diferentes niveles (*El contínuum de metadatos*), que desde la bibliotecología y de parte de los bibliotecólogos debe atenderse, especialmente en: los portales (o sitios web) de las escuelas y facultades de bibliotecología y ciencias de la información, los portales (o sitios web) de las bibliotecas universitarias y públicas (y demás tipologías de unidades de información), y los sitios web de proyectos y recursos digitales (bibliotecas digitales, directorios, proyectos *open source*) claves para esta área del conocimiento; pues son estos recursos los primeros lugares representativos de este *quehacer* buscando una mejor organización, recuperación y almacenamiento de la información digital.

Respecto al concepto metadato se acoge la propuesta de definición elaborada por José A. Senso y Antonio de la Rosa Piñero tras analizar diferentes concepciones sobre el mismo:

*Metadato es toda aquella información descriptiva sobre el contexto, calidad, condición o características de un recurso, dato*

*u objeto que tiene la finalidad de facilitar su recuperación, autenticación, evaluación, preservación o interoperatividad.*

Esta definición implica dos elementos claves, por un lado que los metadatos son una herramienta útil no solo para la *recuperación* de información sino para la *autenticación, evaluación, preservación o interoperatividad*; y que éstos pueden aplicarse según las necesidades, intereses e importancia informativa, a un recurso (página web, portal, directorio, biblioteca digital, etc.), a un dato u objeto (documento, imagen, archivo, etc.).

En general, podríamos decir que algunas de las características principales que definen la importancia de los metadatos respecto a la organización de la información, considerando las particularidades de la sociedad de hoy y el universo de información digital, son:

- Incrementa la accesibilidad: es posible consultar con una única opción de búsqueda
- Disminución del tráfico en la red: se indiza la representación del recurso, objeto o dato y no dicho recurso, objeto o dato en sí
- Expandir el uso de la información: facilita la difusión de versiones digitales de un único objeto
- Control de versiones de esa descripción
- Aspectos legales: permiten establecer claramente las restricciones de explotación, informar sobre los derechos de autor, control del uso de todo, o una parte, del objeto, método de pago por su disfrute, controlar el acceso a información restringida
- Preservación del objeto original: la utilización de metadatos junto al uso de lenguajes controlados permitiría aumentar la precisión en la mayoría de búsquedas en Internet
- Participación y toma de conciencia de la responsabilidad de hacer frente a esta sobrecarga de información, a esta desorganización por parte de todos los responsables de los documentos electrónicos: Editores, Servicios de Información, Proveedores, Bibliotecas, Autores, Servicios de búsqueda de Internet, Servicios de Archivos electrónicos, Depósitos (repositorios) de colecciones de documentos, y bibliotecas digitales, entre otros

Los anteriores elementos claves y características nos conducen, cuando nos acercamos a la práctica de aplicar los metadatos tanto a recursos, como a datos y objetos, a entender que esta aplicación tiene diferentes niveles y desarrollos. Para entender esos niveles y desarrollos es necesario recurrir a la propuesta hecha por *Dempsey y Heery* sobre el *continuum de metadatos* y el *sistema clasificatorio del proyecto Biblink* citados por Eva M<sup>a</sup>. Méndez Rodríguez:

**Cuadro 1: Tipos de metadatos en bandas (basado en Dempsey y Heery)**

	Banda 1		Banda 2		Banda 3	
<b>Características</b>	Formatos simples Sistemas propietarios Indización <i>full-text</i>		Formatos estructurados Estándares de facto Estructura de campos		Formatos ricos Estándares internac, Etiquetas elaboradas	
<b>Formatos</b>	Lycos Alta Vista Yahoo etc.	<b>Etiquetas</b> <META>	DC/DCMI IAFA RFC 1807 SOIF LDIF	<b>EdNa</b> <b>AGLS</b> <b>etc.</b>	ICPSR CIMI LAD TEI MARC	<b>DC-AP</b> <b>DIG3 5</b> <b>IMS</b> <b>etc.</b>

**Cuadro 2: Sistema clasificatorio el en proyecto Biblink**

	Sencillo	-----	----- ► Complejo
Localización	Selección	Evaluación	Análisis
Generados por un robot	Robot con ayuda del hombre	Introducción manual	Introducción y análisis de contenido manual
No estructurados	Valores acompañados de atributos	Cualificadores	Lenguaje de etiquetado estructurado (SGML)
Interface que conecta el formulario con el servicio http mediante CGI	Servicio de directorio (whois++) con enrutamiento mediante CIP (Common Indexing Protocol)	Z39.50	Nuevas versiones de Z39.50 (con SQL, texto completo, browsing...)
Propietario	Apareciendo en la actualidad	Normas genéricas utilizadas actualmente en el mundo bibliotecario	Normas utilizadas en temáticas muy especializadas

Esta tipología y clasificación conlleva a que los recursos, datos y objetos digitales pueden organizarse con determinados niveles y desarrollos que implican grados distintos en las posibilidades de *recuperación, autenticación, evaluación, preservación o interoperatividad*, en las que según el nivel y desarrollo, serán más automatizadas o requerirán mayor intervención del bibliotecólogo (o del editor o del autor pensando en la viabilidad de una web semántica<sup>1</sup>), tendrán estándares, normas y leguajes más sencillos o más complejos y estructurados.

En los últimos años los bibliotecólogos y otros profesionales de la información y profesionales de áreas relacionadas, como los ingenieros de sistemas, han liderado la implementación de recursos, que basados en metadatos, provean de mejor manera el acceso a la información digital (sea general, especializada o temática), mediante la proliferación de Portales y Bibliotecas Digitales, o de recursos semejantes: *repositorios/contenidos + servicios informativos/bibliotecológicos*.

Estas propuestas han potenciado de manera significativa la organización de la información digital, aunque aún falta muchísimo, pero son los primeros pasos necesarios.

No obstante, estas propuestas han descuidado o no valorado suficientemente la importancia en la organización de la información digital de considerar todas las bandas del *continuum de metadatos*, pues se han centrado principalmente en la utilización de esquemas de metadatos para la organización de los datos y objetos (contenidos) que constituyen esos portales y bibliotecas digitales, pero descuidando el primer nivel, el acceso a estos portales (o sitios web) y bibliotecas di-

---

1 Los bibliotecólogos frente a la gestión de metadatos buscando una web semántica, tienen un compromiso fuerte con la formación de usuarios, con la capacitación a otros profesionales de la información, incluso de los autores mismos de la información digital, pues esta web de acceso universal, esta web más útil, con sus diferentes aplicaciones particulares en bibliotecas y organizaciones, requerirá que todos aportemos y nos comprometamos para su organización, como lo indica Eva M<sup>a</sup>. Méndez Rodríguez: *A pesar de que existen esquemas e incluso normas formales para la descripción de recursos de Internet, su aplicabilidad exige persuadir a los creadores de información Web para que utilicen dichos esquemas, así como a los desarrolladores de aplicaciones para que asimilen tales normas en sus productos. Ibíd.*

giales mediante una(s) página(s) web que son en definitiva las que los buscadores, por medio de sus robot, indizan permanentemente. Buscadores que son el principal servicio de recuperación de información que utilizan los usuarios alrededor del mundo.<sup>2</sup>

Utilizando una metáfora, podría decirse, que los profesionales de la información se han preocupado por organizar todo el interior de la casa pero han descuidado la fachada, la portada, el pórtico de la misma.

Este trabajo presenta a continuación una descripción de esta problemática, con el análisis de una muestra representativa de estos recursos (sitios web, portales, directorios, bibliotecas digitales, proyectos *open source*), que evidencian los efectos negativos que conlleva el descuidar ese primer nivel.

## METODOLOGÍA

Para la realización de este análisis, se ha recurrido a los siguientes recursos, en los cuales se ubican sitios web y portales de escuelas y facultades de bibliotecología y ciencias de la información, de bibliotecas universitarias y públicas, y de bibliotecas digitales y demás recursos relacionados:

- *EDIBCIC - Asociación de Educación e Investigación en Bibliotecología, Archivología, Ciencia de la Información y Documentación de Iberoamérica y el Caribe* (<http://www.edibcic.org/Escuelas/Escuelas.htm>)
- *Libweb - Library WWW Servers* ([http://lists.webjunction.org/libweb/CSA\\_main.html](http://lists.webjunction.org/libweb/CSA_main.html))

---

2 Según *Target Marketing Magazine*, las mejores vías para encontrar websites son las siguientes: Anuncios por banners: 1%; Mailings por correo electrónico: 1,2%; Anuncios por televisión: 2,1%; Por casualidad: 2,1%; Anuncios en prensa: 4,4%; Boca a oreja: 20%; Navegación fortuita: 20%; Buscadores de internet: 46%. Según *OneStat* "más del 90% de los internautas utilizan buscadores... más del 50% de las búsquedas hechas en Internet son realizadas con frases de 3 y 4 palabras". Según *Jupiter Research* "los usuarios de Internet situaron la búsqueda como la actividad más importante con 9,1 sobre 10". <http://www.promocionesuweb.com/> [Consultado: 8-07-2007]



- *Infoscience META-OCIB Metasearch Open Acces en Ciencia de la Información y Bibliotecología* (<http://hjr227.googlepages.com/home>)
- *Bibliotecas digitales: directorios de recursos*

La elección de estos recursos, parte de la afirmación que el primer lugar donde los profesionales de la información deben *dar ejemplo* sobre la importancia de los metadatos para la organización de la información es en los propios recursos que ellos usan con más frecuencia en su campo del conocimiento, sea para actividades de docencia, de formación en pregrado o posgrado, de actualización o de desempeño profesional. Es decir, son los recursos donde con mayor claridad hay que evitar, por nuestra propia coherencia, el refrán que dice: “En casa de herrero azadón de palo”.

En dichos recursos, la elección de los sitios o portales se hizo considerando las posibilidades de alcance de este trabajo y la trayectoria bibliotecológica de los países, por lo cual se eligieron para el caso de los sitios web y portales de escuelas y facultades, y de bibliotecas universitarias y públicas, los siguientes: Argentina, Brasil, Colombia, Costa Rica, Chile, Cuba, España, México, Perú y Venezuela.<sup>3</sup>

Tras la elección de estos recursos, se pasó a la utilización de un formato que facilitara la recopilación la información de cada recurso, teniendo en cuenta las etiquetas de metadatos que aparecen en la cabecera (<head>) de las páginas html-htm de esos recursos (*Home page* y páginas del segundo nivel de la arquitectura del recurso, cuando era necesario) tomando como base los principales criterios que definen las metaetiquetas básicas de html-htm (title, keyword, description, http-equiv, autor, generador,) y su interrelación con el estándar Du-

---

3 En este texto se omiten intencionalmente los nombres propios de los recursos analizados, pues se asume que los lectores de este trabajo, harán en su momento, el autodiagnóstico del sitio web o portal de su institución para identificar su estado, en cuando a la gestión de la primera banda del continuum de metadatos, para luego hacer la mejoras respectivas. No obstante, más adelante, para presentar casos de una adecuada gestión en las tres tipologías de recursos que analizamos, se presentan los metadatos de esos recursos, lo cual permite identificar la institución que los ha gestionado en su sitio web o portal.

blín Core (dc.title dc.subject, dc.description, dc.identifier, d.c.creator, dc.date, etc.).

Para esta recopilación, se utilizó diferentes servicios de acceso libre: análisis de metaetiquetas ([http://www.searchengineoptimising.com/metatag\\_analyser/index.htm](http://www.searchengineoptimising.com/metatag_analyser/index.htm)), metadatos DC (<http://www.ukoln.ac.uk/metadata/dcdot/>), y el software Metabrowser, cuyos resultados se validaron con la visualización del código fuente de dichos recursos y de su presentación descriptiva en *Google* y *Yahoo* como los principales recursos de búsqueda de información utilizados por los usuarios de Internet, lo cual se reafirma en cada nuevo estudio.<sup>4</sup>

## ANÁLISIS

Entre las escuelas y facultades de bibliotecología y ciencias de la información, se analizaron en total 80 *Home page* de los sitios web o portales de las instituciones de esta categoría, ubicadas en los siguientes países, y distribuidas así:

País	No.
Argentina:	7
Brasil:	36
Chile:	2
Colombia:	4
Costa Rica	2
Cuba:	1
España	18
México:	6
Perú:	2
Venezuela:	2

---

<sup>4</sup> “Según un artículo publicado el 16 de enero de 2007 en *EarthTimes.org*, Google Inc. ha incrementado su cuota en el mercado de buscadores de los EEUU en un 0,4%, alcanzando en diciembre el 47,4%. A su vez, Yahoo! ha ganado un 0,3% y mantiene la segunda posición con un 28,5%, según los datos del análisis de búsquedas mensual realizado por la empresa comScore Networks”.

Tras el análisis se detectaron los siguientes resultados en cuanto al uso de metadatos, tanto en un primer nivel usando solo los propios de html-htm, o pasando a un segundo nivel utilizando un esquema como el Dublín Core, o combinándolos:

### Sitios y Portales de escuelas y facultades de Bibliotecología y C.Inf. que utilizan la metaetiqueta...

Título	Palabras clave	Descripción	Otras Metaetiquetas no generadas automáticamente	Esquema Dublín Core	Solo metaetiquetas generadas automáticamente
68	17	15	16	2	10

Como evidencia la tabla, de un total de 80 sitios web o portales solo un total de 17 (21.2%) tendrían un manejo “relativamente adecuado” de los metadatos de esta primera banda,<sup>5</sup> 15 de ellas usando las metaetiquetas básicas de html-htm (title, description, keyword, http-equiv, autor, generador), y 2 incluyendo los elementos del estándar Dublín Core (2.5%). Muy preocupante resultó encontrar que 10 de estos sitios o portales (12.5%) no hacían ninguna gestión y que solo aparecían metaetiquetas automáticas generadas por el programa de software en que se diseñó o producto del servidor o plataforma que lo soportan (especialmente gestores de contenido como Mambo y Joomla), a su vez, que en 51 de estos sitios y portales (63.8%), solo se utiliza la metaetiqueta *título*.

En el caso de las 130 bibliotecas universitarias y públicas analizadas, a partir del directorio *Libweb*, según la siguiente distribución por países:

---

5 Se indica que “relativamente adecuado”, pues el alcance de este estudio no entró a evaluar la calidad terminológica de los conceptos usados para la descripción de cada una de estas metaetiquetas, sino su presencia para una más controlada descripción del recurso para efectos de la indización en buscadores y su más fácil y clara recuperación, almacenamiento, etc.

País	No.
Argentina:	12
Brasil:	12
Chile:	11
Colombia:	12
Costa Rica	1
Cuba:	2
España	42
México:	25
Perú:	10
Venezuela:	3

se obtuvieron los siguientes resultados:

### Sitios y Portales de bibliotecas universitarias y públicas que utilizan la metaetiqueta...

Título	Palabras clave	Descripción	Otras Metaetiquetas no generadas automáticamente	Uso Esquema Dublín Core	Solo metadatos generados automáticamente
117	34	30	14	10	3

Estos resultados implican que solo 24 de estos sitios o portales (18.5%) hacen una gestión mínima de estos metadatos en la primera banda, 14 de ellas incluyendo las metaetiquetas básicas de html-htm, y 10 utilizando los metadatos del estándar Dublín Core (7.7%). En este caso, el número de bibliotecas que no hace ninguna gestión (un total de 3, equivalentes al 2.3%) se disminuye ostensiblemente, comparado con los sitios web o portales de las escuelas y facultades, aunque, se sostiene el mismo porcentaje de los sitios o portales de estas bibliotecas que solo usan la metaetiqueta título: un total de 83 (63.8%)

Entre los 19 sitios de acceso abierto, directorios de bibliotecas digitales y demás, directamente relacionados con la bibliotecología y la ciencia de la información, tras el análisis se encontró:

## Sitios de acceso abierto, directorios de bibliotecas digitales y demás recursos del área bibliotecológica que utilizan la metaetiqueta...

Título	Palabras clave	Descripción	Otras Metaetiquetas no generadas automáticamente	Uso Esquema Dublín Core	Solo metadatos generados automáticamente
15	7	7	2	3	1

En este caso, 5 de estos sitios (26.3%) hacen una gestión mínima de estos metadatos en la primera banda, 2 de ellos incluyendo las metaetiquetas básicas de html-htm, y 3 utilizando los metadatos bajo el estándar Dublín Core (15.8%). En este caso, el número de bibliotecas que no hace ninguna gestión es solo de 1 (5.3%), no obstante, ante estos mejores porcentajes en los aspectos anteriores, paradójicamente se presenta que en estos sitios, 13 de ellos solo hacen uso de la metaetiqueta título, lo que equivale a un mayor número porcentualmente (68.4%).

Entre todos estos sitios, portales, directorios y demás (un total de 229 considerados en este estudio), tras analizar la descripción de los mismos en Google y en Yahoo, de acuerdo al link que vincula su *Home page*, se puede decir que solo una tercera de éstos es adecuadamente descrito, es decir, la descripción que presenta el buscador indica lo que el sitio o portal representa (contenidos) y los servicios que incluye, y esa descripción, es producto de los metadatos que los creadores o editores del mismo han determinado que desean que los buscadores les indiquen, no de la lectura indiscriminada del código, lo cual reitera la importancia de no descuidar esta gestión en esta primera banda.<sup>6</sup>

---

<sup>6</sup> Independiente de que recursos informativos, como es el caso de Google y Yahoo, estén mejorando sus procesos de descripción cuando encuentran muy “pobres” o “desconfían” de las metaetiquetas, tras analizar aspectos claves del texto, del contenido completo; esta descripción al no ser controlados directamente por los editores y/o autores de la información, por los bibliotecólogos, generan descripciones erradas o incompletas, además, dichos procesos de descripción no son

## Primer Simposio Internacional sobre Organización...

Algunos ejemplos de esta última situación, de descripción no gestionada directamente, serían los siguientes casos, tanto de sitios web o portales de escuelas y facultades de bibliotecología, de bibliotecas universitarias y públicas, y de recursos de consulta del área bibliotecológica:

Facultad de...  
Cronograma 2007. Revista Alfilo. Campus Virtual. Webmail. Boletín Virtual.  
Sistema Guaraní ...  
www. . . . /index2.php - 5k

NUEVO CONSULTA BASES DE DATOS FUERA DEL CAMPUS UNIVERSITARIO ...  
METABUSCADOR DE RECURSOS BIBLIOGRÁFICOS - NUEVO. CONSULTA BASES  
DE DATOS FUERA DEL CAMPUS UNIVERSITARIO · Horario Carrera 7  
No. 41 - 00 Conm. ...  
www. . . .edu. /biblos/bibliote.html - 13k

| Information Service  
Subject trees arranged alphabetically and by UDC.  
. . . .uk - 6k

## CONCLUSIONES Y RECOMENDACIONES

Los resultados de este análisis confirman que a pesar de esperarse una mejor gestión de metadatos en estos recursos, en todas las *bandas del continuum*, al estar directamente relacionados con el *quehacer* de

---

homogéneos entre tales tipos de recursos para búsqueda de información lo cual genera representaciones diferentes del mismo recurso en Google y en Yahoo, por ejemplo:

Visualización Google	Facultad de Filosofía y Letras - Programa de la Facultad de Filosofía y Letras de la Universidad de . . . . Presentación del programa, plan de estudios, cátedras y departamentos. www. . . . / 2k
Visualización Yahoo	Universidad - Facultad de Filosofía y Letras - Antropología, artes, bibliotecología, educación, historia, filosofía, geografía, y letras. www. . . . - 73k

los bibliotecólogos, las escuelas y facultades de bibliotecología y ciencias de la información, las bibliotecas y diferentes proyectos de acceso a la información digital en esta área del conocimiento, esto aún no se da en una gran mayoría de casos en la primera banda:

- En los sitios web y portales de las escuelas y facultades de bibliotecología analizadas, un 78.8% no utilizaban metadatos básicos para la recuperación de información (sean metaetiquetas html-htm y/o estandar Dublín Core).
- En los sitios web y portales de las bibliotecas universitarias y públicas analizadas, un total de 81.5% no hacen esa utilización
- En las bibliotecas digitales, recursos *open source*, directorios del área y demás, esa no utilización alcanza un 73.7%

Estos resultados, confirman también la tendencia presentada en otro estudio realizado por Cristófol Rovira y Mari-Carmen Marcos sobre “Metadatos en revistas-e de Documentación de libre acceso” en los cuales se encontró que a pesar de partir de la hipótesis que: “Si los profesionales de la Documentación [*Bibliotecología, Ciencias de la Información*] son quienes más interés demuestran en que las páginas web cuenten con metadatos, las revistas digitales [*en el caso de este análisis: los sitios web o portales de escuelas y facultades de bibliotecología y ciencias de la información, de bibliotecas universitarias y públicas, y de proyectos de recursos de consulta del área bibliotecológica*] que sirven de base a su disciplina usarán mayoritariamente metadatos y lo harán correctamente”. Desafortunadamente esa hipótesis, como ellos lo afirman, no es cierta. En ese estudio los resultados indican que “el 59% de esos sitios de revistas del área no utilizaban metadatos útiles para la recuperación de información”.

Todo lo anterior lleva a la conclusión que es necesario, de parte de todos los profesionales que son responsables de los sitios web, portales, bibliotecas digitales, directorios y demás recursos de información publicados en la web; tomar conciencia de la importancia de una adecuada gestión de metadatos desde su primera banda, no dejar este proceso a que los recursos informáticos lo hagan automáticamente o esté a cargo de diseñadores o ingenieros de sistemas que generalmente no son concientes o saben de la importancia de esta gestión.

Es decir, no solo preocuparse por la gestión de metadatos desde la base de datos que vincula o soporta la información que suministran, sino por la gestión en el nivel que permite a los usuarios llegar a los recursos, principalmente por los buscadores; o en el caso en que estos usuarios ya conocieran el recurso y la utilidad de los buscadores no fuera “tan necesaria”, esta gestión en dicha primera banda sigue siendo inmensamente necesaria, no solo por coherencia disciplinar y profesional, sino por aportar al objetivo de lograr una web más semántica: “los profesionales de la documentación deberíamos ser los primeros en cumplir las recomendaciones que se dan desde nuestra área de conocimiento”.

Lograr esa adecuada gestión, no es un objetivo muy difícil de lograr, algunos casos demuestran que es posible hacerlo, e irlo mejorando, integrando las metaetiquetas html-htm y el estándar Dublin Core (u otros estándares) por ejemplo, y así responder positivamente a la tendencia de organización de la información digital de años atrás (Web 1.0), a la actual (Inicios de la Web 2.0) y a la que esperamos que en el futuro se impondrá (Web 3.0 y Web 4.0):

```
<!-- METADATOS -->
<meta name="title" content="Universidad Complutense de Madrid :: Página principal">
<meta name="keywords" content="educación, universidad, estudios, carreras, cultura, universitarios, madrid, españa, ciencia, tecnología, arte, historia, docencia, investigacion, preinscripcion, matricula, doctorado, licenciatura, licenciado, diplomatura">
<meta name="description" content="Facultad de Ciencias de la Documentación de la Universidad Complutense de Madrid">
<meta name="author" content="Universidad Complutense de Madrid">
<meta name="distribution" content="all">
<meta name="resource-type" content="Homepage">
<meta name="expires" content="Tue, 01 Jun 1999 19:58:02 GMT">
<meta name="revisit" content="7 days">
<meta name="robots" content="all">
<meta name="DC.Title" content="Universidad Complutense de Madrid :: Página principal">
```



```
<meta name="DC.Creator" content="Universidad Complutense
de Madrid">
<meta name="DC.Language" scheme="RFC1766" content="Spanish">
<meta name="VW96.objecttype" content="Homepage">
<meta http-equiv="title" content="Universidad Complutense de Madrid ::
Página principal">
<meta http-equiv="keywords" content="educación, universidad, estudios,
carreras, cultura, universitarios, madrid, españa, ciencia, tecnología, arte,
historia, docencia, investigacion, preinscripcion, matricula, doctorado,
licenciatura, licenciado, diplomatura">
<meta http-equiv="description" content="Facultad de Ciencias de la
Documentación de la Universidad Complutense de Madrid">
<meta http-equiv="DC.Description" content="Facultad de Ciencias de la
Documentación de la Universidad Complutense de Madrid">
<meta http-equiv="Content-Type" content="text/html; ISO-8859-1">
<meta http-equiv="Pragma" content="no-cache">
<meta http-equiv="cache-control" content="no-cache">
Title: Biblioteca Universidad Arturo Prat, Iquique - Chile
X-Meta-Author: Héctor Tapia Flores
X-Meta-DateCreated: Thu, 04 July 2002 12:00:54 GMT
X-Meta-DC.Date: 2004-06-04
X-Meta-DC.Date.Availableated: 2000
X-Meta-DC.Date.Created: 2000
X-Meta-DC.Date.Issued&: 2001
X-Meta-DC.Date.Modified: 2007
X-Meta-DC.Date.Valid: 2007
X-Meta-DC.Description: Portal de la Biblioteca Universidad Arturo Prat,
Proyecto MECESUP UAP/9901 - Iquique - Chile
X-Meta-DC.Description.Abstract: En 1997, el Gobierno de Chile definió
una nueva política de desarrollo de la educación superior, creando el
Programa de Mejoramiento de la Calidad y la Equidad de la Educación
Superior (MECESUP), que apunta a complementar el proceso de reforma
de la educación iniciado en la última década.
X-Meta-DC.Description.TableOfContents: Dirección, Secciones de
Circulación, Hemeroteca y Referencia, Biblioteca Digital, Procesos
Técnicos, Atención de Usuarios y Area de Comunicaciones y Marketing
```

Integral

X-Meta-DC.Format: Application/http

X-Meta-DC.Format: Application/pdf

X-Meta-DC.Format: Application/rft

X-Meta-DC.Format: Application/sgml

X-Meta-DC.Format: Application/xml

X-Meta-DC.Format: Message/http

X-Meta-DC.Identifier: <http://biblio.unap.cl/>

X-Meta-DC.Language: es

X-Meta-DC.Relation.HasPart: Universidad Arturo Prat

X-Meta-DC.Source: Biblioteca Universidad Arturo Prat

X-Meta-DC.Subject: Biblioteca, Library, Iquique, Chile, Sudamerica, Tarapaca, Educación, Alto Hospicio, Universidad, Universidad Arturo Prat, Bases de Datos, Publicaciones, Papers, Artículos, Abstract, Tesis, Tesis Doctorales, Magister, Pregrado, Docencia, Educare, Eduxere, Hemeroteca, Referencia, Biblioteca Abierta, Nucleo Central de Información Regional, Mineduc, CNAP, Gobierno de Chile, Revistas, CABID, Bibliotecologos, Archivística, CIRE, INE, UNESCO, OEA, Acreditación, MECESUP UAP/9901, Catalogo Werken, Universidad del Bio-Bio, ISIWEB of SCIENCE, PROQUEST, UNIVERSIA, eLIBRO, eBRARY, EBSCO, OCEANET, Microjuris, BUSINESS MONITOR ONLINE, DIRECTORY OF OPEN ACCESS JOURNALS, OAIster, Scielo, Arturo Prat, Cavancha, Huayquique, Campus, Sedes, Carreras, Técnicos, Liceos, Colegios, Escuelas, Memorias de Titulo, Digitalización, Carlos Merino Pinochet, Consejo de Rectores, PSU, PAA, CEMDIS, INTE, Coordunap, Canchones, Baquedano, W3c, XML, WAI, MITT, Wikipedia, INIA, IIPM, Mathematical, Chemistry, Biology, Education, Repositorio, Informes, Investigaciones, XML, RSS, WAP, Tareas Escolares, Portal Educativo

X-Meta-DC.Title: Biblioteca Universidad Arturo Prat, Iquique, Chile

X-Meta-DC.Type: Library

X-Meta-DC.Type: Collection

X-Meta-DC.Type: Interactive Resource

X-Meta-DC.Type: Service

X-Meta-Revisit-After: 1 days

X-Meta-Revisit-After: 1 days

X-Meta-Robots: index, follow  
X-Meta-Robots: all | index | follow  
X-Meta-Verify-V1:  
mbd5QjETQYnZ8oCTvLKR4PxxgjcyaUN99ny+8j29x5w=

Title: DoIS : Documents in Information Science  
X-Meta-DC.Creator: Jose Manuel Barrueco ; Thomas Krichel; Natividad Noverges; Imma Subirats  
X-Meta-DC.Date: 2001-03-31  
X-Meta-DC.Description: DoIS (Documents in Information Science) is a database of articles and conference proceedings published in electronic format in the area of Library and Information Science. Database contains articles and papers, a lot of them are downloadable from the site. DoIS is a volunteer effort to create a free bibliographic resource of scientific texts specialized in Information Science. It is similar to the highly successful RePEc project for Economics  
X-Meta-DC.Format: text/html  
X-Meta-DC.Format: 3350 bytes  
X-Meta-DC.Identifier: <http://wotan.liu.edu/doi/>  
X-Meta-DC.Language: Multilingua  
X-Meta-DC.Publisher: University of Manchester  
X-Meta-DC.Relation: <http://www.repec.org/>  
X-Meta-DC.Rights: <http://wotan.liu.edu/doi/>  
X-Meta-DC.Subject: Library and Information Science; DoIS; ReLIS ; Documents in Information Science; ReLIS Research Papers in Library and Information Science  
X-Meta-DC.Title: Documents in Information Science  
X-Meta-DC.Type: Service  
X-Meta-GENERATOR: MSHTML 6.00.2722.900

A su vez, que es necesario para alcanzar este objetivo, que desde las escuelas de bibliotecología y ciencia de la información se incluyan asignaturas que apoyen esta formación, investigaciones y actividades de extensión que desarrollen el campo y mayor número de proyectos en esta línea, además, como se ha repetido, de dar ejemplo desde sus mismos sitios y portales web.

En el caso de la Escuela Interamericana de Bibliotecología y de la Universidad de Antioquia este proceso se está iniciando, aunque aún falta mucho:

- Portal EIB (<http://bibliotecologia.udea.edu.co/>)
- “Seminario-Taller Arquitectura de la información y estructuración de metadatos para unidades de información” (<http://aprendeenlinea.udea.edu.co/lms/moodle/course/view.php?id=30>),
- “Seminario de Portales” (<http://aprendeenlinea.udea.edu.co/lms/moodle/course/view.php?id=87>)
- Proyecto Portal U. de A. (<http://www.udea.edu.co/portal.exe>)

Este mismo estudio, refuerza esta afirmación de todo lo que aún nos falta, pues deberían ser nuestros egresados los que estuvieran liderando esta gestión de bandas de metadatos en las bibliotecas del país, buscando una web semántica, pero en el caso de los sitios web, portales y recursos analizados, cuyo origen es Colombia, la carencia de esta gestión, es también evidente.

## REFERENCIAS

- Senso, José A.; de la Rosa Piñero, Antonio. El concepto de metadato. Algo más que descripción de recursos electrónicos recursos electrónicos, En: Ci. Inf., Brasilia, v. 32, n. 2, p. 95-106, maio/ago. 2003 <http://www.scielo.br/pdf/ci/v32n2/17038.pdf> [*Consultado: 10-08-2006*]
- Montoya, Carlos Arturo; Uribe Tirado, Alejandro. La formación bibliotecológica en metadatos desde una perspectiva interdisciplinaria: terminología, informática y comunicación. 8° Congreso Nacional de Bibliotecología y Ciencias de la información. Cartagena, Colombia. Mayo 2006. <http://eprints.rclis.org/archive/00006850/>

Méndez Rodríguez, Eva M<sup>a</sup>. *Metadatos y recuperación de información: estándares, problemas y aplicabilidad en bibliotecas digitales*. Gijón: Trea, 2002.

Merlo Vega, José Antonio; Sorli Rojo, Ángela. Bibliotecas digitales: directorio de recursos. *Revista Española de Documentación Científica*, abril-junio 2000, vol. 23, n. 2, p. 213-220. <http://exlibris.usal.es/merlo/escritos/redcbd2.html> [Consultado: 20-09-2005]

Esta metodología utilizó como base, la metodología utilizada con éxito por los profesores: Rovira, Cristòfol; Marcos, Mari Carmen. Metadatos en revistas-e de Documentación de libre acceso. *El Profesional de la Información* 15(2):pp. 136-144. 2006. <http://eprints.rclis.org/archive/00009452/> [Consultado: 20-03-2007]

Gráfico evolución de la web: [http://novaspivack.typepad.com/nova\\_spivacks\\_weblog/2007/02/steps\\_towards\\_a.html](http://novaspivack.typepad.com/nova_spivacks_weblog/2007/02/steps_towards_a.html) [Consultado: 15-05-2007]



# Software libre en la representación, búsqueda, recuperación e intercambio de información.

DANTE ORTIZ ANCONA  
*Dirección General de Bibliotecas,  
UNAM, México*

## OBJETIVO

Exponer experiencias en la aplicación de herramientas de software libre para la representación, búsqueda, recuperación e intercambio de información.

## RESUMEN:

**I**lustrar las técnicas mas recientes y sofisticadas tales como II, LSI y ECI (de sus siglas en inglés “Inverted Indexing”, “Latent Semantic Indexing” y “Effective Conceptual Indexing” respectivamente) para la creación de índices en los sistemas de recuperación de información.

Presentar algunas herramientas de software libre, tales como Lucene, Zilverline, Lius, Regain , Zebra y Managing Gigabytes, para indexación y búsqueda de información, ilustrando su funcionamiento y operación así como la técnica de indexación que emplean. Se describirá como incorporar un analizador lexicográfico para el lenguaje español.

Presentar un proyecto para mejora de relevancia en recuperación de información en español, utilizando: raíces para formación de pala-

bras (stemming), eliminación de palabras sin importancia en la búsqueda de información (stopwords) y tesauros.

Describir algunos paradigmas de investigación y desarrollo en los sistemas de representación, organización, búsqueda y recuperación de información. Los paradigmas a describir serán: paradigma estadístico, paradigma lingüístico, paradigma bibliotecológico y paradigma de inteligencia artificial (redes neuronales, sistemas basados en conocimiento, procesamiento de lenguaje natural).

Presentar y describir algunas herramientas de software libre, tales como Fedora, DSpace Greenstone para la implantación de repositorios digitales ilustrando sus protocolos para el intercambio de información así como sus aplicaciones en el desarrollo de bibliotecas digitales y para la conservación y preservación de información digital.

## 1. MÉTODOS PARA CREAR ÍNDICES EN LOS SISTEMAS DE RECUPERACIÓN DE INFORMACIÓN

La representación de Índice Invertido (II) es el método dominante para indexar texto, pero no es conveniente en búsqueda de similitud entre documentos. La representación II consiste de una lista de identificadores de documentos, uno por cada palabra en el léxico. Cada palabra *W* tiene asociada una lista de todos los identificadores de documentos que la contienen. Adicionalmente, meta-información que se almacena junto con el identificador del documento, tal como frecuencia de la palabra y posición en el documento. Mediante alguna función de similitud se detectan, en el índice, las palabras correspondientes de una búsqueda. Los detalles de esta función se muestran en [1] y [2]. El desempeño de la representación II empeora cuando se incrementa el número de palabras en un documento o en los casos en que una palabra tiene una lista invertida demasiado grande.

El Indexado Semántico Latente (ISL) es un método para mejorar la calidad en la búsqueda por similitud transformando documentos del conjunto de palabras original a un espacio de conceptos [3]. La idea principal de este método es proyectar los datos en un espacio pequeño, de los datos originales, eliminando los efectos nocivos de sinonimia y polisemia. Trata de minimizar ambigüedad, redundancia y



vocabulario sin comprimir la representación. ISL transforma los datos de una representación indexada dispersa (como en II) con dimensionalidad alta a una representación en un espacio real mucho menos disperso. Desafortunadamente, ISL transforma los datos en un dominio que no es posible brindar técnicas de indexado efectivas.

En el Indexado Conceptual Efectivo (ICE), un documento se representa como conjuntos de atributos que corresponden a conceptos con significado. Cada uno de estos conceptos es definido por una palabra con un peso asociado (frecuencia). La palabra con el peso representa a un conjunto de palabras relacionadas semánticamente. La representación ICE es una representación comprimida que reduce ambigüedad, redundancia y vocabulario no relacionado en un documento. Un vez que se reduce la dimensión de un documento se utiliza el método de índices invertidos para indexar los documentos. El método ICE es mucho mejor que II en búsquedas de similitud y preserva la misma calidad de los resultados, tiene una gran eficiencia computacional y de almacenamiento. Por ejemplo, en una muestra de 167000 documentos se requirió 87.7 Mb usando el método II y 8.3 Mb usando ICE [4].

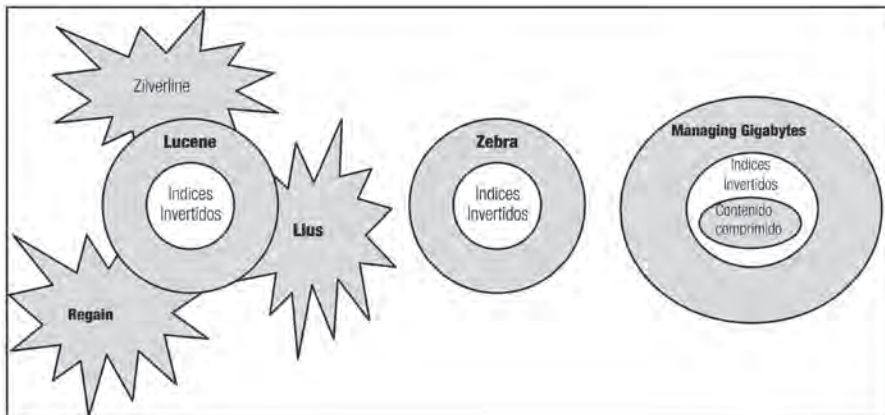
## 2. SOFTWARE LIBRE PARA INDEXACIÓN, BÚSQUEDA Y RECUPERACIÓN DE INFORMACIÓN

Las herramientas de software libre para indexación, búsqueda y recuperación de información más comunes son Lucene [5], Zebra [6] y Managing Gigabytes [7]. Todas estas herramientas tienen como núcleo la representación de índices invertidos (Véase figura 1). Managing Gigabytes incorpora un núcleo de nivel más bajo que maneja algoritmos bastante sofisticados para comprimir la información y manejar de manera más eficiente imágenes, audio y video.

Lucene es una interfaz para programas de aplicaciones que contiene un motor para indexar, buscar y recuperar información tanto de registros como texto completo. Es sin duda, dentro de su clase, el software con mayor respaldo en soporte, documentación y desarrollo de proyectos. Aunque utiliza principalmente el idioma inglés, provee una interfaz de programación que le permite incorporar, con gran facilidad, otros idiomas. Fue desarrollado en el lenguaje de programa-

ción Java, esto le ha permitido una gran portabilidad para funcionar en cualquier sistema operativo y arquitectura de cómputo, así como una gran interoperabilidad con otros sistemas computacionales. No utiliza un sistema de metadatos descriptivo, sin embargo, resulta bastante simple adaptarle cualquier sistema de metadatos.

Figura 1 (Software libre para indexado, búsqueda y recuperación de información)



**Lucene** utiliza por defecto un analizador lexicográfico, para texto en idioma inglés, eliminando del vocabulario palabras sin importancia en búsqueda de información (stopwords) tales como artículos, preposiciones, verbos comunes (is, have, get, etc.), palabras acerca de la estructura del texto, etcétera. El analizador sintáctico permite que un usuario o programador especifique un archivo que contenga esta lista de palabras dando la posibilidad de aumentar o disminuir el diccionario de palabras sin importancia. Además reconoce términos con las características siguientes: secuencias de letras y dígitos (p175waugh), apóstrofes (O'Reilly), acrónimos (H.P), compañías (AT&T), direcciones de correo electrónico (dante@dgb.unam.mx), nombres de servidores WEB (www.dspace.org), números seriales (direcciones IP como 132.248.9.31, números de punto flotante como 3.1416). Puede desarrollarse un analizador lexicográfico para el idioma español o cualquier

otro idioma (que quizá incorpore el uso de raíces de palabras y/o tesauros) y utilizarlo en lugar del que se tiene por defecto. Por ejemplo, sea “vacias.txt” el archivo que contiene las palabras irrelevantes para la búsqueda y sea “SpanishAnalyzer” el Analizador lexicográfico para el idioma español, entonces, el fragmento de código, en el lenguaje de programación Java, para construir dicho analizador lexicográfico sería: **new StandardAnalyzer(new File(“vacias.txt”))**.

La figura 2 muestra una representación en Excel de un subconjunto, del índice construido, utilizando el analizador lexicográfico en idioma español anteriormente mencionado. La primera columna contiene la lista de términos, el primer renglón ilustra parte de una lista con un total de 11 identificadores de documentos de texto en diversos formatos. La segunda columna ilustra la cantidad de documentos en que aparece el término. La celda de intersección entre el término y el identificador del documento muestra el número de veces que aparece el término en el documento de texto. Por simplicidad no se muestran las posiciones del término dentro del documento de texto.

Aunque el motor de Lucene fue desarrollado para indexar, buscar y recuperar información en texto plano. Se han desarrollado otras herramientas de software libre tales como Zilverline [8], LIUS (Lucene Index Update and Search) [9] y Regain [10] que tienen como núcleo a Lucene (ver figura 1) y que amplían su funcionalidad al incorporar filtros que permiten transformar documentos de diferentes formatos (Word, Powerpoint, Excel, Postscript, PDF, HTML, XML, etc.) a texto plano. Estas herramientas proveen una interfaz de usuario vía WEB para administrar el índice, incorporar documentos de texto, realizar búsquedas avanzadas (incorporando operadores lógicos, de agrupamiento, de selección de campos, comodines, de proximidad y de rangos) y recuperar información ya sea textual o descriptiva.

**Zebra** es una interfaz para programas de aplicaciones que contiene un motor para indexar, buscar y recuperar información. Es una herramienta de propósito general y de rendimiento alto, indexa texto estructurado, lee registros en una variedad de formatos de entrada (correo electrónico, XML, MARC) proporcionando acceso a ellos a través de una poderosa combinación de expresiones de búsquedas lógicas y de relevancia. Soporta bases de datos grandes (decenas de

**Primer Simposio Internacional sobre Organización...**

millones de registros, decenas de gigabytes de datos) permitiendo actualizaciones seguras en tiempo real. Soporta el protocolo estandarizado Z39.30 para recuperación e intercambio de información. Cuenta con un amplio respaldo en soporte, documentación [12] y desarrollo. Fue desarrollado en el lenguaje de programación C estándar, esto le ha permitido una gran portabilidad para funcionar en cualquier sistema operativo y arquitectura de cómputo, así como interoperabilidad con otros sistemas computacionales. Provee una interfaz en modo de comandos para administrar el índice y para búsqueda y recuperación de información, permitiendo realizar búsquedas avanzadas (incorporando operadores lógicos, de agrupamiento, de selección de campos, comodines y proximidad).

Figura 2 (Representación en Excel de un subconjunto del índice invertido, generado con un analizador lexicográfico del idioma español).

	A	B	C	D	E	F	G	H	I
1 Término		Frecuencia	173-182.pdf	55-ART_Development.pdf	Bibliografía.htm	Bibliografía.txt	bidi_tac.pdf	pgs-16-22.pdf	repositorios_internacional
2	6	11	6	3	1	1	9	4	
3	9	11	3	2	1	1	5	2	
4 datos		11	11	6	7	7	22	2	
5 es		11	58	18	1	1	92	21	
6 ser		11	8	2	1	1	36	7	
7 uso		11	11	6	1	1	14	3	
8	2000	10	24	4	1	1	1	2	
9 and		10	31		5	5	5	5	
10 autor		10	1	1	1	1	14		
11 digital		10		13	14	14	52	14	
12 digitales		10		13	2	2	28	12	
13 papel		10		2	3	3	9	5	
14 análisis		6	2	2			3	1	
15 artículo		6	1	1			19	4	
16 asi		6	14	2			9	3	
17 autores		6	1				6	1	
18 años		6	1	1			3	1	
19 001-408-9271720		4			1	1			
20 1-58113-231-x/00/0005		4			1	1			
21 p175-waugh		4			1	1			
22 p93-hart		4			1	1			
23 bsandia@ula.ve		3							
24 foster01foster.html		3							
25 iannela/06iannela.html		3							
26 www.adinet.org		3							
27 www.arf.org		3							
28 www.cisco.com		3							
29 www.derechoycultura.com		3							
30 www.dspace.org		3							
31									
32									
33									
34									

**Managing Gigabytes** es una interfaz para programas de aplicaciones que contiene un motor para indexar, buscar y recuperar información de texto completo, archivos binarios, imágenes pictóricas o textuales. Tiene muy poco respaldo en soporte y la única documentación es el libro [14]. A diferencia de otras herramientas, provee algoritmos bastante sofisticados para comprimir texto e imágenes. Fue desarrollado en el lenguaje de programación C y puede interactuar con otros sistemas computacionales. No utiliza un sistema de metadatos descriptivo. Provee un diccionario, en idioma inglés, de palabras sin importancia en búsqueda de información (stopwords) y manejo de raíces de palabras (stemming) para aumentar relevancia en recuperación de documentos y reducir la dimensión del índice. Tiene implantados métodos estadísticos para clasificación y organización de información. Permite realizar búsquedas avanzadas (incorporando operadores lógicos, de agrupamiento, de selección de campos, comodines y proximidad) y recuperación de información de forma interactiva y distribuida.

### 3. PROYECTO PARA MEJORA DE RELEVANCIA EN RECUPERACIÓN DE INFORMACIÓN EN ESPAÑOL

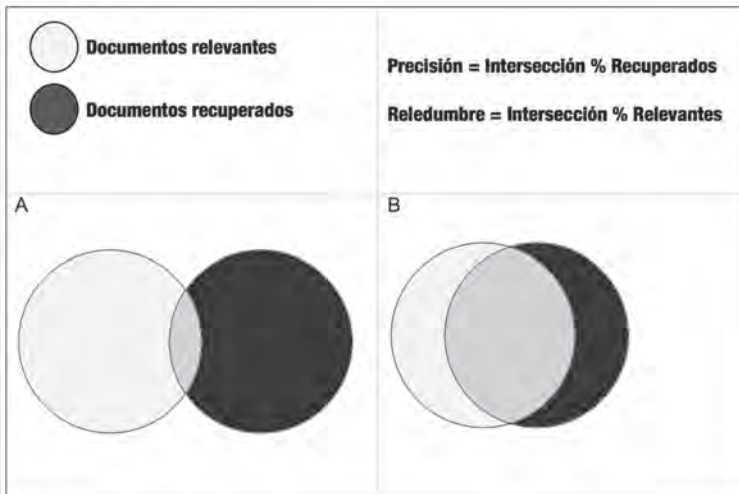
La figura 3 muestra dos medidas básicas en los sistemas de recuperación de información: precisión (precision) y reledumbre<sup>1</sup> (recall) [26]. En color verde se muestra la intersección entre los documentos relevantes para el usuario y los que fueron recuperados en la consulta que planteó. La precisión se define como el número de documentos relevantes recuperados por un usuario, en una búsqueda, dividido entre el total de documentos recuperados. La reledumbre se define como el número de documentos relevantes recuperados por un usuario, en una búsqueda, dividido entre el total de documentos que son de importancia para el usuario. Los sistemas de recuperación de infor-

---

1 Se construyo el término reledumbre (relevancia, certidumbre) usando reglas gramaticales y del mínimo esfuerzo, porque representa más completamente el significado del uso de la palabra recall en lugar de otras traducciones incorporadas tales como exhaustividad, llamada, revocación o recuerdo.

mación tienen como uno de sus objetivos principales incrementar el valor de la precisión y la reledumbre, es decir, transformar el diagrama de la figura 3A por el diagrama de la figura 3B. Algunas técnicas como la eliminación de *stopwords*, filtrado y poda de documentos [14] están orientados a reducir la dimensión de documentos y por tanto la reducción del tamaño del índice; propiciando un mejor desempeño y eficiencia en la búsqueda, recuperación e intercambio de información. Otras técnicas como el tratamiento de la sinonimia, polisemia, raíces de palabras y uso de tesauros, además de reducir la dimensión de los documentos, ayudan a mejorar la precisión y reledumbre en la búsqueda y recuperación de información. Se propone el desarrollo de un proyecto para mejora de reledumbre en recuperación de información en español, utilizando el software snowball [15] que implanta un algoritmo de raíces de palabras (stemming) en español, eliminación de palabras sin importancia en la búsqueda de información (*stopwords*) y un tesoro con términos en español para tratamiento de sinonimia y polisemia. A diferencia de [16] el tesoro propuesto debe agrupar palabras relacionadas semánticamente y que se utilice para implantar el Indexado Conceptual Efectivo en Lucene.

Figura 3 (Medidas básicas en recuperación de información)



#### 4. PARADIGMAS DE INVESTIGACIÓN Y DESARROLLO EN LOS SISTEMAS DE REPRESENTACIÓN, ORGANIZACIÓN, BÚSQUEDA Y RECUPERACIÓN DE INFORMACIÓN

Aunque existen algunas combinaciones de paradigmas, los más comunes son: paradigma estadístico, paradigma lingüístico, paradigma bibliotecológico y paradigma de inteligencia artificial (más generalmente sería paradigma computacional). Existen además muchas herramientas de software libre para filtrado, clasificación, agrupamiento (clustering) y visualización que incorporan todos estos paradigmas, como por ejemplo Weka [17], Yale [18], KDDnuggets [19] y UCI [20].

El paradigma estadístico tiene una gran cantidad de técnicas y metodologías matemáticas, pero que finalmente se basa en la frecuencia u ocurrencia de términos dentro de los documentos. Por ejemplo, los tesauros generados con este paradigma agrupan términos que quizás se manejen como sinónimos porque aparecen en documentos de temas similares y no porque estén relacionados semánticamente. Las técnicas estadísticas tienen la limitante computacional para el análisis de información inmensa (en el orden de terabytes) y para obtener resultados de calidad en búsqueda de información.

El paradigma lingüístico propone soluciones basadas en el conocimiento del lenguaje. Algunos ejemplos son el uso de palabras clave asociadas con el tema de un área del conocimiento, la construcción de tesauros especializados en diversos niveles de abstracción.

El paradigma bibliotecológico está enfocado a proporcionar métodos, técnicas y reglas lógicas para organizar la información. Aportando ideas como normalización, catálogos de autoridades (temas, países, instituciones, dependencias, editores, organismos, etc.) y tesauros. La normalización la definimos como el conjunto de pasos necesarios para implantar sistemas homogéneos y eliminar los heterogéneos (por ejemplo en interfaces de búsqueda, lenguajes para representar información, presentación de información, metadatos, protocolos de intercambio de información, contenidos, etc.) [21].

El paradigma de inteligencia artificial se divide en 3 áreas: redes neuronales, sistemas basados en conocimiento y procesamiento de lenguaje natural. Las redes neuronales artificiales son modelos ma-



temáticos inspirados en la anatomía y fisiología del cerebro humano que permiten hacer computación inteligente realizando procesamiento masivo de datos gracias a su infraestructura paralela y distribuida [25]. Esta área de la inteligencia artificial está relacionada también con temas de estudio de la inteligencia artificial como aprendizaje de máquinas y minería de datos. Los sistemas basados en conocimiento utilizan técnicas computacionales nuevas y herramientas sofisticadas que apoyen a los humanos a almacenar y extraer información útil (conocimiento) de volúmenes de datos inmensos (en el orden de terabytes). Estos sistemas tienen una relación muy fuerte con bases de datos, minería de textos, aprendizaje de máquinas y redes neuronales. El procesamiento de lenguaje natural combina técnicas computacionales con lingüística para procesamiento de texto con el fin de hacer transformaciones de un idioma a otro, introducir texto por voz a una computadora, que un robot entienda instrucciones de un humano y finalmente para el búsqueda y recuperación de información.

## 5. SOFTWARE LIBRE PARA IMPLANTACIÓN DE REPOSITARIOS DE INFORMACIÓN DIGITAL

En la actualidad existe un mínimo de 15 herramientas diferentes, de software libre, para implantación de repositorios de información digital. En el presente trabajo únicamente se presentarán Dspace[22], Fedora [23] y Greenstone [24] tomando en consideración su amplia difusión, soporte técnico, documentación y en especial porque utilizan los métodos de indexación descritos en la sección 1. Estas herramientas aceptan documentos digitales en una gran variedad de formatos tales como Word, Powerpoint, Excel, Postscript, PDF, HTML, XML, GIF, JPEG, TIFF, MP3 y MPEG.

**Dspace** es el software más difundido y de mayor uso en la construcción de repositorios digitales. Tiene un fuerte respaldo en soporte, documentación y está en constante desarrollo. Está programado en Java, esto le ha permitido una gran portabilidad para funcionar en cualquier sistema operativo y arquitectura de cómputo. Lucene forma parte de su núcleo y utiliza sistemas de administración de bases de datos como Postgres y Oracle para el almacenamiento de metadatos y



documentos digitales. Aunque maneja Dublin Core por defecto, puede configurarse para aceptar otro sistema de metadatos como MARC 21. Para el intercambio de información usa el estándar OAI-PMH. Provee una interfaz de usuario, que funciona en WEB, que permite administrar colecciones, definir niveles de acceso a la información y crear usuarios con diferentes permisos para la administración, búsqueda y acceso a los recursos digitales. Se basa en el modelo de referencia OAIS (Open Archive Information System) para la preservación y conservación digital, incorporando algunas estrategias tales como la autenticidad y replicación.

**Fedora** es un software con una gran simplicidad, para instalar y usar, comparado con otras herramientas de software similares. Tiene un fuerte respaldo en soporte, documentación y está en constante desarrollo. Está programado en Java, esto le ha permitido una gran portabilidad para funcionar en cualquier sistema operativo y arquitectura de cómputo. Lucene forma parte de su núcleo y utiliza sistemas de administración de bases de datos como McKoi, MySQL, Postgres y Oracle para el almacenamiento de metadatos y documentos digitales. Su sistema de metadatos es FOXML, Dublin Core y aparentemente, puede configurarse para aceptar otro sistema de metadatos. Para el intercambio de información usa el estándar OAI-PMH y Z39.50. Provee una interfaz de usuario que permite administrar colecciones. Se han desarrollado otras herramientas para ampliar su funcionalidad y definir niveles de acceso a la información y crear usuarios con diferentes permisos para la administración, búsqueda y acceso a los recursos digitales. Ofrece un conjunto de servicios WEB para proveer, a otros programas de aplicación, acceso a la información. Incorpora algunas estrategias de preservación digital como autenticidad, replicación y manejo de versiones.

**Greenstone** es un software que ofrece una gran funcionalidad, tiene un fuerte respaldo en soporte, documentación y está en constante desarrollo. Está programado en varios lenguajes pero principalmente en C estándar y Perl, esto le ha permitido una gran portabilidad para funcionar en cualquier sistema operativo y arquitectura de cómputo. Managing Gigabytes y Zebra forman parte de su núcleo. Cuenta con un sistema de administración de bases de datos propio

(GDBM) para el almacenamiento de metadatos y documentos digitales. Provee Dublin Core por defecto, pero tiene conectores con una gran variedad de metadatos como MARC, XML, METS, CDS/ISIS, etc. Para el intercambio de información usa el estándar OAI-PMH y Z39.50. Provee una interfaz de usuario, que funciona en WEB, que permite administrar colecciones, definir niveles de acceso a la información y crear usuarios con diferentes permisos para la administración, búsqueda y acceso a los recursos digitales. Provee algunas estrategias para la preservación y conservación digital tales como reformateo, rejuvenecimiento y replicación, incorporando algunas herramientas de software para este fin.

## REFERENCIAS

- [1] G. Salton, M.J. McGill. Introduction to Modern Information Retrieval. Mc Graw Hill, New York, 1983.
- [2] C. Faloutsos. Access Methods for Text. ACM Computer Surveys 17, 1, March 1995.
- [3] Dumais S., Furnas G., Landeaur T., Deerwester S., Using Latent Semantic Indexing to improve information retrieval. ACM SIGCHI Conference, 1988.
- [4] C. Aggarwal. On effective conceptual indexing and similarity search in text data. In *IEEE ICDM*, 2001.
- [5] <http://lucene.apache.org>
- [6] <http://www.indexdata.dk/zebra/>
- [7] <http://www.cs.mu.oz.au/mg/>
- [8] <http://www.zilverline.org>
- [9] <http://sourceforge.net/projects/lius/>

- [10] <http://regain.sourceforge.net/>
- [11] Gospodnetic O., Hatcher E., Lucene in action, Diciembre 2004.
- [12] Hammer S., Dickmeiss A., Levanto H., Taylor M., Zebra-User's Guide and Reference, 2005.
- [13] Ian H. Witten, Alistair Moffat, Timothy C. Bell, *Managing Gigabytes, Compressing and Indexing Documents and Images*, Second Edition, Morgan Kaufman Publishers, Inc., Springer 1999.
- [14] J. Lu and J. Callan. Pruning long documents for distributed information retrieval. In *ACM CIKM*, pages 332-339, 2002.
- [15] <http://snowball.tartarus.org/spanish/stemmer.html>
- [16] Ángel F. Zazo, Carlos G. Figuerola, José L. Berrocal, Emilio Rodríguez, Raquel Gómez. *Experiments in Term Expansion Using Thesauri in Spanish*. Grupo de Recuperación Automatizada de la Información (REINA). Depto. de Informática y Automática - Universidad de Salamanca, España. 2003. <http://reina.usal.es>
- [17] <http://www.cs.waikato.ac.nz/ml/weka/>
- [18] <http://yale.cs.uni-dortmund.de/>
- [19] <http://www.kdnuggets.com/>
- [20] <http://www1.ics.uci.edu>
- [21] Castro T. A., García C. E., Ortiz A. D., *Propuesta para interconexión de catálogos*, XVI Asamblea general de IS-TEC, UTPL, Loja, Ecuador, 2007.
- [22] <http://www.dspace.org>

[23] <http://www.fedora.info>

[24] <http://greenstone.org>

[25] Trejo A. M. C., González A. J. G., *Data SOMining Software para el Descubrimiento de Conocimiento en Grandes Bases de Datos de Información Científico Tecnológica*. Tesis de Licenciatura en Ciencias de la Computación, Facultad de Ciencias, UNAM, 2006.

[26] Han J., Lamber M., *Data Mining Concepts and Techniques*, Second Edition, Morgan Kaufman.

SECCIÓN VII  
La terminología en la bibliotecología



# Algunos problemas terminológicos en el dominio de la Bibliotecología y Documentación. Una Babel terminológica-conceptual.

JOSÉ LÓPEZ YEPES

*Centro Universitario de Investigaciones Bibliotecológicas  
Universidad Nacional Autónoma de México,  
Universidad Complutense de Madrid, España*

## 1. ¿QUÉ ES Y CÓMO SE LLAMA ESO QUE SE ENTIENDE POR ¿BIBLIOTECOLOGÍA/BIBLIOTECONOMÍA? ¿DOCUMENTACIÓN? ¿CIENCIA DE LA INFORMACIÓN? CRÓNICA DE UN CONFLICTO TERMINOLÓGICO Y CONCEPTUAL

**U**n denominador común: Información, la palabra mágica. ¿Término polisémico? Las Ciencias de la Información, tal como se las denomina todavía mayoritariamente en España —frente a la alternativa de Ciencias de la Comunicación— tienen como objeto de estudio una naturaleza— la información— de contornos hartamente ambiguos, de universal consideración y, en consecuencia, de dificultosa definición. A efectos de este trabajo, información es —como indica su etimología, la acción de dar forma a algo. Por extensión podemos afirmar que se trata de dar forma a un mensaje de un *modo* determinado para su transmisión a través de un determinado *medio*. Aquí surge el binomio modo/medio de información de tanta trascendencia, lo que permite considerar la información como un modo de adecuación de un mensaje transmisible a través de un medio, en nuestro caso, de un medio de difusión colectiva. Pues bien, son modos informativos y disciplinas consecuentes el periodismo, la comunicación audiovisual,

la publicidad, la propaganda, las relaciones públicas, la documentación... y son medios informativos la prensa, la radio, la televisión, la valla publicitaria, Internet, etc. Modos y medios informativos que se hacen operativos sobre la base de los respectivos procesos informativos. La comunicación, tal y como indica su etimología, se produce cuando sujeto emisor y sujeto receptor en el proceso informativo participan del mismo mensaje, es decir, éste se hace común a ambos estableciéndose así la comunicación. Finalmente, a partir de los mensajes que remedian una necesidad de información en el usuario, éste va construyendo nuevo conocimiento. De entrada, ya podemos proponer que en el origen y conjunción de los tres conceptos, información, comunicación y conocimiento se encuentra la disciplina que cultivamos. Desde esta perspectiva, venimos defendiendo hace años la utilidad de aproximarnos a la naturaleza de la Bibliotecología o Documentación desde los propósitos y metodología de las ciencias informativas, de otro modo, la consideración de la Bibliotecología/Documentación como ciencia informativa del documento.

Efectivamente, esta corriente ha convenido en considerarla ciencia para la ciencia, de un lado y, de otro, ciencia informativa (López Yepes, 1995). En cuanto ciencia para la ciencia, se halla presente en el proceso de comunicación científica haciendo que los logros de las investigaciones lleguen en las mejores condiciones posibles a los destinatarios que, a su vez, se servirán de aquellos para obtener nuevos conocimientos.

Por otra parte la Documentación es ciencia informativa —no es casual que su primera implantación como cátedra fuera en las Facultades de Ciencias de la Información— porque tiene como objeto de estudio un peculiar proceso de información compuesto de sujetos emisores, mensaje, medio y sujeto receptor. Y es, en efecto, información documental la que se produce y se transmite en tal proceso, una información resultante de otra previamente retenida, transformada y recuperada para servir de fuente de nueva información (López Yepes, 2004). De ahí que el profesor Desantes haya definido muy expresivamente la información documental como información de la información o información al cuadrado (1987) adelantándose al juicio de Negroponte cuando afirma que el valor de la información de la información puede ser mayor que el de la propia información (1995).



## Un tema preocupante

Como venimos apuntando, resulta enormemente significativo que los profesionales y estudiosos de las disciplinas relacionadas con el documento, con las fuentes de información y con la transformación de las mismas a fin de remediar necesidades informativas nos preguntemos continuamente —frente a otros saberes y otros profesionales— por la naturaleza del saber que estudiamos y profesamos y cómo debemos denominar los términos más esenciales. A modo de ejemplo, el Congreso de la Asociación Portuguesa de Bibliotecarios, de 2004, tenía como objetivo: “La definición del campo científico, de las relaciones interdisciplinares, del objeto de estudio, de la fundamentación teórica de la Ciencia de la Información y la clarificación de las prácticas y actividades profesionales en conformidad con esos presupuestos es hoy una cuestión crucial para la valoración de un área que hasta ahora se ha afirmado sobre todo por la componente tecnológico. Urge, pues, un debate y un estudio amplio” ([www.apbd.pt](http://www.apbd.pt)). Cabe pensar, a tenor de lo dicho, que estamos ante un tema preocupante que afecta a nuestra credibilidad como estudiosos y como profesionales en el ámbito académico y social respectivamente. Sin embargo, creemos que apenas se ha tenido en cuenta el problema terminológico aunque procede recordar que en la última reunión de las escuelas de bibliotecología de los países de MERCOSUR celebrada en la Universidad de Playa Ancha (Chile) en el mes de octubre del pasado año uno de los objetivos era el de tratar de conciliar los currícula de aquéllas en aras de facilitar los intercambio y de lograr una mayor integración.

Creo honestamente que toda reflexión sobre la figura del profesional de la información, tanto en el aspecto formativo como profesional, debe partir de una idea clara de lo que es la disciplina que cultivamos y, en consecuencia, de cómo debemos formar al estudioso y profesional y cual debe ser su papel en el medio social. En mi caso referido a España, creo que existe un auténtico conflicto conceptual y terminológico sobre ello como lo muestra el hecho de que el problema sigue siendo objeto de estudio en multitud de publicaciones y de reuniones científicas de nuestra especialidad.

Y aquí arranca el primer conflicto: ¿Cómo llamar a esta disciplina? o, tal vez, conjunto de disciplinas: ¿Bibliotecología? ¿Biblioteconomía? ¿Documentación? ¿Ciencia de la Información? ¿Ciencia de los Bienes culturales? ¿Y cómo llamar de modo unánime a los componentes, factores y aspectos diversos de las actividades documentales como, por ejemplo, al profesional? ¿Bibliotecario, bibliotecólogo, documentalista, profesional de la documentación, profesional de la información...? Esto es algo que no ocurre en las disciplinas consagradas como Historia, Medicina, Química, Filosofía, Derecho, etcétera e, incluso, en las todavía relativamente recientes como el Periodismo, el Marketing, la Publicidad, etc. Si grave es la inexistencia de unanimidad en los términos que empleamos, más grave es sospechar que, tras estas dudas, se pueda esconder una falta de unanimidad en su concepto, en su definición, en su fundamentación epistemológica, en suma. Ambos conflictos, si no se resuelven provocan, indefectiblemente, consecuencias de incertidumbre a la hora de planificar líneas docentes y de investigación, a la hora de establecer los objetivos de la disciplina y la figura del estudioso o profesional que debemos formar en nuestros centros. Se impone, pues, la construcción definitiva de un paradigma, en definitiva de una teoría de la documentación siempre actualizada y que sea objeto de estudio preliminar y piedra angular en nuestros centros docentes y en la correcta interpretación por el cuerpo social (López Yepes, 1995, 2005).

## 2. CONSTATACIÓN DEL CONFLICTO TERMINOLÓGICO-DOCUMENTAL. ALGUNOS EJEMPLOS

El conflicto surge del choque o entrecruzamiento de varios términos para designar un concepto y de varios conceptos expresados en un solo término. Esto no ocurre en otros campos del saber: Medicina, Historia, Derecho...

Diversos términos para un concepto:

-Denominación de la disciplina en singular y en plural: Bibliotecología, Biblioteconomía, Ciencia de la Información, Documentación...

- Denominación de las subdisciplinas: Tecnología de la Información, Informática documental...
- Denominación de procesos y operaciones: Catalogación, Clasificación, Análisis documental, Organización del conocimiento
- Denominaciones distintas de los marcos académicos: departamentos, escuelas, facultades: Facultades/escuelas de Bibliotecología, Ciencias de la Documentación, Ciencias de la Información...
- Denominaciones distintas de los títulos académicos: Licenciado, Maestro, Doctor en Bibliotecología, Biblioteconomía, Documentación...
- Denominaciones diversas de la profesión: bibliotecólogo, bibliotecario, documentalista, profesional del documento...

Diversos conceptos para un término:

- De la disciplina: ¿Qué es Bibliotecología?
- De las subdisciplinas: ¿Qué es organización del conocimiento?
- De los procesos y operaciones: ¿Qué es Análisis documental?
- De los marcos académicos: ¿Qué es una facultad de Ciencias de la Información?
- De los títulos académicos: ¿Qué es un Licenciado en Documentación?
- De la profesión: ¿Qué es un bibliotecólogo?

Conflicto de conceptos y términos con campos limítrofes: Sistemas de información (Informática), análisis de contenido (Periodismo), etc.

Consecuencias de esta situación:

1. Positivas: el sabor de la aventura, de seguir investigando en la naturaleza de nuestro campo del saber.
2. Negativas: ambigüedad en el objeto de nuestra investigación, objetivos docentes dispersos a falta de una síntesis razonable, dificultad para establecer modelos formativos de cierta armonía, falta de concreción del papel del profesional en la sociedad, interferencias de otras profesiones, sensación de falta de madurez por

parte de nuestras disciplinas en el sistema de las ciencias, interferencias y solapamientos de asignaturas, etc.

### 3. CAUSAS DEL CONFLICTO TERMINOLÓGICO Y CONCEPTUAL

En mi opinión las causas del conflicto terminológico y conceptual son ocho:

A) La primera, la más remota, radica en la gestación de la disciplina a finales del siglo XIX tanto en el área norteamericana como derivación y conversión de la biblioteca general en biblioteca especializada, de un lado. De otro, en el área europea con el nacimiento del movimiento documental iniciado por Otlet, fundador del Instituto Internacional de Bibliografía, y cuya doctrina aparece sistematizada en su conocido y magno *Tratado de Documentación* publicado en Bruselas en 1934.

B) La segunda causa, mucho más próxima, arranca del cambio social tan intenso que ha producido la eclosión y extensión de ese fenómeno tan presente en nuestras vidas y en nuestras instituciones. Me refiero al gran fenómeno de la información que ha dado lugar al hecho singular de procrear una nueva sociedad con este nombre por la influencia, entre otros factores, del factor tecnológico como motor del cambio y de las crecientes necesidades sociales de información.

C) La tercera causa se deriva directamente de lo que acabamos de esbozar. Efectivamente ha surgido una nueva terminología, de gran impacto, casi a nivel escatológico diría yo, y de fecundidad extraordinaria vinculada a nociones como lo electrónico, lo digital, lo virtual, lo emergente... Pero, sin duda, la nueva terminología es también expresión palpable de que los principios científicos y técnicos de la documentación están experimentando un notable cambio que afecta a todos los elementos del proceso de información documental. En efecto, sujetos emisores, mensaje documentario, sujetos receptores y medios de transmisión pueden experimentar un cambio documentario que hace posible hablar también de sociedad de la documentación. Recordemos que cualquier dato, información o conocimiento son fuentes potenciales de nueva información tan solo cuando se transmi-

ten y se aprovechan en forma de documento, esto es, sobre un soporte físico.

D) La cuarta causa obedece al hecho de que el fenómeno de la información nacido sobre la base de la estructura del proceso vigente desde Aristóteles —emisor, mensaje, medio y receptor— es patrimonio de todas las disciplinas y de sus aplicaciones en la sociedad y provoca una gran convergencia de estudiosos y profesiones que observan el fenómeno desde su articular punto de vista y, claro está, desde una terminología propia. Finalmente, a ello debe añadirse —como quinta causa— las perturbaciones provocadas por el traslado a otras lenguas de términos procedentes del área angloamericana seguidos por muchos de nosotros de modo a veces acríptico y no siempre claramente asimilados a nuestro léxico.

E) La política académica de los países, universidades y centros en cada caso, el diverso origen disciplinar todavía de los autores de los planes de estudio, la influencia de los centros o departamentos donde se imparten estas enseñanzas, etcétera.

F) Las posibles disparidades mostradas por tesauros y diccionarios terminológicos plurilingües o la no utilización de los mismos. Por ejemplo,

G) Las posibles malas prácticas en los procesos de creación del lenguaje científico de la Biblioteconomía y Documentación bien por rendir excesivo tributo a los neologismos, bien por no tener en cuenta directrices de formación de términos (Torres Vargas).

“No hay novedad, por precaria que sea, que la pedagogía patria no absorba a velocidad de vértigo y que no se apresure, a continuación, a explicar mal” (La nueva liturgia, Alvaro Delgado-Gal, ABC, 22-8-2007, 3).

#### 4. PROBLEMAS EN LA CREACIÓN DE LENGUAJE CIENTÍFICO

Como decimos, hay otra fuente de posible confusión en las denominaciones. Efectivamente, en su actividad, el investigador ha de nominar con frecuencia los nuevos conceptos construidos y, en este sentido, se revela como creador de lenguaje e introductor de nuevos términos en su campo de trabajo. Efectivamente, el investigador nombra y, además, define los conceptos descubiertos. Nombrar y definir es oficio

natural del científico. Nominar ya es hacer ciencia. Los términos, como se sabe, son símbolos que describen los conceptos, por lo que la terminología se erige en factor esencial de la definición y de la comunicación de los conceptos.

Los términos, como representación lingüística de los conceptos, se diferencian del lenguaje natural por su uso finalista y preciso y porque son el resultado habitual de un proceso de creación científica que contribuyen a su mantenimiento y uso colectivo —más allá de las modas en tanto en cuanto permanezca vigente la verdad científica por él nominada. Transparencia y consistencia se predicán de los términos correctos. Pues bien, a la hora de proponer términos científicos, es preciso reconocer que hay lenguas como el inglés cuya influencia se deja sentir poderosamente en la formación de la terminología especializada en la mayoría de los campos. El investigador español debe ser muy cuidadoso a la hora de formular o emplear la terminología procurando la mejor adaptación en nuestra lengua, e, incluso, buscando en el acervo lingüístico propio, vocablos olvidados o poco usados que pueden volver a la vida como nombres de los nuevos conceptos, haciendo, pues, innecesario el uso de neologismos o extranjerismos. La cuestión no es baladí. El uso de un extranjerismo comporta el riesgo de una inexacta o parcial comprensión del concepto por él representado. A este propósito, se considera imprescindible que todo investigador tenga a mano a la hora de redactar tres diccionarios: un diccionario de la lengua española, un diccionario de sinónimos y antónimos y un diccionario etimológico, además de léxicos o vocabularios de su campo de conocimiento.

Precisamente, la investigación terminológica y su disciplina, la Terminología, se ocupa de estos problemas. La Terminología es una ciencia interdisciplinar sobre la base de la lingüística, la comunicación, la informática y la teoría del conocimiento y el propio término puede significar la disciplina como tal, lo relacionado con los términos como objetos de estudio y el conjunto de términos que forman un vocabulario especializado. El investigador es usuario de los términos y creador de los mismos conjuntamente por lo que debe conocer los métodos y técnicas de esta rama del saber y, conjuntamente con el terminólogo, hace progresar el lenguaje de su especialidad.

Es, pues, de vital importancia conocer los procedimientos habituales para la formación de términos cuando estos no existan o no deban ser utilizados. Sager señala tres modelos para la formación de términos:

a) “El empleo de las fuentes existentes”, esto es, la adaptación de un término a un nuevo significado. Por ejemplo: nave / nave espacial.

b) “La modificación de las fuentes existentes”, mediante el uso de derivados, combinación de palabras existentes para obtener otras nuevas, etcétera.

c) La creación de nuevas palabras o neologismos es uno de los medios más utilizados por el investigador. Los neologismos son a) resultado de la aplicación de las propias reglas en un idioma determinado; b) procedentes del griego y del latín y, entonces, se denominan *cultismos*, y c) tomados literalmente o adaptados de otras lenguas en forma de *préstamo*.

Además, podemos añadir, d) El uso de la metáfora como elemento creador del lenguaje científico. Ya decía Ullmann que “la metáfora está tan estrechamente entretejida con la textura misma del habla humana que ya la hemos encontrado bajo varios aspectos: como un factor capital de la motivación, como un artificio expresivo, como una fuente de sinonimia y de polisemia, como un escape para las emociones intensas, como un medio de llenar lagunas en el vocabulario y en otros cometidos” ( Cit. por Martín-Municio, 1992 : 224). También vio Ortega el valor de la metáfora como instrumento de la ciencia transplantado de la poesía: “Cuando el investigador descubre un fenómeno nuevo, es decir, cuando forma un nuevo concepto, necesita darle un nombre. Como una voz nueva no significaría nada para los demás, tiene que recurrir al repertorio del lenguaje usadero, donde cada voz se encuentra ya adscrita a una significación. A fin de hacerse entender, elige la palabra cuyo usual sentido tenga semejanza con la nueva significación. De esta manera el término adquiere la nueva significación al través y por medio de la antigua, sin abandonarla. Esto es la metáfora” (1996b : 200-201).

e) En todo caso, y en palabras de García Jiménez, “el científico no aspira a *inventar* un vocabulario original, sino a partir del lenguaje común, depurándolo de tal modo que, sin renunciar a su filiación, res-

ponda al rigor y relativa univocidad que exige su carácter científico” (1993 : 22-23).

## 5. LA TERMINOLOGÍA EN EL ORIGEN DE LAS DISTINTAS ESCUELAS DOCUMENTALES

Nos parece conveniente citar algunas premisas en esta cuestión como son el origen latino de las raíces de la terminología en general, por la vinculación de nuestras disciplinas en origen a los estudios humanísticos. las relación con campos como las tecnologías y también el dominio de la lengua de determinados países.

De modo resumido, podemos observar las aportaciones al modo así como designar el campo en las escuelas más sobresalientes:

### ESCUELAS Y AUTORES

#### **Otlet:**

Documentación, Bibliología, Documentología, Documento, Libro, Bibliografía, Servicio de Bibliografía, Servicio de Documentación, Organización, Biblioteconomía, Organismo documental, Red de información y documentación, Repertorio bibliográfico, Sistema bibliográfico, Trabajo intelectual, Documentación administrativa, Documentación internacional, análisis, Bibliometría, Lenguaje, Método documental, Información documentada, organismo documental, ciencias bibliológicas.

#### **Corriente biblioteconómica:**

Bradford: Documentación, parte de la Biblioteconomía

Shera: Organización bibliográfica=Biblioteca especializada + Documentación (acceso al contenido de los documentos), Bibliotecología especializada (1908), ADI, 1930, Bibliotecología de acuerdo con los tiempos (Shera).

Binomio bibliotecario especializado / documentalista (special librarian/information scientist)



**Corriente documental:**

Briet: Documentación (instrumento de la ciencia)

USA: American Documentations, revista 1955

Pietsch (1966): Documentación

Isakovic (1965): Documentología

Altherton (1965): Informatología

**Corriente informativa:**

**1. Anglo-Americana**

Mooers (1950): Information Retrieval

Vickery(1961);, Information Retrieval, Retrieval Systems, Búsqueda documental

1968: ASIS, Information Science

1945: Vannevar Bush, Information Explosion

Georgia Institute of Technology(1962) Information Science

Technology of Information (¿)

Taylor (1966). Information Sciences

ADI (1966): Information Science

Borko (1968): Information Science

Cronin (1983)

**2. Alemana**

DGD (1954): Documentación

Koblitz (1969): Documentación/Información, Ciencia de la Información y Documentación

Wersig (1975): Ciencias de la Información

**3. Rusa**

Mikhailov (1959): Información científica, Teoría de la información científica

Mikhailov (1966): Informatika

## 6. EL DESARROLLO TERMINOLÓGICO DE LA DISCIPLINA *DOCUMENTACIÓN EN ESPAÑA.*

En nuestro país, donde se ha producido muy recientemente la incorporación de los saberes documentarios a la Universidad y, por tanto, al mundo dinámico de la investigación como actividad promotora de términos científicos, su desarrollo terminológico viene aquejado por las notas generales que afectan a cualquier rama del saber moderno y, especialmente, por la ambigüedad terminológica y conceptual y el colonialismo anglosajón a partir de las siguientes causas

a) Pervivencia de la corriente biblioteconómica tradicional que sigue consagrando términos omnicomprensivos como *biblioteca* y sus derivados.

b) Existencia de focos independientes de estudios documentarios como son los profesionales, los universitarios y los investigadores del C.S.I.C. con intereses aparentemente distintos.

c) Relativa escasez, hasta ahora, de textos y manuales de las disciplinas documentales.

d) Excesivo mimetismo de expresiones del mundo anglosajón, principalmente, sin haber sido digeridas previamente y sin haber contrastado su inserción natural en nuestro sistema lingüístico.

e) Aparente escasa aportación hispánica al vocabulario internacional de la Documentación.

Una cala en la situación podemos llevarla a cabo, en primer lugar, mediante el examen de los términos utilizados para designar el nombre de nuestro sector de conocimiento, denominado oficialmente *Biblioteconomía y Documentación* y de las asignaturas que permiten su dominio. Posteriormente, haremos referencia a los términos que describen algunos aspectos más concretos y a la posibilidad de localizar alguna aportación española.

Hoy día, los nombres mas utilizados para designar el conjunto de las disciplinas que nos ocupan es el de *Biblioteconomía y Documentación*, *Documentación*, *Información y Documentación* y *Ciencias de la Documentación*. El uso del primero se manifiesta sobre todo en el título universitario de Diplomado en Biblioteconomía y Documentación y en la denominación oficial del Área de conociemien-

tos, conjunto integral que, incluye también la Archivística y puede equivaler, por tanto, al de *Ciencias de la Documentación*. El segundo en el título de Licenciado en Documentación y en el modo abreviado de designar nuestro campo. El uso del tercero viene marcado por la influencia de esta denominación por los institutos del C.S.I.C. y por la propuesta para la nueva denominación de la carrera según la reforma de Bolonia (Libro Blanco). El último término hace referencia al deseo de integración de todas las disciplinas documentarias, a la denominación de alguna Facultad universitaria y al título de determinadas obras de conjunto (Manual de Ciencias de la Documentación, Diccionario Enciclopédico de Ciencias de la Documentación).

### Los términos para designar las asignaturas de los estudios de Documentación

En nuestra opinión se produjo una auténtica revolución terminológica en la denominación de algunas asignaturas con motivo de la promulgación del Decreto de diciembre de 1978 sobre líneas directrices para la redacción de los planes de estudio de las Escuelas Universitarias de Biblioteconomía y Documentación. El Decreto se hacía eco de la fundación de las mismas pero aportaba como novedad el añadido del término *Documentación* al de *Biblioteconomía* ( para algunos un mero adorno de modernidad) y además consagraba denominaciones como *Análisis documental*, *Lenguajes documentales*, *Tecnologías de la Información*, etc. Las posteriores Directrices para elaborar planes de estudio correspondientes a la Diplomatura en Biblioteconomía y Documentación y la Licenciatura en Documentación (sin perjuicio del hiato que establece entre ambas carreras por no denominarlas del mismo modo) mantiene la moderna terminología y ha permitido, como se sabe, que las Universidades incluyan en los respectivos planes nuevas asignaturas obligatorias u optativas y, como consecuencia, nuevas posibilidades de nominación. Lamentablemente, y en algunos casos, se registran términos ambiguos bien por pertenecer también a otras áreas de conocimiento o porque en sí mismos no distinguen la materia objeto de estudio. Por ejemplo *Gestión de documentos* (Actividades de la gestión administrativa).- *Comprensión de las fuentes de*

*información en otras lenguas* (Diccionarios en línea. Bases de datos terminológicas).- *Diseño de sistemas de información* (Flujos informativos en las organizaciones) y *Estructura de datos*.

## La normalización terminológica en Documentación

El lenguaje de la documentación puede ser objeto de estudio en sí mismo, como lo hemos venido haciendo hasta ahora o considerado como instrumento de la ordenación y aprovechamiento de los documentos de todas las ramas del saber. En este último caso, no cabe duda que los documentalistas cuando elaboran tesauros están contribuyendo a normalizar el lenguaje científico en general y, en consecuencia, a superar los inconvenientes del mismo derivados de distintas mentalidades, lenguas, dominios geográficos.

En este orden de cosas, y ya refiriéndonos, al primer cometido del lenguaje documental, procede afirmar la necesidad de normalizar la terminología de la Documentación a fin de unificar en la medida de lo posible su vocabulario, definir correctamente los distintos términos y cuidar las traducciones que, procedentes del mundo anglosajón, inundan el mercado español (Díaz Prieto). Ya ha habido intentos de normalización como los representados por la Unesco con el trabajo de Wersig y Neveling (1975) o el más reciente del Cindoc (Consejo Superior de Investigaciones Científicas). En todo caso, la resolución de este problema ha sido uno de los objetivos del **Diccionario Enciclopédico de Ciencias de la Documentación**, gran proyecto de trabajo que, con la colaboración de universitarios, investigadores y profesionales del sector, Editorial Síntesis ha publicado recientemente (2004).

Otro medio de conocer la terminología del campo y de las asignaturas más importantes lo facilita el reciente Libro Blanco elaborado por la Comisión de decanos y directores de centros y departamentos. Se exponen a continuación los nombres de las asignaturas troncales propuestas —es decir, de obligada inclusión en todos los currícula que se programen— y de sus contenidos susceptibles de constituir asignaturas concretas:

**1. Documentos, unidades y sistemas de información, 12 créditos** (Naturaleza de la información. Concepto y evolución de los do-

cumentos. Naturaleza y evolución de los archivos, las bibliotecas y los centros y servicios de documentación. Desarrollo histórico y estado actual de la disciplina. Legislación y políticas nacionales e internacionales. Aspectos culturales, sociales, económicos y legales de la producción y el uso de la información documental. Deontología y perfil del profesional de la Información y Documentación).

**2. Planificación, organización y evaluación de unidades de información, 24 créditos** (Planificación, organización y evaluación de recursos. Procesos, servicios y productos de información. Desarrollo y mantenimiento de fondos y colecciones. Programas de conservación y preservación de documentos. Técnicas de gestión de recursos humanos y financieros. Técnicas de promoción: marketing y formación de usuarios).

**3. Fuentes de información, 12 créditos** (Naturaleza y tipología de las fuentes de información. Acceso y consulta de fuentes de información electrónica. Análisis y evaluación de fuentes. Producción, normalización y difusión de fuentes).

**4. Representación y recuperación de la información, 36 créditos** (Catalogación descriptiva. Indización. Clasificación. Resumen. Desarrollo, mantenimiento y evaluación de instrumentos de organización, control y acceso. Elaboración, mantenimiento y uso de lenguajes documentales. Técnicas de recuperación de información. Sistemas avanzados de recuperación de información. Evaluación de la recuperación de información).

**5. Gestión técnica de documentos de archivo, 18 créditos** (Organización de fondos de archivo. Identificación y valoración. Clasificación y descripción. Acceso y difusión. Gestión del ciclo de vida de los documentos administrativos).

**6. Tecnologías de la información y edición digital, 24 créditos** (Arquitectura de ordenadores. Sistemas operativos. Lenguajes de ordenadores. Sistemas y redes de comunicación. Arquitectura de la información. Programas de aplicación documental. Fundamentos y diseño de bases de datos. Edición de documentos digitales. Bibliotecas digitales).

**7. Fundamentos y metodologías de investigación. Estudios métricos de información, 14 créditos** (Fundamentos, métodos y técnicas de investigación. Recogida y análisis de datos. Estadística des-

criptiva e inductiva. Elementos de análisis multivariante. Estudios métricos de información. Evaluación de la actividad científica. Estudios del uso de la información. Estudios de usuarios).

7. LAS RAÍCES DE LA TERMINOLOGÍA EN LENGUA ESPAÑOLA. LO LATINO MÁS LO ANGLOSAJÓN. LA INFLACIÓN ANGLOSAJONA. PROCEDENCIA Y VÍAS DE LA INFLUENCIA. APROXIMACIÓN A UNA DISTRIBUCIÓN SELECTIVA Y PRELIMINAR DE LOS TÉRMINOS VINCULADOS AL DOMINIO DE LA DOCUMENTACIÓN

**Ex: Extranjerismos**

**Abreviaturas:** CD-ROM, On-Line, KWIC, KWOC, ISBN, ISSN, DSI

**Ex:** Mención de previsible extranjerismo.

**Radicales :**

Anal-Análisis de contenido ex

Anal-Análisis de documentos (document analysis) ex

Anal-Análisis documental ex

Anal-Análisis estadístico ex

Anal-Análisis documental de contenido

Anal-Análisis sociométrico ex

Anal-Analista de información ex

Banc-Banco de datos ex

Banc-Base de datos (database) ex

Bae-Base de datos ex

Biblio-Bibliografía

Biblio-Bibliógrafo

Biblio-Bibliología

Biblio-Bibliometría

Biblio-Biblioteca

Biblio-Biblioteca electrónica

Biblio-Biblioteca tradicional

Biblio-Biblioteca virtual

Biblio-Bibliotecario

Biblio-Bibliotecología

Biblio-Biblioteconomía

Biblio-Biblioteconomía especializada (special librarianship)ex

Busq-Ecuación de búsqueda

Busq-Formulario de búsqueda

Busq-Búsqueda bibliográfica

Busq-Búsqueda documental

Camp-Campo (field) rx

Camp-Campo temático (subject field) ex

Cat-Catalogación colectiva-ex

Cat-Catalogación centralizada-ex

Cat-Catalogación cooperativa-ex

Centr-Centro de análisis de la información ex

Centr-Centro de documentación ex

Centr-Centro de información ex

Centr-Centro de referencia ex

Centr-Centro de suministro de documentos (document supply centre)

Ciencia-Ciencia bibliográfica

Ciencia-Ciencia de la biblioteca

Ciencia-Ciencia de la Ciencia

Ciencia-Ciencia de la Documentación

Ciencia-Ciencia de la Información

Ciencia-Ciencia de la Información y Documentación

Ciencia-Ciencia de los ordenadores

Ciencia-Ciencia documental

Ciencia-Ciencia normativa

Ciencia-Ciencia para la Ciencia

Ciencia-Naukometrija

Ciencia-Scientometry

Ciencias- Ciencias informativas

Ciencias-Ciencias de la Documentación

Ciencias-Ciencias de la Información  
Ciencia-Científico de la información  
Ciencia-Cienciometría

Com-Comunicación científica

Dis-Diseminación de la información ex  
Dis-Diseminación selectiva de la información (selective dissemination of information) ex

Doc-Cuasidocumento

Doc-Docudrama

Doc-Documentación

Doc-Documentación automatizada (mechanized documentation)

Doc-Documentación científica

Doc-Documentación electrónica

Doc-Documental

Doc-Documentalista

Doc-Documentalista de referencia

Doc-Documentalística

Doc-Documentalmente

Doc-Documentario

Doc-Documento

Doc-Documento científico

Doc-Documento primario, secundario.

Doc-Documentología

Doc-Educación en Documentación

Doc-Enseñanza de la Documentación

Doc-Explosión de la documentación

Doc-Formación en Documentación

Doc-Información y Documentación

Doc-Información y Documentación científica

Doc-Medios de documentación

Doc-Mensaje documentario

Doc-Movimiento documental

Doc-Perspectiva biblioteconómica de la Documentación



Doc-Perspectiva documental de la Documentación

Doc-Perspectiva informativa de la Documentación

Doc-Política de información y documentación

Doc-Predocumento

Doc-Semidocumento

Doc-Servicio de Documentación

Doc-Sistema de información documental

Doc-Sistema de información y documentación

Doc-Terminología de la Documentación

Doc-Término Documentación

Doc-Trabajo documental

Doc-Tratamiento documental

Form-Formas de actividad científico-informativa

Fuente-Fuente de información

Gest-Gestión de información y documentación en las organizaciones

Hiper-Hipermedia ex

Hiper-Hipertexto ex

Ind-Indización automática ex

Infor-Ecología de la información

Infor-Economía de la información

Infor-Especialista de la información, documentalista especializado ex

Infor-Evaluación de la información

Infor-Exceso de información

Infor-Explosión de la información

Infor-Información

Infor-Información científica

Infor-Información de la ciencia

Infor-Información digital

Infor-Información electrónica

Infor-Información especializada

Infor-Información online  
Infor-Información registrada (recorded information) ex  
Infor-Información y Documentación  
Infor-Información y documentación científica  
Infor-Informador de la ciencia  
Infor-Informatika  
Infor-Information Management  
Infor-Information Science  
Infor-Information Science and Technology  
Infor-Information scientist  
Infor-Information specialist  
Infor-Informatología  
Infor-Informología  
Infor-Ingeniería de la información  
Infor-Mercado de la información  
Infor-Metaciencia de la información  
Infor-Metainformatika  
Infor-Política de información y documentación  
Infor-Recuperación de la información  
Infor-Servicio de información  
Infor-Sistema de información  
Infor-Tecnología de la información  
Len-Lenguaje controlado  
Len-Lenguaje controlado  
Len-Lenguaje de descriptores ex  
Len-Lenguaje de indización  
Len-Lenguaje de indización (indexing language) ex  
Len-Lenguaje de recuperación  
Len-Lenguaje de recuperación (retrieval language) ex  
Len-Lenguaje documental  
Len-Lenguaje documental (documentary language) ex  
Len-Lenguaje libre  
Len-Lenguaje natural  
  
Lis-Lista autorizada de términos, lista de autoridades, authority list ex  
Lis-Lista de términos ex

Proc-Proceso de resumir  
Proc-Procesamiento de la información  
Proc-Proceso cognitivo  
Proc-Proceso cognitivo  
Proc-Proceso documental  
Proc-Proceso informativo  
Proc-Proceso informativo-documental

Recup-Recuperación de datos (data retrieval) ex  
Recup-Recuperación de documentos (data retrieval) ex  
Recup-Recuperación de información (information retrieval) ex

Resum-Cuerpo del resumen  
Resum-Operación de resumir  
Resum-Proceso general de resumir  
Resum-Referencia del resumen  
Resum-Resumidor  
Resum-Resumidor profesional  
Resum-Revista de resúmenes  
Resum-Servicios de resúmenes  
Resum-Técnica de resumir  
Ser-Servicio de actualización permanente (current awareness service) ex  
Ser-Servicio de alerta (alerting service)

Sist-Sistema de documentación ex  
Sist-Sistema de gestión de la información (information management system) ex  
Sist-Sistema de información ex  
Sist-Sistema de información para la gestión (management information system) MIS ex  
Sist-Sistema de recuperación de datos (data retrieval system) ex  
Sist-Sistema de recuperación de datos (data retrieval system) ex  
Sist-Sistemas de indización  
Sist-Sistemas de recuperación de la información ex  
Sist-Sistemas electrónicos de recuperación de la información

Tel-Servicio de Teledocumentación  
Tel-Teledescarga  
Tel-Teledocumentación  
Tel-Telemática

Teoría-Teoría de la documentación  
Teoría-Teoría de la información científica  
Teoría-Teoría de la información  
Teoría-Teoría general de sistemas

Ter-Término asociado, related term ex  
Ter-Término específico, narrower term ex  
Ter-Término genérico, broader term ex

Trat-Tratamiento automático de datos (automatic data processing) ex  
Trat-Tratamiento de datos (data processing) ex  
Trat-Tratamiento electrónico de datos ex

\*\*\*

Ex-Abstract  
Ex-Acceso  
Ex-Almacenamiento (storage)  
Ex-Boletín de abstracts  
Ex-Byte  
Ex-Citación, cita.  
Ex-Clasificación Decimal Universal  
Ex-Codificación (coding, encoding)  
Ex-Código binario  
Ex-Colegio invisible  
Ex-Conserva de la comunicación  
Ex-Control bibliográfico  
Ex-Correo electrónico  
Ex-Chip  
Ex-Descodificación (decoding)  
Ex-Descriptor

Ex-Eficacia (effectiveness)  
Ex-Eficiencia (efficiency)  
Ex-Enlace, link  
Ex-Entrada (asiento)  
Ex-Estrategia de búsqueda (search strategy)  
Ex-Estudio de usuarios  
Ex-Evaluación del sistema (system evaluation)  
Ex-Hardware  
Ex-Índice de citas, citation index  
Ex-Little Science, Big Science  
Ex-Microforma  
Ex-Necesidades de información  
Ex-Nota de alcance, scope note  
Ex-Notación  
Ex-Palabra clave, keyword  
Ex-Pertinencia (pertinency)  
Ex-Programa  
Ex-Publicación primaria, secundaria.  
Ex-Publicación seriada (serial)  
Ex-Red (network)  
Ex-Referencia bibliográfica (cita)  
Ex-Registro autorizado de autores (registro de autoridades), name authority file  
Ex-Software  
Ex-Thesaurus lingüístico  
Ex-Transferencia de la información (information transfer)

El mundo de la información (comunicación) y de la biblioteconomía/documentación en España. Biblioteca, archivo, información: 3 procedencias. Terminología procedentes de los conceptos de disciplinas afines o disciplinas-marco

**8. EL DESARROLLO TERMINOLÓGICO DE LA DISCIPLINA  
DOCUMENTACIÓN EN EUROPA, ESTADOS UNIDOS  
E IBEROAMÉRICA**

**AREA IBEROAMERICANA**

Bibliotecología

Bibliotecología y Ciencia de la Información

Bibliotecología y Ciencias de la Información

Bibliotecología y Documentación

Bibliotecología y Estudios de Información

Bibliotecología y Estudios de la Información

Bibliotecología, Documentación e Información

Biblioteconomía

Biblioteconomía con habilitación en Gestión de la Información

Biblioteconomía e Información

Biblioteconomía y Documentación

Biblioteconomía y Documentación

Ciencia de la Información

Ciencia de la Información con habilitación en Biblioteconomía

Ciencia de la Información y Documentación

Ciencias bibliotecarias y de la Información

Ciencias de la Información

Ciencias de la Información documental

Ciencias y Tecnologías de la Información

Documentación

**AREA ANGLOSAJONA**

Business Information Technology (Australia)

Information

Information and Communication Sciences

Information and Communications

Information and Library Science

Information and Library Studies

Information Management

Information Management and Systems  
Information Management Systems  
Information Resources and Library Science  
Information Science  
Information Science and Policy  
Information Sciences  
Information Studies  
Information Systems  
Library and Information  
Library and Information Management  
Library and Information Science  
Library and Information Studies  
Library and Information Studies and Archive Studies  
Library and Information Systems  
Library Science  
Library Sciences  
Library, Archival and Informaron Studies  
Library, Archive and Information Studies

**CANADÁ**

Bibliothéconomie et Sciences de l'Information (Canadá)

**EUROPA**

**HOLANDA**

Biblioteks- och Informationsvetenskap

**DINAMARCA**

Biblioteksskole

**ALEMANIA**

Bibliotheks- und Informationswissenschaft

*Primer Simposio Internacional sobre Organización...*

PORTUGAL

Gestão de Informação

FRANCIA

Sciences de l'Information et des Bibliothèques

Techniques de la Documentation

ITALIA

Archivistica, Paleografia e Diplomatica

Beni archivistici e librari

**Beni** culturali

Beni culturali, archivistici e librari

Beni storico-artistici, archivistico-librari e musicali

Biblioteconomia e Paleografia

Culture del testo e del documento

Scienze e conservazione dei beni archivistici e librerie

Scienze archivistiche, librerie e dell'informazione documentarie

Scienze dei beni culturali

**Scienze** della documentazione, archivistiche, biblioteconomiche e vocali

Studio e gestione dei beni culturali

Tecnologie per la conservazione e il restauro dei beni culturali

Teorie e tecniche dell'informazione

UN ESTUDIO DE CASO. TERMINOLOGÍA IBEROAMERICANA  
(PAISES MERCOSUR)

FACULTADES

Filosofía. Humanidades. Educación 10

Biblioteconomía 2

Biblioteconomía y Ciencia de la Información 1

Biblioteconomía y Documentación 1

Comunicación y Biblioteconomía 1

Politécnica 1



ESCUELAS/CENTROS	
Ciencias sociales	7
Bibliotecología	4
Humanidades. Educación	3
Comunicación	3
Ciencias de la Información	2
Archivología	1

DEPARTAMENTOS	
Ciencia de la Información	9
Ciencias de la Información	1
Gestión de la Información	2
Biblioteconomía	3
Bibliotecología	2
Bibliotecología y Documentación	2
Documentación	2
Bibliotecología y Ciencia de la Información	1
Biblioteconomía e Historia	1
Física y Matemática	1

TITULACIONES	
Biblioteconomía	9
Bibliotecología	4
Archivología	2
Gestión de la Información	1
Administración de la Información	1
Ciencia de la Información	1
Ciencia de la Información y Documentación	1

### OBSERVACIONES

En cuanto a denominación de Facultades, se observa a) el predominio notable de facultades relacionadas con filosofía, humanidades y educación; b) el creciente número de facultades específicas de biblioteconomía; c) una de ellas muestra la vinculación con comunicación y d) un único caso de residencia de los estudios bibliotecológicos en una facultad de carácter técnico.

En cuanto a denominación de Escuelas, predominan las ciencias sociales y humanidades y educación pero existen un número notable de facultades específicas con la triple orientación de Bibliotecología, Ciencias de la Información y Comunicación. Por último, se constata un caso de facultad de archivología.

La especificidad es evidente en el nivel de Departamentos. En Brasil predomina de modo notable la denominación Ciencia de la Información. Seguidamente son comúnmente utilizados los términos sinónimos de Bibliotecología y Biblioteconomía. Hay dos casos de Documentación, término muy poco frecuente en el área iberoamericana. Excepcional parece la presencia de un departamento de Física y Matemática.

Se mantienen, no obstante, las denominaciones tradicionales de las titulaciones. En efecto, predominan los títulos de Biblioteconomía/Bibliotecología pero avanzan las titulaciones con el término Información y uno de tradición francamente europea: Ciencia de la Información y Documentación (Brasil). Se registran, asimismo, dos titulaciones en Archivología.

## 9. UN ESTUDIO DE CASO: LAS DENOMINACIONES DE LAS ESCUELAS EN MÉXICO. UN ESTUDIO DE CASO EN QUE CONFLUYEN EL CONFLICTO TERMINOLÓGICO Y CONCEPTUAL. (SLP)

Hay países donde la política académica garantiza uniformidad en el núcleo de los currículos.

### CONCLUSIONES SOBRE CONFLICTOS TERMINOLÓGICO-CONCEPTUALES DE LAS LICENCIATURAS

#### **A) Sobre nombres de centros:**

- Bibliotecología
- Bibliotecología e Información
- Biblioteconomía y Archivonomía

#### **B) Sobre nombre de titulaciones:**

- Bibliotecología (3)

- Bibliotecología y Ciencias de la Información
- Bibliotecología e Información
- Biblioteconomía

**C) Sobre objetivos docentes:**

1. Numerosas funciones:
  - Planificar, diseñar, administrar, comparar, desarrollar, evaluar, organizar, planear, dirigir, supervisar...
2. Objeto de dichas funciones:
  - sistemas de información (bibliotecas, centros de documentación, archivos, etc.)
  - unidades de información
  - bibliotecas
  - centros de documentación
  - centros de información
  - centros de información documental
  - ambientes organizacionales
3. Funciones dispersas
  - educación permanente de usuarios
  - difusión de la lectura
  - uso de recursos informativos en todos los formatos
  - aplicación de nuevas tecnologías

**D) Sobre áreas**

1. No diferenciación de contenidos solapamientos:
  - Usuario de la información (principios básicos del proceso de comunicación y de la expresión oral y escrita)
  - Servicios de información (evolución de la bibliotecología y fundamentos filosóficos y culturales que han dado lugar a la generación de conocimientos).
  - Organización de las unidades de información (Principios generales y particulares de la organización y administración de los servicios de información)

**E) Sobre asignaturas en los diversos currículos**

1. Aparentes repeticiones sin diferenciar en los términos

*Primer Simposio Internacional sobre Organización...*

- Servicios de información
- Unidades de información
  
- Computación
- Análisis de sistemas
- Automatización de unidades de información
- Nuevas tecnologías de información
  
- Computación aplicada a la bibliotecología
- Sistemas y programas de automatización de bibliotecas
  
- Herramientas prácticas de la computación
- Automatización de unidades de información
- Seminario de automatización de bibliotecas
  
- Diagnóstico de servicios de información
- Diseño de sistemas y servicios de calidad
- Diseño de estrategias de servicios de información
  
- 2. Cierta ambigüedad en los términos
- Introducción a la organización del conocimiento
- Redes en el entorno de la información
- Contextualización de las ciencias de la información
- Cuantificación bibliométrica

9. CONCLUSIONES. PROPUESTAS DE ARMONIZACIÓN.

LO QUE HAY QUE RESPETAR. ¿LOS MEJORES TÉRMINOS?

- 1) Nos parece evidente la existencia del conflicto que hemos tratado de mostrar. Ello comporta la existencia de términos ambiguos y equívocos, neologismos innecesarios, etc. lo que repercute desfavorablemente en la comunicación de los estudiosos hispanohablantes y en los intercambios de los alumnos.
  
- 2) En nuestro dominio lingüístico detectamos que el uso dispar de denominaciones y significados procede de la doble concep-

ción de la Bibliotecología/Documentación: el origen bibliotecario (Library Science) y el origen documental que se gesta en Europa y da origen a las escuelas informativas angloamericanas (Documentation, Information Science/ Information Management), alemana (Ciencia de la Información y Documentación) y rusa (Informatika).

- 3) En Iberoamérica ha predominado la corriente bibliotecológica derivada hacia Ciencia de la Información pero sin excluir el término Documentación, presente en Brasil, y cada vez más aceptado en otros países.
- 4) En España, el término Documentación se considera arcaico de alguna manera pero se mantiene por la imposibilidad de utilizar Ciencia de la Información, exclusivo de la comunicación. Se intenta oficializar la expresión Información y Documentación pero aparece como más interesante la tendencia a señalar el matiz integrador de la expresión Ciencias de la Documentación.
- 5) Por las razones presentadas a lo largo del texto, vemos necesaria la posibilidad de armonizar, homogeneizar u homologar en la medida de lo posible la terminología fundamental y su significado en el área hispanohablante, estableciendo un vocabulario común y evitando la colonización indiscriminada.
- 6) La creación de un Grupo de Trabajo sobre armonización terminológica y conceptual —con la colaboración de terminólogos— podría ofrecer como resultado un Diccionario terminológico que contribuyera a alcanzar los objetivos propuestos.



# La estructura sindética como mecanismo de recuperación de la información temática: LCSH y LEMB

AURORA VEGA MARTINEZ, AGEO GARCIA BARBABOSA  
*Universidad de Tulane, Biblioteca Howard-Tilton Memorial, E.U.*

## RESUMEN

Las listas generales de encabezamientos de material constituyen los principales repertorios de terminología controlada que se utilizan como puntos de acceso temático en los catálogos públicos de las bibliotecas y centros de información documental; por ello no puede soslayarse el impacto de su función como sistemas de representación conceptual del universo del conocimiento en los sistemas de recuperación de la información.

El alcance de la estructura terminológica de las listas estandar de encabezamientos de material, tanto en español como en inglés, depende de dos factores críticos: 1) la calidad y la cantidad de términos temáticos que integran cada lista y cada encabezamiento incluido; y la cohesión de la red asociativa que establece las relaciones existentes entre dichos encabezamientos.

La estructura sindética o relacional de cada lista de encabezamientos de materia se construye mediante mecanismos de control terminológico que permiten identificar el tipo de relación que guardan unos términos respecto de los otros; dichos mecanismos tienen como propósito principal hacer posible la representación temática del con-

tenido de los materiales que integran los acervos documentales; y a través de ésta permitir al usuario la identificación de dichos materiales mediante el uso de los sistemas de indización y recuperación de datos que forman parte de los catálogos.

Se presenta un análisis de los mecanismos utilizados para elaborar la estructura sindética de la lista LCSH en inglés y la LEMB en español, con énfasis en los mecanismos de control terminológico utilizados en ambos ámbitos lingüísticos.

## INTRODUCCION

Las herramientas de recuperación de la información documental tales como las listas de encabezamientos de materia, los tesauros y los catálogos en línea presentan, como elemento común, una estructura sindética o de relaciones internas. Las listas estándar de encabezamientos de materia tienen como propósito ayudar al catalogador a crear puntos de acceso temático para la recuperación eficiente de la información contenida en los catálogos. Los catálogos utilizan mecanismos de control terminológico que vinculan encabezamientos similares mediante referencias de y hacia puntos de acceso que permiten identificar en forma eficiente los recursos de información.

La estructura sindética está constituida por el “sistema de vocabulario controlado con todas sus referencias, tal como se utilizan en un catálogo u otras herramientas de recuperación” (Taylor 1999, p. 251). Dicha estructura de referencias relaciona los términos mediante palabras conectivas que establecen vínculos entre diversas formas de los encabezamientos o desde los niveles de relaciones existentes entre ellos.

Las referencias del sistema de Encabezamientos de Materia de la Biblioteca del Congreso de los E.U. A. (LCSH) establece los siguientes tipos de relaciones: de equivalencia, jerárquicas y asociativas (Chan, 2005); estas relaciones dirigen al usuario hacia los términos establecidos como autorizados y sus formas variantes, alternativas y sinónimas; así como a los términos en los niveles inmediatos en las jerarquías inferior, superior y correlativas.



## PROPOSITO Y FUNCIONES

De acuerdo con Taylor, “un vocabulario controlado funciona mediante la elección de una forma preferida de expresar un concepto y asegurando que las formas sinónimas de expresión del concepto se conectarán a la terminología preferida” (Taylor 1999, 143). Mediante el uso de una estructura sindética un vocabulario controlado proporciona al usuario elementos para identificar las formas variantes (Wilson, 1998); sin estos elementos de identificación el sistema de información no sería funcional ya que el acceso a la información representada por dichos términos sería muy difícil o imposible de identificar. Sin un vocabulario controlado el usuario tendría que utilizar como criterios de búsqueda todas y cada una de las formas posibles en que puede nombrarse una obra, una persona o un tópico. El control de autoridades también ayuda a eliminar la información que no es relevante para el usuario mediante mecanismos que acotan el ámbito de referencia de los términos y hacen precisa su alcance conceptual.

Los mecanismos terminológicos aplicados para establecer una forma específica como el único término válido o autorizado para representar una materia y como el núcleo de referencia al que se hacen apuntar los envíos a partir de las referencias de formas variantes, dotan al vocabulario de su principal recurso de sustento del principio de unidad o uniformidad de términos, el cual permite reunir bajo un encabezamiento todos los materiales que tratan principal o exclusivamente sobre un tema.

A continuación se presenta un análisis de los mecanismos utilizados para elaborar la estructura sindética de la lista LCSH en inglés y la LEMB en español, con énfasis en los mecanismos de control terminológico utilizados en ambos ámbitos lingüísticos.

## RELACIONES DE EQUIVALENCIA

La lista LCSH utiliza como mecanismo de vinculación de los términos considerados equivalentes la creación de referencias de envío de “véase”. Los términos equivalentes se conectan mediante los códigos “USE” y “UF” (“Used for”), siendo esta última la entrada recíproca. Es-

tas referencias conectan los términos no utilizados como encabezamientos con el encabezamiento válido. Las referencias “USE” se establecen al vincular los términos sinónimos, las variantes ortográficas, las formas alternativas, las abreviaturas, los acrónimos, los diversos elementos de entrada, los términos opuestos, y los términos de significado “demasiado” restringido (Chan, 2005, p.125). Las referencias “UF” se codifican con etiquetas 4XX’s en el Formato MARC21. Las referencias “UF” constituyen el mecanismo de vinculación por relación recíproca de términos con carga semántica equivalente o muy cercana. Por ejemplo,

**Child rearing**

UF Child raising

Su entrada recíproca se presentaría en la lista como:

Child raising

USE **Child rearing**

Los términos equivalentes que se usan en LEMB como “referencias primarias” o referencias de “véase” van conectadas con el código “V”. Las referencias recíprocas van precedidas de un código formado por una letra equis mayúscula en negrillas (“X”).

Acento

V **ACENTUACION**

La entrada recíproca sería:

**ACENTUACION**

X Acento

- **Términos sinónimos.** Las reglas de construcción de la lista LC-SH establecen que cuando dos o más términos tienen un significado idéntico, uno de ellos debe elegirse como el encabezamiento preferido y los restantes deben consignarse en referencias “USE”. De acuerdo con Chan, “este tipo de referencia se extiende a los términos quasi-sinónimos cuando se considere que no es práctico hacer una distinción de éstos” (Chan, 2005, p.126).

- **Variantes ortográficas.** Se establece una referencia “USE” cuando un término tiene diferentes expresiones ortográficas, ya sea que se trate de formas tempranas o con estructuras gramaticales diferentes. Por ejemplo:

**Databases**

UF Data bases [*encabezamiento anterior*]

y

**Clearinghouses**

UF Clearing houses

Information clearing houses

Y en la lista LEMB:

**MEXICO**

X Méjico

Y su recíproco:

Méjico

V **MEXICO**

- **Abreviaturas, Acrónimos, etc.** En general se establece una referencia “USE” si existe una forma abreviada. Por ejemplo:

**C I S S (Information retrieval system)**

UF Consolidated information storage system

Y en la lista LEMB:

**SIDA**

X Síndrome de inmunodeficiencia adquirida

- **Términos en otros idiomas** En general no se establecen referencias “USE” para los equivalentes en otros idiomas, a menos que tales términos tengan un uso bien establecido en el ámbito lingüístico del catálogo. Por ejemplo:

**Cabañas Bay (Cuba)**

UF Bahía de Cabañas (Cuba)

Cavañas Bay (Cuba)

Puerto de Cabañas (Cuba)

Puerto de Cabañas (Cuba)

y

**Free enterprise**

UF Laissez-faire

Y en la LEMB:

Rock

V **MUSICA ROCK**

- **Términos populares y científicos.** Se utilizan referencias “USE” para vincular los términos equivalentes populares y científicos independientemente de cuales se hayan elegido como términos preferidos. Por ejemplo:

**Crickets**

UF Gryllidae

Y en la LEMB:

**ACIDO ASCORBICO**

X Vitamina C

- **Formas alternativas.** Se establecen referencias “USE” para las formas de los términos que sean más factibles de ser utilizadas por los usuarios en sus criterios de búsqueda. Por ejemplo:

**Airline security personnel**

UF Airline police

y

**Airlines**

UF Air carriers

Air lines [encabezamiento anterior]

Air transportation industry

Airline industry

Aviation industry

Scheduled airlines

Y en la LEMB:

**PAPAYA**

X Fruta bomba

Lechosa  
Mamón

- **Diversos elementos de entrada.** En general se establecen referencias “USE” cuando la forma preferida tiene estructura invertida o estructura ligada por conjunción y cuando se utilizan calificadores para nacionalidad, grupos étnicos y lingüísticos. Por ejemplo:

**Architecture, Classical**  
UF Classical Architecture

y

**Hispanic American decorative arts**  
UF Decorative arts, Hispanic American

Y en la LEMB:

Iglesia y educación  
V EDUCACION E IGLESIA

Este recurso se utiliza con mucho menor frecuencia en la lista LEMB, y en otras listas en español, debido a que las características sintácticas del idioma requieren de un número limitado de encabezamientos de estructura invertida o con calificadores.

- **Referencias generales de envío.** Este tipo de referencias se utiliza para vincular un grupo completo de encabezamientos con un término; o cuando se explica el uso de las subdivisiones flotantes; o cuando se considera que no es práctico incluir todos los términos específicos individuales. Por ejemplo:

**Office, Appointment to**  
USE *subdivision* Officials and employers-Selection and appointment *under names of countries, cities, etc. and names of individual government agencies; and subdivisions* Selection and Appointment *under types of officials*

Y en la LEMB:

### **CATECISMOS**

(Usase para los libros en que se expone la doctrina católica formulada en una serie de preguntas y respuestas y publicados por una autoridad competente); y la subdivisión CATECISMOS Y CREDOS bajo nombres de iglesias, ej.: IGLESIA ANGLICANA - CATECISMOS Y CREDOS.

### **RELACIONES JERARQUICAS**

En la lista LCSH algunos encabezamientos se conectan con dos o más encabezamientos válidos mediante referencias jerárquicas recíprocas: desde términos de significado más amplio, a través de referencias "BT" (Broader term); y desde términos de significado más restringido, mediante referencias "NT" (Narrower term). La lista LCSH introdujo en 1986 el uso de los códigos estándar aplicados en los tesauros (USE, UF, BT, NT y RT) en la notación de su estructura de referencias cruzadas (Dykstra, 1988). Los términos en referencias jerárquicas se codifican con etiquetas 5XX's en el Formato MARC21. Este tipo de mecanismo de vinculación establece los recursos de superordinación y subordinación que sustentan el principio de especificidad de términos, en las actividades de análisis temático y asignación de encabezamientos de materia que se realizan en el contexto de la estructura jerárquica bidimensional que se ha establecido en forma específica dentro del sistema de representación conceptual de la LCSH. Por ejemplo:

#### **Mammals**

BT Vertebrates

NT Primates

y

#### **Cocktails** (*May Subd Geog*)

UF Mixed drinks

BT Alcoholic beverages

NT Margaritas

Martinis

Mint Juleps

Non-alcoholic cocktails

y

**Carpenters** (*May Subd Geog*)

BT Building trades--Employees

NT Joiners

Shipwrights

En la LEMB las relaciones jerárquicas se expresan como referencias “XX” y referencias “Va” (“vease además”). Por ejemplo:

**COCTELES**

X Cocktails

XX BEBIDAS ALCOHOLICAS

**CELULAS Y TEJIDOS VEGETALES**

Va DIFERENCIACION CELULAR

FLOEMA

PLASTIDOS

XILEMA

XX ANATOMIA VEGETAL

CELULAS

CITOLOGIA

Las referencias jerárquicas de la LEMB no siempre coinciden en forma simétrica con la estructura de LCSH, ya que la LEMB no “tesaurizó” su sistema de notación para las referencias; y al no coincidir el uso de los códigos se ha adoptado un uso no diferenciado (y con frecuencia duplicado) de entradas en las referencias jerárquicas, para dar cabida a los términos más específicos (“NT”).

**RELACIONES POR ASOCIACION**

En la Lista LCSH se vinculan mediante el código RT (“Related Term”) a aquellos encabezamientos que tienen una relación diferente a la jerárquica, es decir una relación en la que los términos no tienen una relación género/especie o todo/parte. Por ejemplo:

**Adolescence**

RT Puberty

**Puberty**

RT Adolescence

## REFERENCIAS GENERALES “SA”

Las referencias generales SA (“*See also*”) se utilizan como recurso para señalar al usuario un patrón bajo el cual se ha consignado una clase particular o un grupo de tópicos; y para ayudarle a seleccionar un tipo de encabezamientos que puede ser más apropiado en la búsqueda de información. Se pueden identificar seis tipos de referencias SA:

- a) referencias a encabezamientos que comienzan con la misma palabra o raíz
  - b) referencias a asientos de nombres de instituciones individuales
  - c) referencias a subdivisiones utilizadas bajo encabezamientos generales
  - d) referencias a subdivisiones utilizadas bajo encabezamientos usados como patrón
  - e) referencias a subdivisiones utilizadas bajo nombres personales individuales
  - f) referencias a subdivisiones utilizadas bajo nombres geográficos>
- Tanto LCSH como LEMB utilizan un número creciente de este tipo de notas referenciales como recurso para limitar la extensión física de las listas sin afectar su flexibilidad y expansibilidad.

## CONCLUSION

Las listas generales estandar LCSH y LEMB, desde la perspectiva de sus respectivos ámbitos lingüísticos, utilizan principios de funcionalidad basados en el desarrollo de una estructura interna que establece las relaciones existentes entre los elementos que forma su cuerpo terminológico, constituyendo un vocabulario que controla, mediante referencias cruzadas, la variedad semántica y las relaciones de su asociación subordinada. La consistencia y cohesión de la estructura sindética de ambas listas es fundamental para el cumplimiento de sus objetivos como herramientas de recuperación de la información debido a que su construcción, desarrollada bajo los mismos principios de consistencia y selectividad terminológica, hace posible que el usuario pueda ir acotando sus búsquedas con criterios prede-



terminados para la representación conceptual de los tópicos representados en los recursos documentales de las colecciones.

Los puntos de acceso temático y el control de sus diversas formas de expresión: 1) alivian al usuario de la carga intelectual de identificar cada manera posible en que un concepto puede ser recuperado; 2) le ofrecen la posibilidad de eliminar aquellos términos que no son pertinentes a su búsqueda; y 3) le sugieren términos que podrían ayudarle a recuperar de manera más completa los materiales que son pertinentes a sus necesidades de información.

Es posible identificar dentro de las estructuras sindéticas de las listas LCSH y LEMB la aplicación de mecanismos equiparables que permiten la aplicación de los mismos principios funcionales de control terminológico en dos ámbitos lingüísticos diferentes.

## REFERENCES

- Chan, Lois May. 2005. *Library of Congress Subject Headings. Principles and Application*. Connecticut, Westport:Libraries Unlimited.
- Dykstra, M. 1988. "LC Subject-headings disguised as a thesaurus". In: *Library journal*, 113(4): 42-46.
- Library of Congress Subject Headings*. 2006. 29<sup>th</sup> ed.
- Lista de Encabezamientos de Materia para Bibliotecas*. 1998. 3era. ed. Santa Fé de Bogotá, D.C. (Colombia): Rojas Eberhard LTDA.
- Taylor, Arlene G. 1999. *The organization of information*. Englewood, CO: Libraries Unlimited.
- Taylor, Arlene G. 2000. *Introduction to cataloging and classification*. 9th ed. Englewood, CO: Libraries Unlimited.
- Wilson, Mary Dabney 1998. Specificity, Syndetic Structure, and Subject Access. *Library Resources and Technical Services*, 42 (4), 272-281.



Catálogo de Autoridad Terminológico para el análisis  
de la información documental de la Comunidad  
Universitaria de Unidades de Información  
Especializadas en Ciencias Sociales (CUUICS).  
Una experiencia metodológica costarricense

LUCÍA MATAMOROS SEGURA  
*Universidad Nacional (Costa Rica)*  
*Facultad de Ciencias Sociales*

## 1. ANTECEDENTES

**E**n noviembre del 2001 inicia un proyecto de desarrollo institucional del esfuerzo de distintas unidades de información especializadas en el ámbito de las ciencias sociales, tanto de la Universidad Nacional (Costa Rica) y de la Universidad de Costa Rica, que tienen como horizonte y propósito brindar en cada una de las bibliotecas y centros de documentación los mejores servicios a sus respectivos usuarios, sean estos estudiantes, docentes, investigadores o extensionistas.

Este propósito lo consiguen por medio de una red de información en Internet denominada: “Comunidad Universitaria de Unidades de Información Documental Especializadas en Ciencias Sociales (CUUICS)” y esta conformada por las siguientes Unidades de Información Documental: por parte de la Universidad de Costa Rica: Biblioteca Eugenio Fonseca Tortós, Biblioteca de Trabajo Social, Centro de Investigación de Estudios de la Mujer (CIEM), El Centro Integrado de Documentación Centroamericana de Ciencias Sociales, El Centro de Información del Laboratorio de Etnología, El Centro de Investigaciones Históricas de América Central (CIHAC), El Centro de documenta-

ción del Centro Centroamericano de Población (CCP), el Instituto de Investigaciones en Ciencias Económicas, la Biblioteca de la Facultad de Educación, por parte de la Universidad Nacional se cuenta con El Centro de Información Documental de la Facultad de Ciencias Sociales (CIDCSO), el Fondo Documental de la Escuela de Historia (FBEH) y la Biblioteca del Centro Internacional de Política Económica, donde todas estas unidades constituyeron una metabase que se puede consultar en la siguiente dirección: : <http://cuuics.fcs.ucr.ac.cr/>

Siendo la CUUICS una red de información documental fue necesario establecer políticas conjuntas, lineamientos, normas y procedimientos estandarizados para normalizar los datos y para que los sistemas de las unidades de información documental fueran compatibles para facilitar la difusión y el intercambio automatizado de información, cada una de las unidades de información documental participantes trabajaban en forma independiente de acuerdo con las políticas particulares, con su propio vocabulario libre, para el análisis de difusión de la información documental. Y si bien existían grandes similitudes, tal como el uso generalizado del Macrothesaurus de la OCDE, existía, también el uso de términos muy heterogéneos para referirse al mismo contenido temático, así como el uso del mismo término para diferentes contenidos temáticos.

Sin embargo, a pesar de que se contaba con varios y buenos tesauros, estos no cubrían todas las necesidades de descriptores que las unidades de información requerían. Esta situación se daba con frecuencia debido a factores como: la especificidad de los temas de interés, la singularidad de la realidad social y la historia costarricense o regional, el desarrollo de nuevos enfoques teóricos o la aparición de nuevos fenómenos y situaciones sociales. De tal forma, las unidades no encontraban, en los tesauros disponibles, todos los descriptores que requerían, teniendo así que ir incorporando nuevos términos conforme su trabajo de procesamiento documental lo necesitaba: vocabulario libre o términos no controlados.

Es así como empieza a darse en las bases de datos de las unidades de información documental la presencia de una gran cantidad de términos no controlados, usados como descriptores del contenido de la información y como medio de recuperación de los documentos que

iban siendo incorporados a las bases de datos. Cada centro de documentación y biblioteca especializada buscó de esta manera subsanar las necesidades que los tesauros no satisfacían.

Pero la incorporación de nuevos términos no controlados conllevaba a importantes problemas que atentan contra la calidad del procesamiento documental y el servicio que se brindaba al usuario, además, suponía un grave obstáculo en el esfuerzo por crear un espacio de trabajo colectivo entre las distintas unidades de información documental, incluyendo el compartir bases de datos e integrar procesos técnicos.

Se daban situaciones negativas relacionadas con la ambigüedad y hasta contradicciones en el significado de los términos; un mismo término era usado como referencia a contenidos distintos, en la misma unidad de información documental que lo introdujo, o como referencia al uso que se hacía de ese término en otros centros de documentación o bibliotecas. Así, existían inconsistencias con respecto a la jerarquía y relaciones horizontales entre términos.

Es por eso que fue necesario aplicar una metodología para la normalización de los términos que no se encontraban en los tesauros, trasladándolos a un lenguaje controlado, para que hubiera una unidad de criterio en los términos y métodos de trabajo, y unificar una sola política en el proceso de indización para la CUUICS.

## 2. PROPÓSITO DE LA INVESTIGACIÓN

El propósito de esta investigación fue construir y posibilitar a las tres unidades de información documental de un vocabulario libre normalizado que sirviera como herramienta homogénea en el proceso de indización de los documentos. Esta normalización posibilitó, a su vez una base para la homogeneización en el procesamiento y transferencia de información documental dentro de la CUUICS, para facilitar una pertinente y eficaz recuperación de la información documental por parte de los usuarios.

La revisión, corrección y normalización del vocabulario libre de las unidades de información documental permitirá un intercambio de información documental cooperativamente y eficiente entre las unida-

des de información documental, además de que, trabajen con un mismo vocabulario y se procure una mejor difusión y recuperación de la información documental existente, logrando con ello una indización homogénea de los documentos que permita su recuperación de una forma fácil y exhaustiva, que eliminara ambigüedades y sinonimia entre los términos a la hora de realizar búsquedas en la red.

### 3. OBJETIVOS

Los objetivos generales planteados en esta investigación fueron los siguientes:

#### Objetivo general 1

Realizar un diagnóstico sobre el estado actual de los procesos de indización de la CUUICS.

#### Objetivos específicos:

1. Verificar la existencia de catálogos de autoridad de materia para la asignación de términos en el proceso de la indización en el área de Ciencias Sociales, en cada una de las unidades de información documental.
2. Identificar las políticas y criterios de las unidades de información documental sobre la indización de los documentos de las unidades de información documental: Biblioteca Eugenio Fonseca Tortós de la Universidad de Costa Rica (BEFT), Centro de Información Documental de Ciencias Sociales (CIDCSO) y el Fondo Bibliográfico de la Escuela de Historia de la Universidad Nacional (FBEH).
3. Identificar el tipo de instrumentos normalizados para la indización de los documentos utilizados por las unidades de información documental.
4. Describir cuál es el proceso que siguen las unidades de información documental para la selección de términos nuevos que no se encuentran en los tesauros.

## Objetivo general 2

Sistematizar y normalizar el lenguaje libre empleado por las unidades de información y proponer su uso dentro de la CUUICS.

### Objetivos específicos:

1. Identificar y contrastar, en los tesauros, el lenguaje controlado que utilizan las unidades de información documental para conocer el lenguaje libre.
2. Corregir las incongruencias del vocabulario controlado de las unidades de información documental.
3. Identificar los términos que están en lenguaje libre de cada una de las unidades de información documental.
4. Determinar los problemas de normalización y de uniformidad que representa el uso del vocabulario libre en los procesos de indización.
5. Normalizar los términos que están en lenguaje libre de cada una de las unidades de información documental.
6. Analizar con especialistas en la disciplina de ciencias sociales los términos libres que representan “conceptos” de las unidades de información documental para determinar la incorporación al Catálogo de Autoridad Terminológico para el análisis de la información documental de las Unidades de Información Especializadas en Ciencias Sociales.
7. Proponer los términos normalizados del lenguaje libre que formarán parte del vocabulario controlado de las unidades de información documental.
8. Proponer los procedimientos para la revisión y actualización del vocabulario.
9. Diseñar una base de datos del lenguaje libre normalizado para la CUUICS, utilizando el software Winisis.
10. Publicar los términos que fueron normalizados en forma impresa y digital.
11. Validar los resultados con un grupo de especialistas en Ciencias Sociales.

#### 4. MUESTRA

Para realizar esta investigación se seleccionaron como muestra tres (3) Unidades de Información Documental, de las doce (12) Unidades de Información que actualmente integran la CUUICS, que son: La Biblioteca Eugenio Fonseca Tortós de la Universidad de Costa Rica (BEFT), el Centro de Información Documental de la Facultad de Ciencias Sociales de la Universidad Nacional (CIDCSO) y el Fondo Bibliográfico de la Escuela de Historia de la Facultad de Ciencias Sociales de la Universidad Nacional (FBEH). Las razones para esta escogencia fueron:

- El volumen y complejidad de los datos. Se trata de 3 de las mayores Unidades de Información Documental, por el volumen de sus colecciones, como son el CIDCSO con una base de datos 27272 registros y la BEFT con una base de datos de 26410 registros y el FBEH con una base de datos de 5048 registros.
- La diversidad de las Unidades escogidas. La BEFT ofrece sus servicios para la docencia a toda la Facultad de Ciencias Sociales de la Universidad de Costa Rica, el CIDCSO especializado en la investigación social sobre la realidad nacional y el FBEH que tiene por usuarios privilegiados a los profesores y estudiantes de una Escuela en particular. Esta diversidad incide en el tipo de documentos y el tratamiento en el procesamiento de la información documental que se la da en cada una de estas Unidades.
- La carencia de normalización del lenguaje libre que poseían cada una de estas Unidades de Información Documental.
- La organización y posición Institucional diferenciada.

#### 5. INSTRUMENTOS PARA LA RECOLECCIÓN DE LA INFORMACIÓN

Los instrumentos que se utilizaron para recolectar la información correspondiente a la elaboración del diagnóstico fueron dos cuestionarios con las siguientes características:



**Cuestionario No 1:** elaborado para directores o coordinadores de las unidades de información documental con un total de 9 preguntas: 4 preguntas abiertas, 3 preguntas semiabiertas y 2 preguntas cerradas. Pretendió obtener información sobre las políticas y procesos de indización de cada una de las unidades de información documental.

**Cuestionario No. 2:** elaborado para los funcionarios que realizaban el proceso de indización y el análisis de datos en las unidades de información documental con un total de 12 preguntas: 3 preguntas cerradas, 6 abiertas y 3 semiabiertas. Para recuperar información sobre cómo realizaban el proceso de indización en las unidades de información documental.

**Técnica de grupos focales:** se utilizó una metodología participativa y los resultados se recopilaban mediante la observación, la anotación de las situaciones analizadas y los comentarios, además de las anotaciones y aportes que los participantes hicieron en los documentos analizados.

## 5. TIPO DE INVESTIGACIÓN:

Esta investigación fue exploratoria y descriptiva, con un enfoque cuantitativo por cuanto se revisaron 6150 descriptores del Centro de Información Documental de Ciencias Sociales, 3575 del Fondo Bibliográfico de la Escuela de Historia y 7462 descriptores de la Biblioteca Eugenio Fonseca Tortós. También fue cualitativa porque consistió en las apreciaciones del grupo focal compuesta por un equipo multidisciplinario. Se investigó la situación de las unidades de información sobre los procesos de indización, los instrumentos y la terminología empleada en cada una de ellas. Con el propósito de determinar las políticas de indización.

## 6. SUJETOS Y OBJETOS DE ESTUDIO

En esta investigación se determinaron tres (3) poblaciones como sujetos de información: las dos primeras aportaron información para elaborar el diagnóstico y se consultó a la totalidad de las personas. Y la cuarta (4) población como objeto de estudio. Estuvieron integradas de la siguiente forma:

- **Primera población: directores y coordinadores de las unidades de información estudiadas**
- **Segunda población: documentalistas encargados del análisis documental en las unidades estudiadas**
- **La tercera población estuvo constituida por los usuarios de las unidades de información estudiadas, que son un promedio de 500 por mes.**

Esta población incluyó estudiantes, profesores, investigadores y particulares. Para consultar esta población, se aplicó la metodología cualitativa mediante la técnica de grupos focales y, por lo tanto, se seleccionaron cinco usuarios de cada universidad que cumplieron con las siguientes características:

- Usuarios recurrentes de los centros de documentación.
- Expertos en el área de Ciencias Sociales.
- Estudiantes de tercer nivel y licenciatura en Ciencias Sociales.
- **Cuarta población estudiada compuesta por los descriptores de las tres unidades de información documental.**

Se estudió la población completa que esta distribuida de la siguiente forma: 8216 descriptores controlados, 8328 términos libres y 643 términos ambiguos.

## 6. METODOLOGÍA:

La investigación se desarrolló en trece (13) etapas que fueron las siguientes:

1. Investigación bibliográfica para el desarrollo del marco teórico.
2. Diagnóstico de las tres Unidades de Información Documental, que comprendió investigar sobre la naturaleza de cada una de las unidades y las políticas de indización.
3. Compilación de los términos temáticos de cada una de las bases de datos de las Unidades de Información Documental.
4. Identificación y contrastación en los tesauros del lenguaje con-

- trolado que utilizan las unidades de información documental para conocer el lenguaje libre.
5. Descripción de los tesauros utilizados para identificar el vocabulario controlado.
  6. Corrección para la normalización de los términos del lenguaje controlado.
  7. Selección del lenguaje libre. Análisis, corrección y normalización. Se realizó una tipología por índices geográficos, nombres personales, siglas y conceptos).
  8. Análisis con especialistas de la disciplina de ciencias sociales de los conceptos del lenguaje libre utilizados por las unidades de información documental.
  9. Características y propuesta de la terminología libre para el Catálogo de Autoridad Terminológico para el proceso de Indización de la Comunidad de Unidades de Información Especializada en Ciencias Sociales (CUUICS).
  10. Diseño de la base de datos y digitación de los términos del vocabulario libre normalizado.
  11. Revisión del Catálogo impreso para su respectiva depuración.
  12. Impresión del Catálogo definitivo.
  13. Validación de los resultados.

La información se recopiló por medio de entrevistas, cuestionarios y archivos generados por las bases de datos del programa winisis que se integraron para realizar una revisión minuciosa de los términos controlados contra los instrumentos normalizados para luego identificar el vocabulario libre y proceder a su normalización. Mediante la revisión y análisis de los términos, se detectaron los problemas que presentaban y se propusieron los términos apropiados.

Por último se aplicó la técnica de grupos focales para validar la información obtenida luego del proceso de normalización.

Para analizar la información se utilizó el programa Excel, donde se introdujeron los resultados del diagnóstico para sistematizar los datos que posteriormente, se analizaron y se interpretaron.

Para la elaboración del catálogo, se utilizó el programa Winisis, mediante la construcción de una base de datos donde se digitaron los términos.

## ***Primer Simposio Internacional sobre Organización...***

El cuadro No. 1 muestra un resumen del total de términos de lenguaje libre revisados para la normalización de las bases de datos de las unidades de información documental: CIDCSO, BEFT y FBEH

**Cuadro No. 1. Resumen del total de términos de lenguaje libre revisados para la normalización de las bases de datos de las unidades de información documental: CIDCSO, BEFT y FBEH**

**1977-2004**

<b>Nombre de términos</b>	<b>Total</b>	<b>Porcentaje</b>
Términos controlados que fueron identificados en los tesauros.	5927	36.62%
Términos controlados que presentaban incongruencias	2289	14.14%
Términos de nombres propios	1328	8.20%
Términos geográficos.	1754	10.84%
Términos- siglas-	91	0.56%
Términos libres reubicados a un vocabulario controlado.	439	2.71%
Términos que representan las administraciones de los gobiernos.	107	0.66%
Términos ambiguos-	344	2.13%
Términos de instituciones	196	1.21%
Términos que representan conceptos.	3152	19.47
Términos que representan fechas.	299	1.85%
Mezclas de términos.	129	0.80%
Términos de lenguaje libre reubicados a un lenguaje documental.	132	0.82%
<b>TOTAL</b>	<b>16187</b>	<b>100%</b>

**Fuente: Matamoros Segura, L (2005) Catálogo de Autoridad Terminológico para el análisis de la información documental de la Comunidad Universitaria de Unidades de Información Especializadas en Ciencias Sociales (CUUICS) / Heredia, C.R.: L. Matamoros**

## **7. CONCLUSIONES:**

Esta investigación permitió concluir que:

- Las unidades estudiadas utilizaban herramientas para normalizar y realizar el análisis documental siguiendo los procedimientos establecidos y contando con personal profesional y especializado

en documentación, en el área de Ciencias Sociales; sin embargo, se demostró la falta de normalización. Por ejemplo, 2289 términos controlados presentaban incongruencias tanto en el uso del plural y singular, como digitación y la no utilización de las referencias de los tesauros.

- En la compilación de los términos temáticos de cada una de las bases de datos de las unidades de información documental para su análisis se detectó que la terminología contenía muchos errores ortográficos, mala digitación, diferentes entradas de nombres propios, diversos usos del plural y singular de los términos y ambigüedad.
- El mismo término está en plural y en singular indistintamente en una u otra unidad de información documental.

Ejemplo:

Término	Siglas de la unidad de información documental
Suicidio	CIDCSO
Suicidios	BEFT

Fuente: Matamoras Segura, L (2005) Catálogo de Autoridad Terminológico para el análisis de la información documental de la Comunidad Universitaria de Unidades de Información Especializadas en Ciencias Sociales (CUUICS) / Heredia, C.R.: L. Matamoras

- Términos desconocidos o que generan confusión

Ejemplo:

Término	Siglas de la unidad de información documental
Vicariato de los sentidos	BEFT
Columna	BEFT
Interacción del Estado	CIDCSO

Fuente: Matamoras Segura, L (2005) Catálogo de Autoridad Terminológico para el análisis de la información documental de la Comunidad Universitaria de Unidades de Información Especializadas en Ciencias Sociales (CUUICS) / Heredia, C.R.: L. Matamoras

- Faltan referencias de algunos términos

Ejemplo:

Término	Siglas de la unidad de información documental
A y A	CIDCSO
SAL	CIDCSO
CHUMAD	BEFT

Fuente: Matamoros Segura, L (2005) Catálogo de Autoridad Terminológico para el análisis de la información documental de la Comunidad Universitaria de Unidades de Información Especializadas en Ciencias Sociales (CUUICS) / Heredia, C.R.: L. Matamoros

- En lo que se refiere al lenguaje controlado, se determinó que las unidades de información documental no aplicaban correctamente las referencias sugeridas en los tesauros, no empleaban bien el plural y el singular de los términos a la hora de ingresar la información a la base de datos y hay errores digitación.
- La BEFT lleva un control de su lenguaje libre en una lista en el programa Excel; el CIDCSO incluye estos términos en una base de datos, sin embargo, a través de la revisión en los tesauros, se identificó que algunos de estos términos ya se encontraban en los tesauros. El FBEH no llevaba ningún control.
- Al igual que el lenguaje controlado, el lenguaje libre presentaba muchas incongruencias como las descritas seguidamente:
- En lo que se refiere a los nombres propios, se verificó que existían diferentes entradas para un mismo nombre propio, muchos nombres están incompletos y, en algunos casos, no se dejaba espacio entre la coma y el nombre.
- Los términos geográficos diferían mucho entre las unidades de información documental investigadas. Por ejemplo, CIDCSO realizaba en su proceso de indización una jerarquía de los términos geográficos, donde identificaba el país, provincia, cantón, ciudad, barrio y poblado. La BEFT anotaba solamente la sigla del país y provincia de donde era el término, y en algunos casos no se anotaba el lugar de procedencia. En el FBEH se anotaba el término, pero no se especificaba el país o la provincia de donde provenía.

- Las unidades de información empleaban en sus procesos de indización indistintamente el nombre y la sigla de instituciones u organizaciones, pero en ningún caso existían referencias que indicaban cuál era el término usado.
- En cuanto al nombre de instituciones, se detectó que tenían varias entradas para una misma institución; además, no se especificaban los países. En algunas instituciones, al existir ambos términos, se localizaba por los dos usos del término, pero no se lograba un resultado satisfactorio del 100% de una institución
- Se detectó que existían 99 mezclas de términos, es decir, términos que estaban unidos cuando se digitaron, por ejemplo:

Término	Sigla de la unidad de información documental
HISTORICAMAPANAS NACIONALES	BEFT
HUMANISMO BURGUÉS<PROLETARIADO	FBEH
IGLESIA ESTADO	FBEH

Fuente: Matamoras Segura, L (2005) Catálogo de Autoridad Terminológico para el análisis de la información documental de la Comunidad Universitaria de Unidades de Información Especializadas en Ciencias Sociales (CUUICS) / Heredia, C.R.: L. Matamoras

Estas incongruencias se presentaban en las tres unidades, pero especialmente en el FBEH.

- En cuanto a los términos que se usaban para indicar períodos de gobiernos se identificó que la BEFT utilizaba Administración y el CIDCSO Gobierno, por lo cual no existía uniformidad. Se propuso el uso del término gobierno por ser un término usado en el Tesouro de Historia.

Ejemplo:

Término	Siglas de la unidad de información documental
Administración Arias Sánchez	BEFT
Gobierno Arias Sánchez	CIDCSO

Fuente: Matamoras Segura, L (2005) Catálogo de Autoridad Terminológico para el análisis de la información documental de la Comunidad Universitaria de Unidades de Información Especializadas en Ciencias Sociales (CUUICS) / Heredia, C.R.: L. Matamoras

### *Primer Simposio Internacional sobre Organización...*

- Los términos que representaban “conceptos”, como por ejemplo “Columna”, son ambiguos; algunos no tenían relación con la temática de los documentos y otros eran parte de los títulos de los artículos.
- Una fortaleza que se detectó en cada una de las unidades de información documental es el apoyo de un equipo multidisciplinario para el apoyo en los procesos de indización.

### RECOMENDACIONES:

Como recomendaciones se proponen las siguientes:

- Corregir las incongruencias detectadas en la base de datos de las unidades de información documental, para contar con un catálogo adecuado de autoridades que permita simplificar y agilizar el trabajo de la CUUICS.
- Proponer el uso del Catálogo de Autoridad Terminológico para la normalización de las demás unidades participantes en el proyecto CUUICS.
- Crear una Comisión de Normalización, conformada por analistas y responsables en la toma de decisiones de las unidades participantes, en cuanto a la selección de términos controlados y libres, con el fin de establecer políticas y criterios a la hora de crear términos libres para mantener un lenguaje común.
- Depurar, alimentar y dar un mantenimiento continuo a la base de datos del Catálogo de Autoridad Terminológico para el análisis de la información documental de la Comunidad Universitaria de Unidades de Información Especializadas en Ciencias Sociales.
- Establecer controles de calidad de los registros de la base de datos.
- Utilizar como uno de los instrumentos básicos de normalización las Reglas de Catalogación Angloamericanas.
- Utilizar correctamente el uso de los términos de los tesauros y sus referencias.
- Establecer notas de alcance a los términos que generen confusión o son desconocidos.
- Instalar el Catálogo en formato electrónico en la página de la CUUICS para que este disponible al usuario y al documentalista.



## LOGROS:

- Esta investigación permitió elaborar una propuesta a la “Comunidad Universitaria de Unidades de Información Especializadas en Ciencias Sociales (CUUICS)”, que consistió en un procedimiento de implementación para la normalización de la terminología y un catálogo de autoridad terminológico, para facilitar a los investigadores, docentes y estudiantes el acceso, eficiente y eficaz a la información, disponible a través de la red de información en el ámbito de las Ciencias Sociales.
- Actualmente, se realiza un trabajo conjunto con los bibliotecólogos encargados del proceso de indización de cada una de las Unidades de Información de la Comunidad Universitaria de Unidades de Información Especializadas en Ciencias Sociales (CUUICS), donde se han capacitado para iniciar los procesos de normalización de su vocabulario, de acuerdo a un documento titulado “Algunos elementos para la normalización de la Terminología para la Comunidad Universitaria de Unidades de Información Especializadas en Ciencias Sociales (CUUICS).
- También se ha nombrado una Comisión de Normalización que establece las políticas para la incorporación de la terminología al Catálogo de Autoridad Terminológico.
- Una vez que las unidades de información participantes en la CUUICS hayan depurado todo su vocabulario el producto final será la edición conjunta de un Tesauro Costarricense de Ciencias Sociales.

## BIBLIOGRAFÍA:

- Barité, M. (2002). La terminología: implicaciones y aplicaciones respecto a los lenguajes documentales. Recuperado 11 feb. 2004 del World Wide Web: <http://161.116.1698.9/docs/346/Barité.htm> 2002

**Primer Simposio Internacional sobre Organización...**

- Blaxter, L., Hughes, C., Tight, M. (2000). *Cómo se hace una investigación*. Madrid, España. Gedisa.
- Chain Navarro, C. (2004). "Coincidencia y equiparación en los modelos de recuperación de información". *Documentación de las Ciencias de la Información* (27). 219-236
- CONPROLAT (2002). *Directrices para el establecimiento y desarrollo de tesauros monolingües*. Normas utilizadas en el taller para la construcción de tesauros auspiciado por ILANUD bajo la dirección académica de Conprolat. S.A.
- Correia Goncalves, C. (2001). "La función documentación/ biblioteca: una perspectiva en Portugal". *Documentación de las Ciencias de la Información*. 24. 323-341.
- De la Rosa, A. (1999). Instrumentos terminológicos en el www: xml. *El profesional de la Información*. 8(10). 14-36
- García Jiménez, Antonio (2004). "Instrumentos de representación del conocimiento: tesauros versus ontologías". *Anales de documentación*. 7. 79-95.
- Gil Leiva, Isidoro (2002). Consistencia en la indización de documentos entre indicadores noveles. *Anales de documentación* 5. 99-111.
- Herrero Pascual, C. (1999). "El control de autoridades". *Anales de Documentación* 2. 121-136
- Lafuente López, R. (2003). "Sobre el análisis y representación de los documentos". *Investigación bibliotecológica* (15)30. 163.193.
- Lancaster, F. W. (1995). *Control del vocabulario controlado en la recuperación de información*. Valencia, España: Universidad de Valencia.

- López-Huertas, M. J. (1999). "Potencialidad evolutiva del tesoro: hacia una base de conocimiento experto". *La Representación y Organización del Conocimiento en sus distintas perspectivas: su influencia en la recuperación de la información*. España: EOCNSID
- Martínez, A., Mongaterra, N., Ristuccia, C., Pichinini, M. Y Pené, M. (1997). "Control de autoridades en catálogos en línea". *Investigación bibliotecológica* (11) 23. 80-101.
- Martínez, M. (2004). "Los grupos focales de discusión como método de investigación". *Heterotopia. Tejiendo el Pensamiento desde el Otro Lugar* 26. 59-72.
- Méndez Rodríguez, E. M. (2003). "Metadatos y recuperación de información: estándares", de *Bibliotecarios* 54. 33-57.
- Merlo Vega, J. A. y Sorlí Rojo, A. (2000). "Organizaciones de normalización de Internet". *Revista Española de Documentación Científica* 23(2). 327-340.
- Naumis Peña, C. (2000) "Análisis de la confluencia entre término y descriptor en la elaboración de tesauros". *Investigación bibliotecológica* (14) 29
- Naumis Peña, C. (2003). "Indización y clasificación: un problema conceptual y terminológico". *Documentación y Ciencias de la Información* 26, p. 23-40.
- Naumis Peña, C. (2003). "Tesauro documental para la representación de contenidos educativos en videos". *Investigación bibliotecológica* (17)34. 144-174.
- Rodríguez Luna, C. (2003). *Lenguajes documentales*. Argentina: Universidad de León.
- Sánchez Díaz, M. y Vega Valdés, J.C. (2003). Algunos aspectos teórico-conceptuales sobre el análisis documental y el análisis de información. *Ciencias de la Información* 34, (2)



# La recuperación de la información en la Red Especializada en Ciencias Sociales “CUUICS” en Costa Rica

CRISTINA ROCHA PALMA

*Escuela de Bibliotecología y Ciencias de la Información,  
Universidad de Costa Rica*

## RESUMEN:

Se presenta información concerniente a la creación de la red especializada en Ciencias Sociales (CUUICS) por las Universidades de Costa Rica y Universidad Nacional de Heredia. Se mencionan las normas y políticas que se han establecido para la recuperación de la información, así como la metodología empleada.

**Descriptores:** Redes Especializadas/ Recuperación de la Información/ Ciencias Sociales

## INTRODUCCIÓN

Con la llegada de un nuevo modelo económico y social como la globalización, la información adquiere gran relevancia y desempeña un papel medular en el conocimiento. Los cambios que generan las tecnologías de la información y de las comunicaciones (TIC's) están transformado campos como la ciencia, economía, educación, los estilos de vida, la socialización, la cultura de las personas.

Según Rojas M. (2006) *el conocimiento hoy es información organizada, con una coherencia lógica y empírica, es decir, un conjun-*

*to de afirmaciones que articulan datos, hechos o ideas de forma sistemática y metódica* (p.5). Este conocimiento una vez decodificado e interpretado por aquellos que lo reciben genera un valor, un nuevo saber y, mediante el proceso de comunicación, promueve el aprendizaje, capacidad para tomar decisiones así como facultades para desarrollar actividades intelectuales y manuales, sirviendo de vehículo para otros sectores de la sociedad.

Este fenómeno de la globalización del conocimiento, también ha permeado a las bibliotecas y centros de información, por lo que la información pasa de sistemas manuales como los catálogos, a modelos más interactivos como son las bases de datos en línea y en formato digital convirtiéndose en un recurso importante. Este hecho provoca que en las instituciones, la gestión del conocimiento se organice de forma que permita ser un puente entre las necesidades de información de los usuarios y los recursos existentes, propiciando nuevos conocimientos. Lo que plantea que los esfuerzos se deban dirigir a fortalecer todas las estructuras de gestión e información y que junto al recurso tecnológico constituyan un medio de productividad y competencia.

El desarrollo acelerado de la tecnología permite el establecimiento de redes de información, lo que a su vez proporciona la facilidad en muchas ocasiones de obtener esta información en cualquier tipo de formato, rompiendo las barreras del tiempo y espacio.

El incremento del recurso información promueve el surgimiento de bibliotecas y centros de información especializados, adquiriendo gran relevancia, específicamente, las que se encuentran dentro de instituciones de educación superior, donde la información y el conocimiento ahí contenidos, apoyan el cumplimiento de los planes y programas de estudio, los proyectos y programas de investigación, la acción social y la docencia. La comunidad de usuarios es amplia, variada y requiere de diversos servicios y productos, por consiguiente; se debe estar preparado para enfrentar estas demandas, potenciando estrategias de administración y elevando los niveles de eficiencia y eficacia.

Dada la proliferación de información hoy día en muchos campos del saber y de la diversidad de carreras que florecen a la luz de este

aumento por el conocimiento; surge la necesidad de aunar esfuerzos y establecer vínculos de cooperación entre una serie de unidades de información especializadas en Ciencias Sociales, para potenciar el uso de los recursos de información y brindar mejores servicios a los usuarios.

## 1. RED CUUICS

Desde el año 2001, la Universidad de Costa Rica, específicamente la Facultad de Ciencias Sociales y la Biblioteca Eugenio Fonseca Tortós, junto con su homóloga en la Universidad Nacional de Heredia, han contribuido a establecer una red de información especializada en el área de las Ciencias Sociales, denominado: Comunidad Universitaria de Unidades de Información Especializadas en Ciencias Sociales [CUUICS] <sup>1</sup> como proyecto inscrito a la Vicerrectoría de Investigación de la Universidad de Costa Rica .

Según el documento “Proyecto Conjunto de Desarrollo Institucional” propuesto colectivamente por las dos universidades, la “Comunidad”<sup>2</sup> tiene el *propósito de crear un espacio de trabajo colectivo para todas las unidades proponentes, que posibilite la consecución del interés común de optimizar el uso de los recursos que en información, para el ámbito de las Ciencias Sociales, invierten las universidades públicas por medio de la construcción conjunta de nuevos procedimientos, herramientas e instrumentos para la recuperación, procesamiento y difusión de la documentación que satisfaga oportuna y eficientemente las demandas académicas de la comunidad universitaria* (2001, p.7).

Por lo anterior se pretende dar a conocer, cómo ha sido la recuperación de información en esta red especializada, sus procesos, normalización, herramientas, ventajas y la participación de todas las unidades de información en esta iniciativa.

---

1 Véase <http://cuuics.fcs.ucr.ac.cr>

2 Se utiliza en referencia a la Comunidad Universitaria de Unidades de Información Especializadas en Ciencias Sociales

Según Arteaga-Fernández (1996) citado en Matamoros S. (2005), *la red al igual que un sistema es un conjunto de unidades de información que interactúan entre sí para alcanzar fines y objetivos comunes...*(p.71). Implica un compromiso mutuo de intercambio de información, el establecimiento de relaciones cooperativas y la normalización en el uso de herramientas y procedimientos de forma organizada y responsable.

El establecimiento de redes según (Matamoros S., p.73.) permiten:

- *Actuar conjunta y coordinadamente para lograr un beneficio común*
- *Participar, integrar y multiplicar recursos*
- *Desarrollar tareas comunes en la adquisición, verificación e intercambio de información.*
- *Asegurar la calidad, la pertinencia y la eficacia de los servicios.*

Con la creación de la red y el propósito mencionado, se establecieron otros objetivos particulares para la consecución del proyecto<sup>3</sup>:

- Acceso a toda persona por medio de Internet, a todas las bases de datos especializadas participantes en la red.
- Mejorar la productividad en la recuperación, procesamiento y análisis documental de los nuevos materiales, evitando la duplicación de tareas.
- Racionalizar y optimizar la inversión de recursos en la obtención de nuevos materiales documentales basándose en el conjunto de las colecciones de las unidades participantes para el establecimiento de políticas, procedimientos y mecanismos colectivos en el desarrollo de las colecciones.
- Aumentar la productividad del recurso humano por medio de la normalización de procesos y desarrollo de nuevas herramientas informáticas comunes a todas las bibliotecas y centros participantes.

---

3 Proyecto Conjunto de Desarrollo Institucional, p.5



- Capacitación del personal en herramientas informáticas para el manejo y difusión de la información.

Los componentes de la red son centros cooperantes ubicados en Bibliotecas, Centros de Documentación e Institutos de Investigación, que aportan información relacionada con las áreas de: educación, derecho, antropología, sociología, etnología, historia, ciencias de la comunicación colectiva, trabajo social, psicología, historia, ciencias económicas, ciencias políticas, política económica, deportes, población, museología, estudios de la mujer. Cada centro participa según sus propias condiciones y necesidades, pero aplicando normas y procedimientos comunes para el procesamiento de la información.

El centro coordinador se encuentra ubicado en la Biblioteca Eugenio Fonseca Tortós<sup>4</sup> de la Facultad de Ciencias Sociales de la Universidad de Costa Rica, desde donde se promueven y supervisan todas las actividades. Las decisiones se toman en las reuniones de la red.

En cuanto a los usuarios de la red, son generalmente, estudiantes, docentes, investigadores, administrativos y público externo, que la consultan desde diversas ubicaciones dentro y fuera de la universidad, para solventar una demanda de información de carácter específica y especializada.

Esta situación de especificidad, conduce a la “Comunidad” a trazar algunas estrategias para integrar el procesamiento y almacenamiento de la información, y por lo tanto, que su recuperación y uso sean asequibles a aquellos que consulten las bases, con el fin de recuperar la información lo más precisa y pertinente posible, de acuerdo con sus necesidades.

Como es sabido desde hace varios siglos, la bibliotecología ha mostrado un interés en el establecimiento de normas que regulen su actividad. El aumento constante de documentos propone la creación de criterios, para las operaciones relacionadas con el tratamiento intelectual de éstos. Según Garrido Arilla (1996) citado en Pené (2006) alude que el término normalización documental consiste en *estable-*

---

<sup>4</sup> Véase <http://beft.fcs.ucr.ac.cr/>

*cer reglas que aseguren la interconexión de sistemas que permitan una mayor facilidad en el tratamiento documental y en la transferencia de la información* (Parra. 4.1). Consiguientemente, se han establecido procesos, metodologías y normas estándar.

## 2. PROCESOS TÉCNICOS

En cuanto a la transferencia de información, se ha establecido que las unidades de información participantes, empleen en el procesamiento de la información la siguiente metodología:

- Todo ingreso de información a las bases de datos bibliográficas se efectúa mediante el uso de las Hojas de Descripción Bibliográfica (HDB) y Hojas de Análisis de Contenido (HAC) del Sistema CEPAL que permiten la visualización de información de los registros así como la obtención de catálogos, además de la guía para la entrada de datos denominado Manual de Procedimientos de CEPAL.
- El manejo de la información se realiza mediante el aporte de la UNESCO del programa computacional ISIS (Winisis versión 5.1) que permite el diseño y manejo de bases de datos textuales. Asimismo la adquisición del programa WXIS, permite que estas bases de datos sean acezadas a través de Internet tanto a nivel nacional como internacional, ocupando un posicionamiento en el amplio campo de la información.
- La descripción bibliográfica se efectúa a través del uso de las Reglas de Catalogación Angloamericanas. 2ª. Edición.
- En los procesos de indización, el control terminológico se realiza mediante la utilización como lenguaje documental principal, el Macrothesaurus de la OCDE. Es un vocabulario controlado, que permite procesar con mayor cobertura, información relativa al desarrollo económico y social y que se adapta a las necesidades temáticas de información de los centros participantes. Igualmente se cuenta con el apoyo de otros lenguajes, según sea la especialidad requerida.

Aspecto que ha tomado mayor relevancia es la indización de los conceptos contenidos en los documentos; como política, se ha establecido la indización de tipo exhaustiva.

Esta exhaustividad permite que el usuario, en la recuperación de información encuentre todos los posibles puntos de acceso, que respondan relevante y pertinentemente a una pregunta realizada.

### 3. RECUPERACIÓN DE LA INFORMACIÓN

En las Ciencias Sociales, la recuperación de información es de suma importancia, pues los avances y cambios en esta área, así como las investigaciones constantes por parte de diversos tipos de usuarios, obligan a mantener una depuración de la terminología utilizada. En ocasiones los términos seleccionados para la indización, no son recuperables a través los lenguajes documentales; esta circunstancia motiva a realizar permanentemente consulta a especialistas, docentes e investigadores específicos según el área, con el fin de normalizar la entrada de nuevos términos al sistema.

Los usuarios de la CUUICS, tienen la opción de recuperar la información bajo tres modalidades de búsqueda : palabra clave, autor, título, orientando su demanda hacia una o todas las bases de datos del sistema.

Por las características de especialización de los centros participantes en Ciencias Sociales, y que la recuperación de información es muy específica avocada a esta especialización, se establece que tanto la indización como la clasificación se sustente en el macrothesaurus de la OCDE como vocabulario controlado.

El usuario en la búsqueda de información, recuperará aquellos conceptos o términos que hallan sido seleccionados previamente, como los mas representativos de los contenidos temáticos según el lenguaje documental que utilicen los analistas, conformado la mayoría de las veces por tesauros especializados. El lenguaje controlado según Lancaster (1992, p.180) es más práctico ya que proporciona al usuario un punto de búsqueda y reduce el silencio.

De igual forma, el usuario siempre enfrentará problemas de sinonimia, polisemia y otros accidentes lingüísticos, producto de las estrategias de búsqueda al interactuar con el sistema, por el uso de un lenguaje natural contra los lenguajes controlados empleados en las distintas unidades de de información. Se requiere entonces que el

usuario obtenga como resultado todos aquellos documentos relevantes, pertinentes, que respondan con precisión y exhaustividad a esa petición formulada. Sin embargo, la mayoría de la veces los usuarios no saben plantearle al sistema su necesidad real de información.

Autores como Moreiro (2004) y Lancaster (1995) mencionan que los sistemas con lenguajes naturales, ofrecen una terminología más amplia, con gran especificidad en la recuperación. En la "Comunidad", cada centro ha creado listas de términos propias, con la finalidad de llenar vacíos que los vocabularios controlados no cubren, sea por su especialidad, por el aumento constante de nueva información y por la necesidad de crear nuevos términos que se ajusten a las necesidades actuales de información.

El manejo de este tipo de vocabulario libre, utilizado sin ninguna normalización, dio origen en el año 2005, a que se formulara un proyecto de graduación, que regulara y normalizara ese lenguaje para toda la red. El proyecto de graduación de la Universidad Nacional de Heredia, desarrollado por la Señora Lucía Matamoros Segura y denominado: **CATÁLOGO DE AUTORIDAD TERMINOLÓGICO PARA EL ANÁLISIS DE LA INFORMACIÓN DOCUMENTAL DE LA COMUNIDAD UNIVERSITARIA DE UNIDADES DE INFORMACIÓN ESPECIALIZADAS EN CIENCIAS SOCIALES (CUUICS)**, pretende normalizar todos aquellos términos que no se encuentran en los tesauros y trasladarlos a un lenguaje controlado, para que exista una unidad de criterio en los términos, métodos de trabajo y unificar las políticas de indización en la red.

#### 4. RESULTADOS OBTENIDOS

Las distintas unidades participantes en la Red CUUICS mediante el esfuerzo mancomunado de todas, han manifestado algunos resultados como:

- Intercambio de recursos bibliográficos
- Cooperación entre todas las unidades de información
- Acceso a todas las bases de datos
- Diseminación de información

- Capacitación y asistencia técnica
- Posicionamiento en el mercado de la información
- Mayor interacción entre usuarios y las unidades de información
- Capacitación profesional
- Interés de las autoridades universitarias por esta red de información
- Conocimiento del entorno institucional

## 5. CONCLUSIONES

Con base en la argumentación anterior, se pueden establecer las siguientes conclusiones:

- Hoy por hoy, la implementación del catálogo no ha sido posible cumplirla a cabalidad; pues los centros no han incorporado el catálogo a sus propios sistemas, ya que la revisión, análisis y comprobación de este vocabulario con respecto a los tesauros que poseen, ha resultado una labor ardua dadas las actividades propias de cada unidad y por la escasez de personal profesional en ellas.
- Sin embargo la tarea sigue pendiente y el grupo coordinador está a la expectativa del cumplimiento de esta fase, para unificar los procesos de indización y el logro de una pertinencia terminológica
- A pesar de que existen limitaciones para un excelente desempeño de CUUICS; la red ha funcionado para los usuarios tanto internos como externos, que requieren obtener información especializada en el campo de las Ciencias Sociales.
- Los usuarios en general, tiene la posibilidad de acceso desde cualquier sitio donde se encuentren, lo que les permite ahorro de tiempo en la búsqueda de información.
- La red ha llenado vacíos de información que frente a otros sistemas existentes en las universidades, con mayor capacidad y con mejor tecnología, no han podido cubrir dada la terminología tan específica que requiere el usuario de las Ciencias Sociales.
- El profesional de estas unidades de información, debe estar en constante actualización, ya que los cambios y el aumento de la información son constantes y obligan a estar al día en cuanto al

surgimiento de nuevos términos y de nuevas aplicaciones especialmente en los sistemas de recuperación de la web.

**BIBLIOGRAFIA CONSULTADA:**

Adeltein, A.; Cabré. MT.(2003).*Representación lexicográfica y terminográficas de las unidades terminológicas*. Ponencia VII Simposio Iberoamericano de Terminología RITerm. (2000) [Recuperado el 20 de junio 2007] de <http://www.upf.edu/pdi/df/teresa.cabre/docums/ca-03al.pdf>

Aja Q., L. *Gestión de información, gestión del conocimiento y gestión de la calidad en las organizaciones*. ACIMED. [online]. sep.-oct. 2002, vol.10, no.5 [Recuperado el 20 de junio 2007] de [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S1024-94352002000500004&lng=es&nrm=iso&tlng=es](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352002000500004&lng=es&nrm=iso&tlng=es)

Carrion G.,MI.(1988) *Manual de bibliotecas*. Madrid: Fundación Germán Sánchez Ruipérez.

Castells, MI.(1996) *La era de la Información. vol.1 La Sociedad Red*. Madrid: Alianza Editorial. p.550

*Comunidad Universitaria de Unidades de Información Especializadas en Ciencias Sociales*. (Proyecto Conjunto de Desarrollo Institucional) (2002) (Documento no publicado) San José, Costa Rica: Universidad de Costa Rica, Universidad Nacional. Facultad de Ciencias Sociales.

Lancaster, FW. (1995).*El control del vocabulario en la recuperación de información*. Madrid: Universitat de Valencia.

\_\_\_\_\_.(1996). *Indización y resúmenes: teoría y práctica*. Buenos Aires: EB Publicaciones.

López Y., J.(Coord.) *Manual de Ciencias de la Documentación*. Madrid: Ediciones Pirámide.

Matamoros S., L. (2005) *Catálogo de autoridad terminológico para el análisis de la información documental de la Comunidad Universitaria de Unidades de Información Especializada en Ciencias Sociales (CUUICS)* Tesis para optar al título de Licenciada en Bibliotecología y Ciencias de la Información. Universidad Nacional. Heredia, Costa Rica.

Moreira G., J.A. (2004). *El contenido de los documentos textuales: su análisis y representación mediante el lenguaje natural*. Madrid: Ediciones Trea.

Pené. Monica G. Panorama general de las herramientas de gestión., Unidad 2.(2007) Extraído el 20 de junio 2007 desde [www.r020.com.ar/docs.php?id=336](http://www.r020.com.ar/docs.php?id=336)

Rojas M., Y. (2006) *De la gestión de información a la gestión del conocimiento*. ACIMED. [online]. ene.-feb. 2006, vol.14, no.1 [Recuperado el 27 Julio 2007], de [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S1024-94352006000100002&lng=es&nrm=iso](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352006000100002&lng=es&nrm=iso)





# La normalización de nombres geográficos contenidos en encabezamientos de materia

NOÉ ÁNGELES ESCOBAR  
*El Colegio de México*

## RESUMEN

Los nombres geográficos son parte del patrimonio cultural de los pueblos y por tanto vale la pena el esfuerzo para evitar que la información que transmiten, y que encontramos en las colecciones de nuestras bibliotecas, se disperse y pueda comunicarse de manera efectiva. Nadie mejor que cada país para definir las formas válidas que deben adoptar sus propios nombres geográficos. En este reporte se analiza el ámbito que caracteriza a los nombres geográficos como entidades portadoras de información y su tipología, tanto en términos generales como en el entorno bibliográfico, en su integración como parte de EM, y se describe el desarrollo de la investigación que tuvo como objetivo establecer una propuesta para su normalización en México.

## PANORAMA GENERAL

En México en general no existe suficiente consenso normativo en la representación de encabezamientos de materia. En este momento se cuenta principalmente con cuatro listas que son la fuente de

consulta para la asignación de encabezamientos de materia (EM) en español:

*Bilindex-general: a list of Spanish-English bilingual subject heading equivalents to Library of Congress Subject Heading List*, Oakland, CA: Floricanto Press, [CD-ROMs].

Consejo Superior de Investigaciones Científicas (España), Unidad de Coordinación de Bibliotecas, *Lista de encabezamientos de materia de la red de bibliotecas del CSIC*, Madrid: Consejo Superior de Investigaciones Científicas, 2000.

Escamilla González, Gloria, *Lista de encabezamientos de materia*, 2a ed., México: UNAM, Instituto de Investigaciones Bibliográficas, 1978.

*Lista de encabezamientos de materia para bibliotecas*, 3a ed., Santafé de Bogotá: Rojas Eberhard, 1998.

El hecho de que existan cuatro diferentes vocabularios controlados a nuestra disposición no significa necesariamente que esto resulte pernicioso, pues en el último de los casos podría considerarse que redundaría en una riqueza léxica de fuentes que empleadas convenientemente se complementarían.<sup>1</sup> Sin embargo la situación cambia cuando se les mira desde el punto de vista estructural y de uso: en tal caso, si los EM de estas listas no pasan por un proceso de adaptación en los catálogos bibliográficos, se genera dispersión (descontrol) de la información que cada uno por su parte controla. Esto ocurrió (y ocurre)

---

1 Así resulta para la lengua francesa cuando se compara el *Répertoire de vedettes-matière* de la Université Laval (Canadá) con el *Répertoire d'autorité-matière encyclopédique et alphabétique unifié* (RAMEAU) de la Biblioteca Nacional de Francia y se justifica la existencia de dos lenguajes documentales que tienen como base una misma lengua natural: se reconocen las circunstancias nacionales, lingüísticas, culturales e incluso las administrativas, sobre las que se construyó cada uno de ellos y, al mismo tiempo se mantiene un espíritu de colaboración internacional y de respeto por las necesidades locales. Véase J.A. Bélair, F. Bourdon y M. Mingam, "The Répertoire de vedettes-matière and RAMEAU: two indexing languages in French: a necessary luxury?", en *World Library and Information Congress: 71th IFLA General Conference and Council: "Libraries - A voyage of discovery"*, August 14th - 18th 2005, Oslo, Norway.

con bibliotecas que originalmente empleaban la lista de Escamilla (la más arraigada en el país,<sup>2</sup> pero que no cuenta con actualizaciones) y que posteriormente introdujeron el uso de la *Lista de encabezamientos de materia para bibliotecas* o de otro u otros vocabularios controlados, sin que mediara un proceso de cambio o asimilación de su estructura, lo que provocó la coexistencia de dos o más estructuras formales en un mismo catálogo<sup>3</sup> y, por lo tanto, que hubiera en algunos casos dos formas diferentes para representar un mismo concepto; esto dificulta al usuario el acceso eficiente a la información y al catalogador la toma de decisiones en el proceso técnico de la misma.

Ante este panorama la solución del conflicto pasa por dos procesos: 1) la documentación de criterios que normalizan la elección y asignación de EM y 2) el control de autoridades de materia.

La situación descrita indica que vale la pena hacer el esfuerzo de generar un instrumento que auxilie a la normalización del proceso de asignación de EM, en virtud de la utilidad que podría aportar tanto al trabajo profesional como a la capacitación o formación profesional de los catalogadores. Sin embargo, en vista de que nos enfrentamos a un trabajo de enormes proporciones, parece lo más prudente iniciar con las pautas de normalización de uno de los componentes que integran un EM: en este caso en particular se ha elegido la normalización de nombres geográficos mediante la presentación de un manual porque, al ser encabezamientos que de manera habitual no se localizarán en listas de EM, es decir que tendrán que ser construidos por el catalogador en mayor medida que otros elementos, se hace entonces necesaria una guía que garantice su uniformidad y consistencia.

El tratamiento de nombres geográficos como elementos del lenguaje que deben ser controlados se presenta con características especiales: es relativamente sencillo identificar cuando una obra presenta

---

2 A. Quijano Solís, P.M. Moreno Jiménez y R. Figueroa Servín, "Automated authority files of Spanish-language subject headings", en *Cataloging & classification quarterly*, 2000, v. 19, no. 1-2: 212.

3 No hay más que observar los índices de los catálogos en línea de las bibliotecas más representativas del país para notar la falta de consistencia en la asignación de encabezamientos de materia; dicha falta de consistencia es tanto formal como léxica.

un tratamiento geográfico y, a diferencia del análisis que se hace para definir otro tipo de elementos de un EM, los nombres geográficos no involucran demasiado tiempo para su determinación como parte del contenido de una obra (salvo en el caso en que no esté definido en las partes más evidentes, como el título o la tabla de contenido, y que entonces deba profundizarse la inspección en el resto de la obra); por lo tanto hay una menor complicación en el paso que va desde el momento en que se revisa la obra a que se determina que presenta un tratamiento geográfico. No obstante, el verdadero problema se presenta en el momento de traducir ese análisis a la estructura de un lenguaje controlado (documental) pues si dichos nombres no aparecerán como parte integrante del vocabulario de una lista de EM,<sup>4</sup> de ahí entonces la necesidad de contar con lineamientos para establecer la estructura que deben mantener.

## DEFINICIÓN

Los nombres geográficos, como cualquier signo lingüístico, encuentran su origen en la necesidad del hombre de referirse a la realidad que le rodea, de poder evocarla aun cuando no está presente. En términos generales un nombre geográfico es lingüística o gramaticalmente un nombre propio,<sup>5</sup> es decir, un signo lingüístico que identifica entidades únicas en su tipo. La disciplina encargada del estudio de los nombres propios es la onomástica, la cual a su vez se divide en dos ramas: la antroponomía, que se encarga del estudio de los nombres de personas, y la toponimia, que es el estudio taxonómico de los nombres de lugares o topónimos, su origen y significado.<sup>6</sup> El propósito principal de un

---

4 J. Jiménez Pelayo y R. García Blanco, *El catálogo de autoridades: creación y gestión en unidades documentales*, Gijón, España: Trea, 2002: 513.

5 I. Guzmán Betancourt, "La toponimia : introducción al estudio de los nombres de lugar", conferencia pronunciada el 7 de marzo de 1983 en las instalaciones del Centro de Estudios Científicos y Tecnológicos. Citado por Víctor M. Ortiz Martínez, *Manual metodológico para la investigación documental de nombres geográficos*, México: INEGI, 1989: 29.

6 También se le llama toponimia al conjunto de nombres geográficos de un territorio determinado.

nombre geográfico es proporcionar información acerca de accidentes naturales o espacios creados por el hombre sobre la superficie de la tierra, así como de ubicaciones en el lecho marino y superficies planetarias.<sup>7</sup> Los nombres geográficos representan una muy amplia fuente de información, en especial para la lingüística, la historia y la geografía, pero también son importantes para la antropología, la ecología, la botánica, la zoología, entre otras disciplinas. A través de su investigación y estudio informan acerca de la extensión territorial de una lengua desaparecida, reducida, o modificada; informan también acerca de sus patrones de formación y de sus contactos fundamentales con otras lenguas. Son claves en la hipótesis y corroboración de algunos hechos históricos, como incursiones, desplazamientos y asentamientos de determinados grupos humanos. Indirectamente informan de las costumbres, economía, filosofía, tradiciones, religión, política, guerras y sucesos de los pueblos.<sup>8</sup>

Orth ofrece la definición más detallada de nombre geográfico:

Los nombres geográficos son sonidos convencionales o símbolos visuales representativos que se originan y son empleados en el lenguaje diario para referirse a lugares, accidentes o áreas de la superficie terrestre y constituyen un sistema primario de referencia empleado por la gente en todo el mundo. Su uso ilustra una facultad humana única para abstraer y clasificar entidades geográficas en el continuum espacial que percibimos como realidad; y entonces identifican ítems específicos, dentro de esa clasificación, a partir de nombres propios. El acto de nombrar es una parte de nuestra necesidad intelectual para clasificar y para referir. Sin embargo, está en la esencia del lenguaje el ser flexible, y por lo tanto el permitir la variación en las formas de los nombres y su aplicación a entidades geográficas. Aunque la naturaleza de esta variación

---

7 R. R. Randall, *Place names: how they define the world - and more*, Lanham, MD: Scarecrow, 2001: 27.

8 V.M. Ortiz Martínez, *Manual metodológico para la investigación documental de nombres geográficos*, México: INEGI, 1989: 40.

es compleja, normalmente no causa mayores problemas en el lenguaje hablado en la vida diaria. Los expertos del lenguaje dicen que el significado y aplicación de los nombres, así como el de otras palabras, no tiene necesariamente que ser preciso pues son las oraciones y no las palabras la esencia del acto del habla. Obviamente esto deja de ser cierto cuando los nombres son removidos de su contexto original y sus formas escritas son empleadas de manera aislada en mapas, señales o en registros legales.<sup>9</sup>

Esta característica de los nombres geográficos (en realidad de cualquier signo lingüístico) de falta de univocidad<sup>10</sup> es lo que tiende a propiciar la dispersión de información y la ambigüedad en la comunicación, particularmente cuando se les extrae de o no se les usa en su contexto natural, y es esto lo que genera la necesidad de su normalización. En la vida cotidiana, a través del lenguaje hablado, el empleo de nombres geográficos no precisa una intervención normativa rigurosa porque a partir del contexto de uso el referente aparece con toda claridad para el emisor y para el receptor, además de que la interacción inmediata contribuye a la aclaración de cualquier laguna en la comprensión del mensaje que se transmite. Pero cuando la comunicación tiene que efectuarse entre personas o comunidades alejadas en el tiempo y en el espacio, el contexto es mínimo o inexistente, por lo que se hace preciso encontrar un mecanismo que garantice la efectividad de la comunicación. Este mecanismo se constituye a partir de varias acciones encaminadas al control de las diferentes formas por

---

9 D. J. Orth, "Organization and functions of a National Geographical Names Standardization Programme: a manual", en *World Cartography*, 1990, v. 21: 12.

10 En la teoría de la clasificación la univocidad es "la cualidad que debe presentar todo concepto (sea un término, un descriptor, una palabra clave o un encabezamiento de materia), y que consiste en la atribución de un único sentido o un único significado al mismo dentro de un determinado contexto. La univocidad facilita la desambiguación de los conceptos y el control adecuado de las relaciones jerárquicas, de sinonimia y cuasi-sinonimia". Véase M. Barité, *Diccionario de organización y representación del conocimiento*, [citado (7 Septiembre 2005)], disponible en <[http://www.eubca.edu.uy/diccionario/letra\\_u.htm](http://www.eubca.edu.uy/diccionario/letra_u.htm)>.

las que se reconoce una entidad geográfica, siendo la más importante, hacia la que se orienta el resto de ellas, la normalización.

La normalización es importante en la medida que establece una forma de tratar la información de manera eficiente, evita la dispersión de la misma y evita también la duplicación de esfuerzos en diferentes organismos que, aunque desarrollen actividades diferentes, tienen que ver con el tratamiento de entidades geográficas y por tanto tienen que referirse a ellas de una misma manera.

En el ámbito bibliográfico un nombre geográfico es el nombre que más comúnmente se usa para identificar una ubicación, un accidente o un área geográficos, el cual es seleccionado por el catalogador para establecer la forma de un asiento (que no necesariamente coincide con el nombre oficial).<sup>11</sup> En el campo propiamente dicho de la asignación de encabezamientos de materia los nombres geográficos son materias que designan contenidos relativos a lugares y que requieren directrices especiales para su normalización: están sometidos a las mismas variaciones y ambigüedades que los nombres de personas o entidades corporativas: problemas de variaciones lingüísticas (New York, Nueva York), de escritura (ortografía: México, Méjico; y transliteración: Beijing y Pekin), de extensión (Santiago y Santiago de Compostela) y de redacción (orden directo o inversión de términos: Mar Adriático vs. Adriático, Mar).<sup>12</sup>

El análisis temático para la identificación del tratamiento geográfico en una obra sigue el mismo proceso que para los encabezamientos de materia generales pero, una vez definido ese tratamiento geográfico, no se recurre a listas de encabezamientos de materia sino que deberá consultarse otras fuentes de información: los sistemas de información geográfica (SIG), los nomenclátors (*gazeteers*), diccionarios, enciclopedias, tesauros, etcétera, para retomar de ahí los nombres autorizados. Posteriormente estos nombres tendrán que adaptarse al lenguaje de indización temática. Usar un lenguaje de indización equi-

---

11 J. M. Reitz, *Dictionary for library and information science*, Westport, CT: Libraries Unlimited, 2004: 310.

12 Cfr. J. Jiménez Pelayo y R. García Blanco, *El catálogo de autoridades: creación y gestión en unidades documentales*, Gijón, España: Trea, 2002: 513

vale a traducir a ese lenguaje los textos que se analizan<sup>13</sup> y por lo tanto está sujeto a las mismas dificultades inherentes a cualquier proceso de traducción: la necesidad de una visión general del vocabulario de ambos lenguajes y del significado de cada unidad léxica del lenguaje de origen y del lenguaje de destino. En este caso están implicados, por un lado el lenguaje natural de los documentos que se analizan, el lenguaje normalizado, pero del contexto de los sistemas de información geográfica y de otras fuentes de referencia (como lenguajes de origen) y el lenguaje de representación bibliográfica de nombres geográficos. Si se da como un hecho que los lenguajes de origen escapan a la intervención directa del catalogador (es decir, que le llegan como algo ya dado), de ahí entonces la necesidad de contar con un lenguaje de indización que favorezca los objetivos de un sistema de información bibliográfica, y de ahí, también, la importancia de desarrollar ese lenguaje con claridad y precisión.<sup>14</sup>

## TIPOS DE LUGARES GEOGRÁFICOS

Las entidades geográficas a las que normalmente se les asigna un nombre propio pueden dividirse en un cierto número de categorías. Las que con mayor frecuencia son tratadas con propósitos de normalización de sus nombres son:

- a) accidentes del paisaje natural (ej. corrientes de agua, valles, colinas, lagos, bahías, montañas.
- b) sitios y localidades poblados (ej. ciudades, pueblos, estaciones de tren).
- c) divisiones civiles y políticas (ej. estados, condados, provincias, departamentos).

---

13 R. Fugmann, "The complementarity of natural and indexing languages", en Lois Mai Chan, Phyllis A. Richmond, Elaine Svenonius (eds.), *Theory of subject analysis: a sourcebook*, Littleton, CO: Libraries Unlimited, 1985: 390-402.

14 Se puede decir que para conformar un lenguaje no basta con elaborar un vocabulario de las palabras o términos que lo constituyen, es necesario además establecer las reglas que rigen su combinación y uso.



- d) áreas administrativas (ej. parques, reservas forestales, sitios históricos).
- e) rutas de transporte (ej. calles, carreteras, autopistas, veredas).
- f) construcciones (ej. presas, edificaciones, monumentos).<sup>15</sup>

En el ámbito bibliográfico la tipología establecida para nombres geográficos considera primeramente dos grandes categorías: nombres jurisdiccionales y nombres no jurisdiccionales. Los de la primera categoría se establecen de acuerdo con las convenciones de la catalogación descriptiva; cuando este tipo de nombres geográficos se asignan como encabezamientos de materia, es decir como asientos de autor corporativo (campo 610 del formato MARC y en el campo 110 del formato MARC de autoridades) deben seguir las disposiciones de un código de catalogación descriptiva. Este tipo de nombres geográficos representa a países o a divisiones políticas o administrativas dentro de países (provincias, estados, municipios, ciudades y equivalentes) y en las Reglas de Catalogación Angloamericanas siguen las disposiciones de los capítulos 23 (Nombres geográficos) y 24 (Encabezamientos de entidades corporativas).

En cambio, en la segunda categoría se incluyen los nombres geográficos que son objeto exclusivo de la catalogación por materia (codificados en el campo 651 y en los subcampos \$z del formato MARC bibliográfico y en el campo 151 del formato MARC de autoridades). La investigación que se presenta se limitó al tratamiento de los nombres de la segunda categoría y a los de la primera cuando no son autores corporativos.<sup>16</sup> De esta manera más específica, las tipologías establecidas en diferentes sistemas de indización analizados dividen a los nombres de la siguiente manera:

---

15 D. J. Orth, "Organization and functions of a National Geographical Names Standardization Programme: a manual", en *World Cartography*, 1990, v. 21: 16.

16 De acuerdo con los *Principles underlying subject headings languages*, de la IFLA, debe haber coincidencia entre la forma que guarda un nombre en 610 y en 651, pero la caracterización es diferente: en el primer caso se habla de un gobierno, mientras en el segundo se alude a una entidad como espacio físico-cultural.

**Subject cataloging manual : subject headings<sup>17</sup>**

- Sitios arqueológicos, históricos, etc.
- Regiones geográficas
- Canales
- Presas
- Ciudades desaparecidas (anteriores a 1500)
- Granjas, ranchos, jardines
- Bosques, praderas, etc.
- Accidentes geográficos (por ejemplo, cuevas, desiertos, islas sin jurisdicción, lagos, montañas, corrientes oceánicas, planicies, ríos, mares, llanuras, sitios geográficos submarinos)
- Depresiones y formaciones geológicas, etcétera
- Minas
- Parques, reservas, refugios, áreas recreativas
- Embalses
- Carreteras, calles, caminos
- Valles

**Jiménez Pelayo y García Blanco<sup>18</sup>**

- Accidentes geográficos
- Continentes y grandes agrupaciones territoriales
- Países, regiones, provincias, ciudades y pueblos
- Comarcas, parajes y zonas de interés turístico, económico, etc.
- Nombres geográficos formulados por siglas o introducidos por artículo
- Nombres de delimitaciones geográficas por su situación según los puntos cardinales
- Espacios naturales protegidos, yacimientos arqueológicos y cuevas
- Nombres geográficos de espacios urbanos: barrios, jardines, plazas y calles

---

17 Library of Congress. Office for *Subject Cataloging Policy*, *Subject cataloging manual : subject headings*, 5th ed., Washington, DC: Cataloging Distribution Service, Library of Congress, 1991: H690.

18 J. Jiménez Pelayo y R. García Blanco, *El catálogo de autoridades: creación y gestión en unidades documentales*, Gijón, España: Trea, 2002: 515. El sistema que presentan los autores está expresado en Biblioteca Nacional (España), *Encabezamientos de materia: normativa para su redacción*, Madrid, España : Ministerio de Cultura, 1991, aunque ellos lo interpretan y sintetizan a partir de su propia experiencia.

- **Maroun y Neves<sup>19</sup>**
- Nombres geográficos con categoría administrativa
    - Brasileños
      - Unidades de la federación
      - Municipios
      - Lugares dentro de ciudades
      - Comarcas, fronteras, parroquias
    - Extranjeros
      - Países, posesiones territoriales, dependencias, etc.
      - Estados, provincias, ciudades
      - Lugares dentro de ciudades
      - Regiones
  - Nombres geográficos sin categoría administrativa
    - Ríos
    - Accidentes geográficos
    - Islas
    - Áreas de levantamientos topográficos
    - Cuencas sedimentarias
    - Parques, reservas, monumentos nacionales
    - Calles
    - Ciudades desaparecidas y sitios arqueológicos

En los sistemas de indización temática los tipos de lugares identificados obedecen, como se ha visto, a una diferenciación entre nombres jurisdiccionales y nombres no jurisdiccionales (o con categoría administrativa y sin categoría administrativa). También, aunque en esta categorización general no es evidente en todos los casos, establecen un trato especial para los nombres locales de los países en donde se desarrollan estos sistemas y los nombres extranjeros.<sup>20</sup> Las categorizaciones son importantes porque establecen la primera pauta temática para decidir cuándo un nombre de lugar es considerado como materia y cuándo no.

---

19 M.C. dos Santos Bou Maroun y M.L.T. Pacheco Neves, *Nomes geográficos: normas para indexação*, Río de Janeiro: Fundação Instituto Brasileiro de Geografia e Estatística, Ministério de Planejamento e Orçamento, 1996.

20 Esta última característica resulta relevante para la comparación entre diferentes lenguajes de indización. Cada sistema le confiere un trato especial al los nombres del país en el que se establecen y un trato diferenciado para los nombres de otros países.

## DESARROLLO DE LA INVESTIGACIÓN FUENTE

Para lograr la normalización de la situación descrita líneas arriba se desarrolló una investigación<sup>21</sup> que aportara las directrices bajo las cuales controlar en México los nombres geográficos y su integración como parte de EM. Los ejes sobre los que se condujo esta investigación se desarrollaron en dos dimensiones: 1) los parámetros para la elección o determinación de la forma de un nombre geográfico y 2) la redacción del encabezamiento. El primero gira en torno a la decisión del nombre que se integra como elemento geográfico de un EM (principalmente las fuentes para establecer un nombre único entre las variantes que puedan existir) y el segundo se dirige a establecer su estructura en la representación bibliográfica (es decir, una vez decidido el nombre, su adecuación a un lenguaje documental).

La investigación se inició con el análisis de la relevancia de los nombres geográficos como entidades lingüísticas y culturales de referencia espacial y el reconocimiento de la importancia estratégica de su normalización en diferentes ámbitos de la actividad humana, más allá de los objetivos de nuestra especialidad.<sup>22</sup> En esta parte se presenta además una exposición sobre la necesidad de la normalización de nombres geográficos en general y la generación de fuentes que sistematizan los repertorios de nombres, principalmente mediante los sistemas de información geográfica (SIG); se explica brevemente en qué estado se encuentra esta función en México.

Posteriormente se aborda el análisis de la estructura de un nombre geográfico para su redacción e integración como parte de un encabezamiento de materia. Se parte de la definición de nombre geográfico en el ámbito bibliográfico y la tipología establecida para su control,

---

21 La investigación original fue presentada como la tesis para obtener el título de licenciado en bibliotecología en la Universidad Nacional Autónoma de México. Véase N. Ángeles Escobar, *Tratamiento de nombres geográficos como parte de encabezamientos de materia*, México: El autor, 2006.

22 Existen coincidencias entre los objetivos y principios de la normalización, tal como los establece el Grupo de Expertos en Nombres Geográficos de las Naciones Unidas, con los principios de normalización bibliográfica, aunque a la larga su uso y necesidad de representación resulte un poco diferente.

para llegar a la revisión de los principios que fundamentan su creación (con base en los *Principles underlying subject headings languages*<sup>23</sup> que establece la IFLA para EM generales). Se efectuó después un análisis comparado de estructuras de redacción de nombres geográficos en diferentes sistemas de encabezamientos de materia. Los objetivos de este análisis fueron: 1. fundamentar las decisiones y la aplicación que se establecerá posteriormente; 2. revisar y, en lo posible, mantener estructuras generales presentes de manera consistente en los diferentes sistemas; 3. mostrar patrones de representación de nombres geográficos en los diferentes sistemas revisados, de modo que al encontrarse con uno, éste se reconozca y pueda traducirse a la estructura que se propone. De esta manera, los diferentes sistemas no dejan de ser fuentes de consulta válidas.

Del análisis mencionado anteriormente resulta como producto final el *Manual para la asignación de encabezamientos de materia que contienen nombres geográficos*. Este manual establece la forma que tomarán los nombres geográficos como parte de un encabezamiento de materia válido. Determina además su representación como encabezamientos principales, como calificadores<sup>24</sup> y como subdivisiones, además de plasmar la política a seguir para el caso de nombres geográficos de México. Sus objetivos particulares son:

- documentar el control bibliográfico de nombres geográficos como parte de encabezamientos de materia para su uso en sistemas de información bibliográfica.
- conformar una fuente de consulta para completar el proceso de análisis temático con los lineamientos para la investigación y representación de encabezamientos de materia que contienen nombres geográficos.

---

23 M. J. Lopes y B.J. Julianne (eds.), *Principles underlying subject headings languages* (SHLs), Munich: K. G. Saur, 1999.

24 Un calificador es un término o términos que se agrega(n) entre paréntesis a continuación de un EM, para diferenciarlo de otros con la misma escritura o para eliminar ambigüedades.

- presentar las subdivisiones temáticas válidas asociadas a un encabezamiento de materia principal de tipo geográfico. Para cumplir este objetivo se incluyó un índice inglés-español, español-inglés (con algunas subdivisiones en francés cuando estuvieron disponibles) de subdivisiones flotantes asociadas a nombres geográficos, el cual se construyó a partir del *Subject cataloging manual : subject headings*, 5th ed., de la *Library of Congress*, y su traducción a partir de diferentes fuentes en español: los *Subepígrafes de materia para bibliotecas* de la Biblioteca Nacional de Perú<sup>25</sup>; el *Catálogo CSIC de autoridades (en línea)*<sup>26</sup> del Consejo Superior de Investigaciones Científicas de España; la *Lista de encabezamientos de materia para bibliotecas* del Banco de la República de Colombia; y la *Lista de encabezamientos de materia* de Gloria Escamilla.
- presentar un repertorio de fuentes asociadas a la investigación para la normalización bibliográfica de nombres geográficos contenidos en encabezamientos de materia.
- En cuanto a su forma, mantener una fuente de consulta actualizable para las decisiones de tipo técnico en la asignación de encabezamientos de materia.

## PROPUESTA

A partir de la situación descrita, se propone desarrollar un proyecto que integre en la práctica los esfuerzos normativos expresados en la investigación precedente, considerando las siguientes etapas:

### 1. Difusión

- a. Revisión y publicación del manual
- b. Retroalimentación con los catalogadores

---

25 Biblioteca Nacional del Perú. Centro Bibliográfico Nacional, Subepígrafes de materia para bibliotecas, 2a ed., Lima : Biblioteca Nacional, 2003

26 Consejo Superior de Investigaciones Científicas (España). Red de Bibliotecas, Catálogo CSIC autoridades [en línea], [Madrid : Consejo Superior de Investigaciones Científicas], [Citado (Octubre 2005)], disponible en <[http://aleph.csic.es/F?func=file&file\\_name=find-b&local\\_base=MAD10](http://aleph.csic.es/F?func=file&file_name=find-b&local_base=MAD10)>.

2. Uso

- a. Creación de registros bibliográficos y de autoridad
- b. Discusión e integración, del lenguaje documental propuesto, como parte de las políticas del Proyecto de Control de Autoridades de Encabezamientos de Materia.

3. Consolidación del lenguaje de EM que propone el manual y mantenimiento de actualizaciones del mismo.

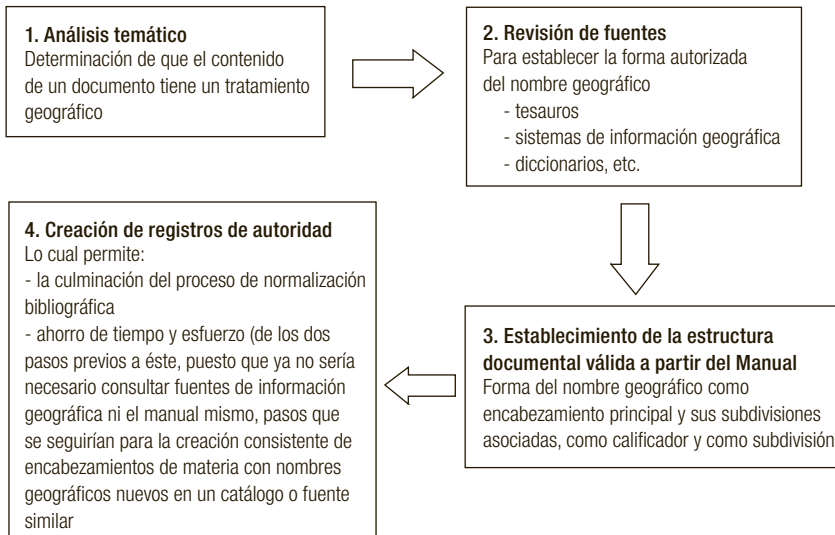
Como parte del primer punto se tiene proyectada la publicación de la investigación fuente por la Biblioteca Daniel Cosío Villegas de El Colegio de México. En estos momentos se encuentra en proceso de revisión. A partir de la publicación se espera propiciar un ambiente de comunicación y captación de impresiones sobre el manual, en asuntos como: comentarios de casos y experiencias, documentación sobre fuentes de consulta de nombres geográficos, etcétera.

Del punto 2 se obtendrá la verificación práctica del proceso de normalización de acuerdo con la situación en México sobre los encabezamientos de materia mencionada al principio de este reporte. También en este punto se incluye la intención de integrar formalmente los postulados de la investigación como parte del Proyecto Colectivo de Control de Autoridad de Materia, en el que participan representantes de las bibliotecas de la Universidad Nacional Autónoma de México, de la Universidad Iberoamericana, de El Colegio de México y del Instituto Tecnológico de Estudios Superiores de Monterrey, quedando pendiente por ahora la presentación de esta propuesta ante los representantes del proyecto y la gestión de la adopción del manual.

## CONCLUSIONES Y EXPECTATIVAS

Como para cualquier fuente que establece criterios de normalización, se espera que el instrumento que se presenta establezca las pautas de decisión y acción acerca de un aspecto que no ha sido suficientemente abordado en nuestro medio. En él se ha intentado cubrir el espectro de la toma de decisiones que acompaña al control de nombres geográficos que son parte de encabezamientos de materia. Aun así no

puede entenderse este esfuerzo como algo definitivo ni exhaustivo en su más alto nivel. Sólo el uso del Manual será el indicador de que las decisiones adoptadas son las más adecuadas para la representación, organización, recuperación y acceso a la información que contienen las entidades geográficas en un sistema bibliográfico. El mejor de sus destinos será que se mantenga una interacción de ida y vuelta entre el manual como instrumento normativo y la práctica de asignación de encabezamientos de materia y que los hallazgos de esta última lo completen o lo corrijan. El proceso en el que se integra el trabajo desarrollado se puede ver en el siguiente esquema:



En la investigación presentada se atendió principalmente a los pasos 2 y 3 del esquema, los que establecen una forma de trabajo que busca consolidar un proceso técnico de calidad. Los elementos que se han tenido en cuenta para su elaboración son:

- la preeminencia del contexto local (las bibliotecas mexicanas) en la concepción general de control de información. Esto involucra la revisión, adaptación y documentación consciente



y consistente de prácticas normativas de diferente origen y naturaleza. Algo que no admita ni la aceptación total de una norma externa ni su descalificación absoluta.

b). la lengua española como base del análisis temático.

c). la convergencia entre el lenguaje del usuario y el de presentación de encabezamientos de materia.

## BIBLIOGRAFÍA

*Archivo histórico de localidades* [en línea], México: Instituto Nacional de Estadística, Geografía e Informática, Sistema Nacional Estadístico y de información Geográfica, [2007], [citado (Febrero 2007)], Disponible en <<http://mapserver.inegi.gob.mx/dsis/ahl2003/index.html?c=424>>.

Bélair, J.A. ; Bourdon, F. ; Mingam, M., *The Répertoire de vedettes-matière and RAMEAU: Two Indexing Languages in French: A Necessary Luxury?* [en línea]. Presentado en el World Library and Information Congress: 71th IFLA General Conference and Council : "Libraries - A voyage of discovery", August 14th - 18th 2005, Oslo, Norway. [citado (Enero 2006)], disponible en <[http://www.ifla.org/IV/ifla71/papers/145e-Belair\\_Bourdon\\_Mingam.pdf](http://www.ifla.org/IV/ifla71/papers/145e-Belair_Bourdon_Mingam.pdf)>.

Bermello Crespo, L., "Tendencias de desarrollo de los lenguajes de encabezamientos de materia en las bibliotecas", en *ACIMED*, 2001, 9, no. 2: 109-120.

Codina, L., "Modelo conceptual de un sistema de información documental", en *Revista de documentación científica*, 1994, 17, no. 4: 440-441.

Drabenstott, K. M., Facilitating geographic subdivisions assignment in subject headings: need for indirect forms of geographic names in authority files, en *Library Resources & Technical Services*, Oct. 1992, 36: 411-425.

***Primer Simposio Internacional sobre Organización...***

Fugmann, R., "The complementarity of natural and indexing languages", en Chan, L. M.; Richmond, P. A.; Svenonius, E. (eds.), *Theory of subject analysis: a sourcebook*, Littleton, Colorado : Libraries Unlimited, 1985: 390-402.

Grupo de Expertos de las Naciones Unidas en Nombres Geográficos, *Uso uniforme de topónimos*, Nueva York : Secretaría del Grupo de Expertos de las Naciones Unidas en Nombres Geográficos, Departamento de Asuntos Económicos y sociales, División de Estadística, 2001. Tríptico informativo.

*Introduction to geographic names (document in progress)* [en línea]. New Orleans : Tulane University, revised draft 18 May 2003. [<http://www.tulane.edu/~techserv/geogintrod.html> ; consultado en octubre de 2005].

Jiménez Pelayo, J. y García Blanco, R., *El catálogo de autoridades : creación y gestión en unidades documentales*, Gijón, España : Trea, 2002.

Kadmon, N. (ed.), *Glossary of terms for the standardization of geographical names = Glosario de términos para la normalización de los nombres geográficos*. New York : United Nations, 2002.

Kerfoot, H., "Geographical names : some current issues in the context of the United Nations", en *The Cartographic Journal*, 2004, 41, no. 2: 89-94.

Library of Congress, *Free-floating subdivisions : an alphabetical index*, 16th ed., Washington, D.C. : Cataloging Distribution Service, Library of Congress, 2004.

Library of Congress. Office for Subject Cataloging Policy, *Subject cataloging manual : subject headings*, 5th ed., Washington, D.C. : Cataloging Distribution Service, Library of Congress, 1991- .

Orth, D.J., "Organization and functions of a National Geographical Names standarization Programme : a manual", en *World Cartography*, 1990, 21: 11-40

Lopes, M.J. y Julianne, B. J. (eds.), *Principles underlying subject headings languages (SHLs)*, Munich : K. G. Saur, 1999.

Proyecto Colectivo de Control de Autoridad de Materia, Comité técnico, *Manual de políticas para la elaboración de registros de autoridad de materia*, México : El Comité, 2003.

Quijano Solís, A.; Moreno Jiménez, P.M.; Figueroa Servín, R., "Automated authority files of Spanish-language subject headings", en *Cataloging & classification quarterly*, 2000, 19, no. 1-2: 209-223.

Randall, R.R., *Place names : how they define the world - and more*, Lanham, Maryland : Scarecrow, 2001.



SECCIÓN VIII  
Normas y uso del lenguaje  
en un ambiente digital



# Las tecnologías del lenguaje en el manejo de información y en la generación del conocimiento

GERARDO SIERRA MARTÍNEZ  
*Instituto de Ingeniería,  
Universidad Nacional Autónoma de México*

## EL MANEJO DE INFORMACIÓN

**E**n nuestra sociedad actual, la información y el conocimiento son componentes esenciales para el progreso económico y para el bienestar personal. A fin de lograr la generación de conocimiento para tener las respuestas más adecuadas ante las diversas circunstancias que se van presentando, las personas necesitan disponer de información oportuna y verídica.

Con la creación de una política inteligente de acceso a la información, se logra un avance importante para una interacción eficiente entre los sectores industriales y estatales —con necesidades concretas de conocimiento—, los sectores académicos —productores y gestores del mismo— y la sociedad. Al establecer un sistema de procesamiento eficiente de intercambio de conocimiento entre las partes mencionadas, se pueden realizar acciones concretas con resultados óptimos que ayuden realmente a los diferentes estratos de la población, por ejemplo pequeñas y medianas empresas (PyMEs) con objetivos de comercio internacional que coadyuvan al desarrollo económico y sustentable del país.

Existe, en México, una gama considerable de empresas que, de una u otra forma, tienen que tratar con la gestión de información. Tan

sólo en el D.F., hay alrededor de 116,000 empresas registradas<sup>1</sup> que requieren de un tratamiento a mediana o gran escala para actualizar su información y gestionarla de manera que puedan tener una mejor competencia en el mercado y logren generar un mayor número de divisas. Igualmente, los institutos y entidades requieren de una gestión y una explotación de calidad de información eficiente, por lo que es menester tomarlos en cuenta para que México siga en su ruta de avance y progreso.

## INFORMACIÓN, CONOCIMIENTO Y LENGUAJE

Si bien la información se ha tomado muchas veces como equivalente de saber o de conocimiento, hay varias diferencias entre estos dos conceptos que debemos considerar. Entre las características principales de la información podemos mencionar las siguientes: se trata de un bien de naturaleza intelectual, inmaterial, aunque se registra y presenta en soportes físicos. No es escasa; cada vez es más abundante y su uso no es excluyente. La pueden utilizar varias personas a la vez, salvo cuando su valor está precisamente en la privacidad y exclusividad. Por otro lado, respecto al conocimiento, decimos que éste es fruto de las asociaciones que el cerebro es capaz de hacer a partir del análisis de información almacenada. Además, el conocimiento se elabora para dar respuesta a alguna cuestión de interés particular; por tanto, el conocimiento siempre supone más que la información de la que se ha partido para construirlo; tiene una funcionalidad.

Pese a sus diferencias, existe una relación indisoluble entre información, conocimiento y lenguaje; tanto la información como el conocimiento se expresan y transmiten gracias al lenguaje (oral, escrito, gesticular, etc.) La “sociedad de la información” [1] se define en relación a mecanismos como la producción, el tratamiento y la distribución de la información y, por ende, está directamente relacionada con el lenguaje. Esta sociedad exige, desde un punto de vista técnico, la infraestructura necesaria para la utilización de la información en

---

1 Puede consultarse la página del Gobierno de México: [http://www.gob.mx/wb/egobierno/egob\\_conozca\\_la\\_concentracion\\_geografica\\_de\\_empres](http://www.gob.mx/wb/egobierno/egob_conozca_la_concentracion_geografica_de_empres)



todos los ámbitos de la economía y de la vida social. Es gracias a esta infraestructura que es posible encontrar información pertinente en el entorno diario que, en conjunto con la información en la memoria individual o colectiva, permite la construcción continua de conocimientos útiles para afrontar los retos cotidianos.

En la sociedad actual, que pone a nuestro alcance cantidades ingentes de información de todo tipo, una de las exigencias básicas a la que todo ciudadano se enfrenta es el aprovechamiento de esta información para construir conocimientos. Las actividades cotidianas involucradas en la construcción de conocimientos son las siguientes:

- Buscar información en diversas fuentes, tales como bibliotecas, diarios, enciclopedias, Internet, revistas, etc.
- Analizar la información que nos permita relacionarla, valorarla, seleccionarla y ordenarla en función de nuestras necesidades.

Estas actividades, reunidas bajo el concepto de “gestión de la información”, son indispensables para la construcción de nuevos conocimientos. Las sociedades de la información basan su valor en su capacidad para obtener, clasificar, sistematizar, sintetizar, difundir y aprovechar la información. Toda persona en esta nueva sociedad es productora de información y una adecuada gestión de la misma permitirá su aprovechamiento por cualquier otra persona.

Es importante destacar que, dada la complejidad de estas tareas, es prácticamente imposible realizarlas de forma manual; de ahí la necesidad de apoyarse en métodos automáticos.

El conocimiento ha estado presente en sistemas que operan de manera particular frente a cada problema; a diferencia de la información, la cual se emplea en sistemas rígidos cuya acción es general para los problemas que resuelven [2]. La operación central que se realiza en el conocimiento es la inferencia, esto es, se parte de ejemplos para generalizar el conocimiento.

Las formas en que se representa el conocimiento son variadas; principalmente relaciones, reglas, grafos, y jerarquías [3]. Por consecuencia, se requiere de especialistas, científicos computacionales

para llevar a cabo la realización e implementación de las diferentes representaciones de conocimiento en ambientes artificiales, es decir mediante sistemas de cómputo.

Un subconjunto de este tipo de sistemas, referidos como sistemas basados en conocimiento lingüístico (SBCL), es el formado por aquellos que utilizan conocimiento recopilado en léxicos, reglas morfosintácticas, heurísticas semánticas, etc. [4] La siguiente sección expone la importancia y beneficios de dichos sistemas, todo ello englobado en las llamadas tecnologías del lenguaje.

## TECNOLOGÍAS DEL LENGUAJE

El lenguaje es el medio principal a través del cual los humanos intercambian información. Ningún otro sistema de información es tan eficaz para compartir ideas, transmitir conocimientos, expresar emociones y crear, en definitiva, lo que se conoce como cultura.

Las llamadas tecnologías del lenguaje constituyen la base de la ingeniería lingüística, entendida como la aplicación de los conocimientos sobre la lengua en el desarrollo de sistemas informáticos que puedan reconocer, comprender, interpretar y generar lenguaje humano en todas sus formas. Por lo anterior, el primer punto a destacar es que el desarrollo de las tecnologías del lenguaje requiere de equipos multidisciplinarios de investigación y desarrollo, en los que participan científicos en computación, inteligencia artificial, lingüistas, terminólogos e ingenieros, entre otros muchos especialistas.

En la actual sociedad del conocimiento, el uso de herramientas y servicios como portales de voz, correo electrónico leído a través del teléfono celular, dictado de documentos a la computadora, consulta de páginas en Internet en diferentes lenguas mediante traductores automáticos, e incluso tareas tan comunes como el uso de correctores automáticos y gramaticales durante la elaboración de documentos, se ha convertido en un conjunto de actividades cotidianas realizadas por millones de personas en todo el mundo. Sin embargo, no existe conocimiento mínimo de las necesidades económicas, técnicas y culturales requeridas para el desarrollo y mejora de estas herramientas, producto de la investigación y desarrollo en materia de tecnologías del lenguaje.

Las tecnologías del lenguaje son todas aquellas que se integran en aplicaciones informáticas para permitir el tratamiento de textos escritos —como el caso de la traducción automática o la búsqueda y recuperación de información—, o el procesamiento del habla —requerido para el dictado automático de documentos o la lectura de un correo electrónico en voz alta por la computadora—. Se trata, en conjunto, de tecnologías que hacen posible la creación de herramientas pensadas para auxiliarnos en el uso de las computadoras —y en general de las máquinas— sin renunciar por ello al uso habitual del lenguaje como medio de interacción y de intercambio de información [5].

Estas tecnologías abren un nuevo capítulo en la historia de la Sociedad de la Información; pueden enriquecer nuestro entorno informático con muchas de las funciones expresivas naturales propias del lenguaje humano, que hasta ahora no habían estado presentes en la interacción entre personas y máquinas. Combinando el potencial tecnológico con la profundidad y la flexibilidad del procesamiento del lenguaje humano, las aplicaciones y los servicios que ofrecen las tecnologías de la lengua pueden mejorar radicalmente la eficacia y la facilidad de los mecanismos y sistemas usados en las actividades cotidianas relacionadas con la información y la comunicación. Esto incluye el ocio y los negocios, la administración y la educación, los servicios y la vida diaria. Algunas aplicaciones típicas de las tecnologías del lenguaje en la actualidad son:

- Servicios de traducción automática, a los que se puede acceder a través de una página web o del teléfono móvil.
- Sistemas de dictado automático, que permiten transcribir cartas, informes o notas en diversos campos profesionales.
- Buscadores electrónicos con la integración de recursos lingüísticos, capaces de localizar y suministrar de forma inmediata información actualizada, a partir de bases documentales multilingües.
- Pantallas hablantes, que permiten a usuarios con discapacidades visuales superar las limitaciones que habían tenido hasta ahora.
- Diccionarios electrónicos multilingües, que facilitan la lectura de textos en lenguas extranjeras.

Como se ha mencionado, las tecnologías del lenguaje involucran una serie de técnicas relacionadas con el tratamiento del lenguaje por medios computacionales. Pueden distinguirse dos grandes vertientes, las que se aplican al tratamiento del lenguaje escrito y aquellas para el tratamiento del lenguaje oral o hablado. Las siguientes secciones profundizan en dichas vertientes [6].

## TECNOLOGÍAS PARA TRATAMIENTO DE TEXTO

Actualmente, el acceso a la información debe ser selectivo y guiado por objetivos; si no es así, a pesar de disponer de una gran cantidad de datos al alcance, no se llegarán a convertir en información que se pueda integrar como conocimiento. Este hecho plantea la necesidad de definir métodos de almacenaje y codificación de la información a fin de que posteriormente pueda ser debidamente recuperada.

El objetivo último es que los documentos puedan ser indexados por conceptos, que se pueda acceder a la información mediante interacciones en lenguaje natural y que sea posible interactuar y recuperar la información en la lengua deseada, todo esto independientemente de la lengua en la que se encuentre originalmente la fuente de consulta.

Sin embargo, actualmente los documentos se indexan por formas o por palabras clave, sin tener en cuenta las relaciones entre las variantes flexivas de las palabras o las relaciones de significado. Tanto las empresas editoriales, en cuanto creadoras de contenidos estructurados (diccionarios y enciclopedias), como los sistemas de búsqueda en Internet o la gestión de los grandes bancos de datos requieren sistemas que permitan el acceso al contenido desde diferentes perspectivas, según las necesidades de los usuarios.

Entre otras técnicas asociadas a la gestión de información se encuentran la extracción de información, la recuperación de información, la traducción automática y la elaboración de resúmenes. Todas ellas requieren una serie de recursos para el procesamiento de los datos, entre los que destacan los analizadores morfológicos, los desambiguadores, los analizadores sintácticos, las redes semánticas, los corpus etiquetados y los lexicones computacionales y multilingües.

El desarrollo de los distintos sistemas que utilizan de los recursos de ingeniería lingüística permite optimizar los resultados de las búsquedas. Las mejoras se pueden producir en diferentes sentidos:

- Permitiendo expresar las búsquedas en lenguaje natural y sin restringirlas a un metalenguaje, que es una combinación de palabras clave y operadores booleanos de aprendizaje complicado.
- Ampliando el espacio de búsqueda mediante la ampliación de la pregunta del usuario incorporando en él sus formas flexivas y las palabras semánticamente relacionadas.
- Restringiendo los documentos seleccionados a aquellos que realmente responden al sentido de las búsquedas.

## TECNOLOGÍAS PARA EL TRATAMIENTO DEL HABLA

El uso de la lengua oral como modo de comunicación con los sistemas informáticos permite, en primer lugar, liberar las manos y la vista, lo que facilita la recepción de la información desde cualquier punto y hace posible la movilidad, de manera que pueden realizarse otras actividades simultáneas. La entrada de datos en la gestión de los fondos de un almacén, las operaciones relacionadas con el mantenimiento y el control de la calidad en entornos industriales, la recepción de información en entornos como cabinas de avión —con una gran cantidad de sistemas de control e indicadores—, el control de la radio, del teléfono o de la climatización de un coche y la obtención de información sobre itinerarios —el prototipo desarrollado en el proyecto *Vodis*<sup>2</sup> (*Advanced Speech Technologies for Voice-operated Driver Information Systems*) ofrece estas posibilidades— o bien el dictado de documentos son algunas de las aplicaciones para las cuales se ha demostrado que la utilización de la lengua hablada es una ventaja importante.

En segundo lugar, utilizar el habla para acceder a la información almacenada en un sistema de cómputo brinda la posibilidad de usar el teléfono en vez de la pantalla y el teclado tradicionales, con todo lo

---

2 <http://cordis.europa.eu/ist/98vienna/xvodis.htm>

que ello implica en cuanto a rapidez. Hay que pensar, además, en la independencia que este procedimiento proporciona con respecto a la distancia y la disponibilidad de una computadora. Algunas aplicaciones en este ámbito son los servicios de información sobre horarios de transportes públicos, las consultas en las páginas amarillas u otros directorios, la información ciudadana, la banca telefónica o las centrales automatizadas.

Finalmente, no hay que olvidar que guardar y recibir información mediante el habla es una alternativa práctica cuando hay que prescindir, debido al tamaño, del teclado, como sucede con las agendas electrónicas y los comunicadores personales. La voz, como los lápices electrónicos y las pantallas táctiles, es una modalidad viable para hacer aumentar la funcionalidad de estos aparatos. Además que esto es un nicho de mercado abierto para aplicaciones en la lengua española.

Hoy en día disponemos de programas comerciales que posibilitan el dictado automático, por ejemplo los desarrollados por Dragon Systems, IBM o Philips, la navegación por Internet, por ejemplo el Free Speech Browser<sup>3</sup>, y la recuperación de información mediante el habla. También es posible la lectura en voz alta de un texto en formato electrónico, como podría ser un mensaje de correo o el contenido de una base de datos. Es posible, además, con los sistemas de diálogo que incorporan síntesis, reconocimiento y un cierto grado de comprensión, obtener información o efectuar transacciones comerciales mediante el teléfono.

Estos últimos son, seguramente, sistemas que en la actualidad despiertan más interés y, por este motivo, se mencionan algunas de las aplicaciones desarrolladas en el contexto de la comunidad europea, donde el fenómeno del acceso a la información multilingüe es crucial para el funcionamiento adecuado y fluido de gran parte de la vida cotidiana.

Entre otras aplicaciones de las tecnologías del habla podemos mencionar los sistemas de diálogo (algunos complementados con información multimodal), la información ciudadana de forma automática mediante el teléfono, los sistemas de traducción automática de con-

---

3 [http://findarticles.com/p/articles/mi\\_m0NEW/is\\_1999\\_Nov\\_4/ai\\_57296080](http://findarticles.com/p/articles/mi_m0NEW/is_1999_Nov_4/ai_57296080)

versación, tanto telefónica como cara a cara, la enseñanza de lenguas asistida por ordenador y la ayuda para los sectores con necesidades especiales.

Tal como queda manifiesto, al incorporarse al mundo de la informática y las telecomunicaciones, las tecnologías del lenguaje están creando un nuevo mercado para productos y servicios especializados.

## IMPORTANCIA DE LAS TECNOLOGÍAS DEL LENGUAJE EN EL MANEJO DE INFORMACIÓN

La clave para el manejo futuro de la información y el conocimiento son los textos. Una prueba de ello es el impacto que tiene Internet en el trabajo cotidiano a todos los niveles, pues gestiona el acceso al conocimiento aunque todavía de manera rudimentaria. En este sentido, la gestión de textos resulta prioritaria en el actual desarrollo de las tecnologías de la información e incide favorablemente en el crecimiento tecnológico de la nación.

El tratamiento de información oral y escrita requiere de estudios teóricos del lenguaje natural (realizado por profesionales como son lingüistas, bibliotecólogos, comunicólogos y otros pertenecientes al área humanística) así como de una serie de metodologías y técnicas de análisis (lo que involucra a profesionales de campos como computación, informática, estadística, ingeniería, etc., con un perfil de ciencias exactas). En este sentido, la interdisciplina incide favorablemente en la interacción y colaboración de sectores del conocimiento que, desde un punto de vista tradicional, se han visto como desligados [7]. Por ello, resulta vital ayudar a aumentar el crecimiento de una sólida planta de investigadores, docentes y técnicos competitivos tanto para el sector de las ciencias exactas como el de las ciencias humanísticas.

El fruto de esta fusión de disciplinas ya se ha visto concretado en el importante auge que ha tenido en otros países lo que se conoce como *tecnologías e industrias del lenguaje y la información*. Este sector ha dado lugar a una gran cantidad de áreas de investigación y estudio fundadas desde mediados del siglo pasado. Estas áreas se han consolidado hasta lograr un crecimiento exponencial en los 80 y 90. De entre ellas, figuran entre las más relevantes la inteligencia artificial, la robó-

tica, el aprendizaje-automático, el procesamiento de lenguaje natural, la ingeniería lingüística, la lingüística computacional, la extracción y recuperación de información, el procesamiento de voz, la representación y generación de conocimiento, la terminótica y la minería de textos. En el caso de nuestro país, la relevancia estratégica que tienen todas estas áreas es aún emergente, debido en gran medida, a la falta de conocimiento y difusión. Por ello, para lograr un alto grado de calidad y competitividad en la generación y aplicación de tecnologías de la información en el país, es prioritario fomentar y destinar fondos para el desarrollo de estas interdisciplinas en nuestro medio.

La creación de una plataforma para el procesamiento del lenguaje natural en español, así como el desarrollo de recursos lingüísticos en nuestro idioma, como son los corpus lingüísticos, los bancos terminológicos, los tesauros y las ontologías, será de alto provecho para la generación propia de múltiples aplicaciones como sistemas de resumen automático, interfaces inteligentes, sistemas pregunta-respuesta, sistemas de extracción de información, buscadores inteligentes, entre otros de los que actualmente se empiezan a crear en el país.

Los beneficios de contar con tecnologías del lenguaje y el aprovechamiento de sus recursos derivados son varios para la gestión de información, entre los que podemos encontrar:

*Búsqueda inteligente de información*, que permita a los usuarios de bases de datos documentales realizar búsquedas textuales en dichos repositorios de información, utilizando frases y contexto, reduciendo con ello las deficiencias de los buscadores actuales, las múltiples búsquedas realizadas manualmente y el gran número de resultados carentes de sentido a los requerimientos de información.

*Distribución selectiva de información* a través de asistentes automáticos personales encargados de monitorear fuentes de información en línea, detectar nueva información, enviar una notificación al usuario con un resumen de dicha información y los accesos necesarios para que pueda consultarla.

*Poblado automático de bases de datos*, que permite a los usuarios obtener, a partir de textos (por ejemplo, noticias, informes, etc.), conjuntos de datos de forma automática, como nombres de personas y sus cargos, nombres de eventos, cifras relevantes, etc. Estos datos



son almacenados en bases de datos para la generación de estadísticas, análisis de datos o gráficas.

## SITUACIÓN ACTUAL EN MÉXICO

La falta de un proyecto a nivel nacional que rijan, gestione y coordine el trabajo de Investigación y Desarrollo (I+D) de los grupos, laboratorios y centros relacionados con las tecnologías de la información provoca una enorme desorganización e inhibe avances en la solución de problemas específicos. La consecuencia lógica es un estancamiento tecnológico y un elevado gasto para el país, pues se ve obligado a adquirir tecnología extranjera ante la poca o nula producción nacional, esto sin mencionar los altos costos de implementación y difusión de tecnologías de información que de inicio son ajenas a la cultura e idiosincrasia de los usuarios del país. Un ejemplo de esto son los llamados sistemas para inteligencia de negocios, diseñados y producidos principalmente por firmas de tecnologías de información extranjeras, que combinan complejos mecanismos computacionales para la explotación de los datos almacenados en sistemas empresariales tradicionales u otras fuentes potenciales de información; sistemas que además requieren de inversiones importantes en el orden de varios cientos de miles de pesos para su adquisición e implantación en territorio nacional. Este tipo de herramientas son inaccesibles y en el mayor de los casos prohibitivas para el grupo empresarial más representativo en México, las micro, pequeñas y medianas empresas [8].

Otro factor que debe ser abordado es la gran deficiencia formativa y de visión interdisciplinaria a nivel académico, aunado a la carencia en la promoción laboral para profesionistas egresados tanto del nivel medio-superior (técnico), superior (licenciatura) y postgrado (maestría, doctorados y post-doctorado). Cabe señalar que a nivel nacional, un gran porcentaje de las carreras afines a las Tecnologías del Lenguaje, la Información y el Conocimiento (TLIC) carecen de un perfil formal y vinculado a la industria para la investigación aplicada, el diseño y desarrollo de nuevas herramientas de producción nacional y en beneficio directo del consumo y los mercados internos.

Aunado a los factores expuestos, el papel que juega la sociedad mexicana sigue siendo uno de los temas decisivos en el crecimiento

y desarrollo de las TLIC en México. La sociedad, siendo primordialmente consumidora, está habituada a la adquisición de tecnología aún cuando ésta no haya sido diseñada para resolver sus necesidades particulares. Como consecuencia, la demanda por nuevos y mejores productos nacionales, así como el uso de las TLIC se ve obstaculizada por el hecho de que los usuarios de éstas deben adecuarse a la tecnología y en muy pocas ocasiones la tecnología se adecua a las necesidades de los usuarios o de los potenciales consumidores. Otro perfil de este problema es que la sociedad también se ha habituado y considera, en un porcentaje muy amplio, que los productos y servicios de las TLIC no merecen una remuneración económica, desvirtuando la industria y provocando que, ante el fenómeno de la piratería, sólo las grandes empresas transnacionales gocen de algún factor de crecimiento, mientras las pocas empresas nacionales sobreviven -en el mejor de los casos- del mercado nacional.

Finalmente, es importante reconocer y tratar el problema relativo a los recursos humanos ubicados en las organizaciones que representan los usuarios potenciales de las TLIC. Al tratarse de tecnologías emergentes, estos recursos humanos no están preparados para aceptar dichas tecnologías, ya sea por desconocimiento de las mismas, o bien por la falta de habilidades y capacitación para llevar a cabo las implementaciones necesarias. Como consecuencia, se opta por la adquisición de otro tipo de tecnologías que tal vez no resuelven el problema de los usuarios, o que sólo ofrecen soluciones parciales, además de implicar costos más elevados ya sea en el corto, mediano o largo plazo, dado que generalmente dichas soluciones provienen de empresas extranjeras.

## EL FUTURO EN MÉXICO

Dada la situación actual descrita, resulta necesario concretar un proyecto a nivel nacional para contribuir a la gestión de información que conduzca al descubrimiento, generación y administración de conocimiento, con base en el uso de tecnologías de lenguaje, partiendo de la inmensa base de información existente en medios electrónicos tanto en formatos textuales como orales.

Una propuesta para ello es crear un centro de ámbito nacional para la investigación aplicada y desarrollo tecnológico, con carácter formativo y de vinculación empresarial, que encamine los esfuerzos para la creación de nuevas tecnologías para la gestión de información y conocimiento con base en las tecnologías del lenguaje. El fin último será coordinar los diversos esfuerzos que existen en I+D en el área para que converjan hacia un fin común, mediante un plan de desarrollo estratégico conforme a las necesidades del país. Entre los objetivos particulares de esta propuesta se proponen:

- Incrementar la I+D nacional de productos para la gestión y explotación de información científica, tecnológica, administrativa y legal para llevar a cabo los más diversos objetivos y, especialmente, para generar conocimiento a fin de alcanzar niveles competitivos y de calidad, logrando con ello un impacto a nivel internacional.
- Facilitar los medios para captar y usar de un modo inteligente la información oral, escrita o visual, proveniente de cualquier campo del conocimiento y de cualquier sector, sea público, privado o social.
- Encauzar la investigación básica y aplicada para crear tecnologías capaces de procesar y manejar volúmenes gigantescos de información que se requieran en la generación de conocimiento.
- Desarrollar productos y recursos tecnológicos (por ejemplo, corpus lingüísticos, bancos terminológicos, ontologías y tesauros) que faciliten el procesamiento de información requerida para fines específicos.
- Poner al alcance de las empresas e instituciones, así como de la población en general, sistemas para recuperar y extraer información, mediante el uso de lenguaje natural.
- Proporcionar los canales adecuados para el consumo de tecnología desarrollada dentro del país a través de una adecuada gestión de cambio organizacional que dé apoyo a este tipo de propuestas.

- Vincular los sectores públicos, privados y sociales para el fomento del desarrollo y uso de las tecnologías, recursos y productos desarrollados en el país. Asimismo, coadyuvar al acercamiento entre el ámbito académico y el ámbito empresarial para la emergencia de colaboraciones entre los dos rubros y para la optimización de la tecnología desarrollada.
- Fomentar la formación de profesionales competitivos y de recursos humanos de alta especialización en el país, sobre todo considerando que pueden llegar a ser más competitivos y rentables que sus similares de países del primer mundo. En este rubro se crearán programas de licenciatura y posgrado que apoyen la especialización y la competitividad de los integrantes del proyecto y de los cuadros que a largo plazo trabajen en este rubro para el gobierno y las empresas.
- Generar ingresos mediante la venta en el extranjero de nuestra tecnología, en lugar de causar gastos en la compra de la misma.
- Cabe señalar que la creación de un centro de esta naturaleza fue propuesta recientemente al Consejo Nacional de Ciencia y Tecnología por iniciativa del Instituto de Ingeniería de la UNAM, con la participación de grupos de investigación de diversas instituciones en el país, como son el Instituto Politécnico Nacional, el Instituto Nacional de Astrofísica, Óptica y Electrónica, la Benemérita Universidad Autónoma de Puebla, El Colegio de México, y de la misma UNAM.

## REFERENCIAS

- [1] R. Trejo Delarbre, "Vivir en la Sociedad de la Información. Orden global y dimensiones locales en el universo digital," en *Comunicación Educativa en la Sociedad de la Información*, (Ed. R. Aparici), Madrid: Universidad Nacional de Educación a Distancia de España, 2003.

- [2] Norvig, Peter y Russell, Stuart, *Artificial intelligence: a modern approach*, (2nd ed.) , Prentice-Hall International, New Jersey, 2003.
- [3] John F. Sowa: Knowledge Representation: *Logical, Philosophical, and Computational Foundations*. Brooks/Cole: New York, 2000.
- [4] Mitkov, Ruslan, *The Oxford Handbook of Computational Linguistics*, Oxford University Press, New York, 2003.
- [5] Cole R.A., J. Mariani, H. Uszkoreit, G. Varile, A. Zaenen, V. Zue, A. Zampolli. *Survey of the State of the Art in Human Language Technology*, Cambridge University Press and Giardini, Cambridge, 1997.
- [6] M. A. Martí y Joaquim Llisterri, “La ingeniería lingüística en la sociedad de la información”, *Digitum, Revista digital d’humanitats* (3), 2001.
- [7] G. Sierra y A. Medina, “Ingeniería Lingüística: Interdisciplina aplicada al desarrollo humano” en *Memorias del Primer Congreso de la Academia de Ingeniería*, Academia de Ingeniería, México, 2003.
- [8] L. Orozco, “Sistemas y tecnologías de información en las micro y pequeñas empresas”, *Razón y Palabra*, núm 39, junio-julio 2004.



# Organización y estructuración del conocimiento en la normalización internacional: el caso de las normas ISO 704 y ANSI/NISO Z39.19-2005

MARÍA POZZI  
*El Colegio de México*

## 1. INTRODUCCIÓN

**D**os disciplinas que tradicionalmente se han ocupado de la organización y estructuración del conocimiento son la terminología y la documentación, cada una de ellas desde su propia perspectiva. Para la documentación, lo más importante es la identificación y recuperación de documentos mediante un vocabulario controlado (en una o más lenguas) constituido por una serie de términos que pertenecen al campo del conocimiento en cuestión y las relaciones que existen entre ellos. Para la terminología, por otra parte, sus dos principales funciones consisten en la representación y la transmisión del conocimiento especializado mediante la identificación de conceptos y las relaciones que existen entre ellos, el análisis y modelación de sistemas de conceptos con base en los conceptos identificados y las relaciones que existen entre ellos, la representación de sistemas de conceptos mediante diagramas de conceptos, la definición de conceptos, la atribución de designaciones —predominantemente términos— a cada concepto en una o más lenguas. Esto significa que la documentación trabaja principalmente, aunque no exclusivamente, en el nivel de la lengua mientras que la terminología lo hace en el nivel del concepto.

Ya que las aplicaciones tanto de la terminología como de la documentación son de naturaleza muy variada y extendida, ambas fueron de las primeras en ser objeto de normalización tanto a nivel nacional como internacional. Las normas están diseñadas para proporcionar reglas, lineamientos y directrices para lograr el grado máximo de orden en el contexto en que han de aplicarse. Por lo tanto, es de esperarse que las normas se apliquen de manera uniforme y en donde no haya lugar para interpretaciones subjetivas.

## 2. LAS NORMAS PUBLICADAS POR LA ORGANIZACIÓN INTERNACIONAL DE NORMALIZACIÓN

La normalización es el proceso de formular y aplicar reglas para tratar de manera ordenada una actividad específica para el beneficio y con la cooperación de todos los interesados, y en particular para la promoción de una economía óptima en su conjunto tomando en cuenta las condiciones de funcionamiento y los requisitos de seguridad. Se logra mediante el consenso de las partes interesadas.

El principal organismo de normalización internacional es la Organización Internacional de Normalización (ISO), cuyos miembros son los organismos nacionales de normalización de cada país. La ISO está constituida por cerca de 230 comités técnicos cada uno de los cuales se ocupa de la elaboración, revisión y publicación de las normas internacionales ISO de su actividad específica. En el caso de la terminología, el responsable de estas actividades es el Comité Técnico ISO/TC37 *Terminología y otros recursos lingüísticos y de contenido* y, en el caso de la documentación, estas actividades corresponden al Comité Técnico ISO/TC46 *Información y documentación*. Una de las primeras tareas de cada uno de estos dos comités fue la organización y estructuración del conocimiento para sus propios fines: en el caso del ISO/TC37 fue la norma ISO 704 *Terminology work - Principles and methods* de la que en la actualidad se está preparando la tercera edición (la primera es de 1987 y la segunda de 2000). El ISO/TC46 publicó en 1974 la primera edición de la norma ISO 2788 *Documentation - Guidelines for the establishment and development of monolingual thesauri* y la segunda edición es de 1986 y la norma ISO



5964 *Documentation - Guidelines for the establishment and development of multilingual thesauri* se publicó en 1985. Debido a que estas dos normas han estado vigentes durante más de 20 años y en respuesta a los grandes adelantos que ha habido en materia de manejo de información, en la actualidad están en las primeras etapas de un importante proceso de revisión que culminará en la publicación de la tercera y segunda edición respectivamente y en donde se esperan cambios sustanciales.

Por otra parte, el American National Standards Institute (ANSI) en colaboración con la National Information Standards Organization (NISO) publicó en 2005 la norma ANSI/NISO Z39.19-2005 *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*, cuyo texto está sirviendo de base para la revisión de las dos normas ISO (2788 y 5964). Entonces, por ser la norma Z39.19-2005 la más utilizada internacionalmente y la más actual, me basaré en ella para llevar a cabo este trabajo.

### 3. ORGANIZACIÓN Y ESTRUCTURACIÓN DEL CONOCIMIENTO EN ISO 704 Y ANSI/NISO Z39.19-2005

Los objetivos principales y la forma de organizar y estructurar el conocimiento desde las perspectivas de la terminología y de la documentación se ven bien reflejados en las normas ISO 704 y ANSI/NISO Z39.19-2005. Algunos de ellos coinciden, otros coinciden sólo parcialmente y otros difieren. En este trabajo me ocuparé de hacer un análisis comparativo de la forma en que ambos lo llevan a cabo.

El conocimiento se estructura mediante el establecimiento de relaciones que pueden ser de muy diversa naturaleza. La terminología, enfocada principalmente al estudio de los conceptos prefiere utilizar *relaciones interconceptuales* y la documentación, más preocupada por la información, prefiere utilizar *relaciones semánticas*. Si bien la esencia de estas relaciones difiere de un campo al otro, en la práctica son muy similares. Estas relaciones pueden ser:

- jerárquicas
  - genéricas

- partitivas
- asociativas

### 3.1 Relaciones jerárquicas

Tradicionalmente las relaciones jerárquicas que se manejan tanto en terminología como en documentación son las genéricas ( $x$  es un tipo de  $y$  /  $x$  es  $y$ ) y las partitivas ( $a$  es una parte de  $b$  /  $a$  es un elemento de  $b$ , etc.).

ANSI/NISO Z39.19 (2005:47) define las relaciones genéricas como “el vínculo entre una clase y sus miembros o especies”.<sup>1</sup> La forma de establecer estas relaciones es “[narrower term] *is a* [broader term]” La naturaleza genérica de una relación puede identificarse mediante los códigos BT/NT (broader term/narrower term) y, en caso de que se requiera mayor precisión se utilizan los códigos BTG (broader term generic) y NTG (narrower term generic) o BTI/NTI (broader term instance/ narrower term instance).

Algunos ejemplos de la forma en que se manejan estas relaciones:

1. BT = coníferas NT = pinos NT = abedules | NT = cipreses
  
2. roedores  
    NTG ratones  
    ratones  
    BTG roedores
  
3. regiones montañosas  
    NTI Alpes  
    NTI Himalayas  
    NTI Andes

De manera similar, según la norma ANSI/NISO Z39.19 (2005:49), las relaciones partitivas (o relaciones todo-parte) existen cuando un

---

1 “This relationship identifies the link between a class and its members or species”.

concepto está incluido en otro sin importar el contexto de manera que los términos se pueden organizar en jerarquías lógicas, esto es, que el todo se puede considerar como el *broader term* y cada una de las partes como el *narrower term*. Estas relaciones se indican explícitamente con los siguientes códigos:

BTP = broader term partitive y

NTP = narrower term partitive

Algunos ejemplos de relaciones partitivas:

1. sistemas y órganos del cuerpo humano:

sistema nervioso

sistema nervioso central

cerebro

médula espinal

sistema nervioso central

BTP sistema nervioso

sistema nervioso

NTP sistema nervioso central

2. lugares geográficos

Canadá

Ontario

Ottawa

Toronto

Ontario

NTP Ottawa

NTP Toronto

Toronto

BTP Ontario

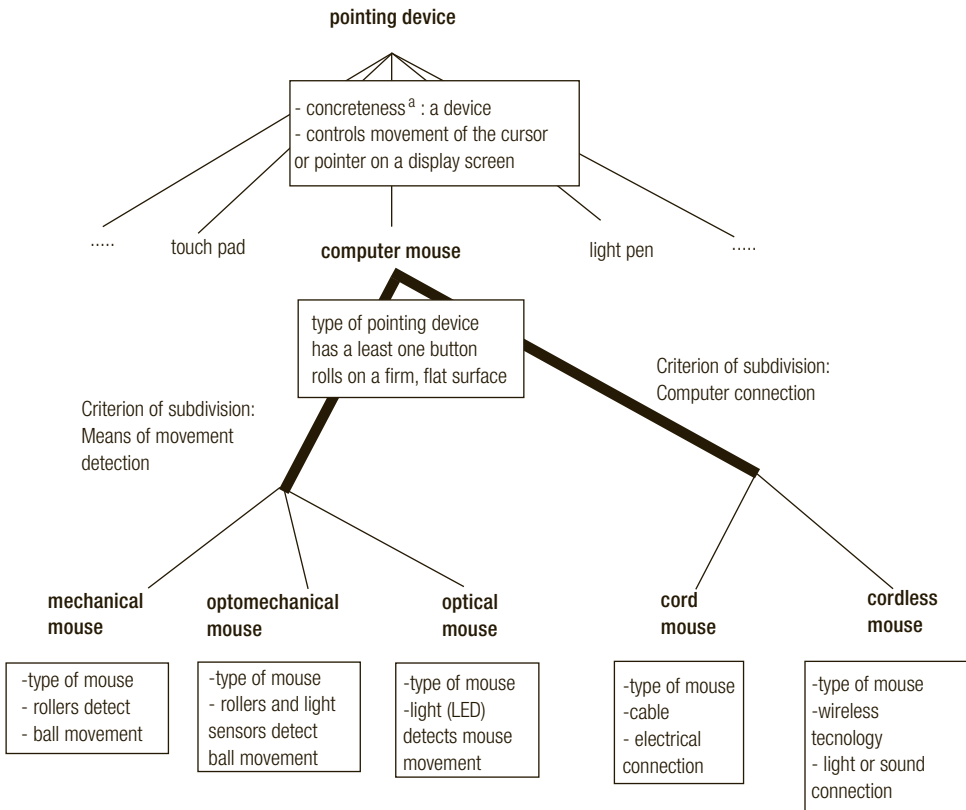
Ottawa

BTP Ontario

Por otra parte, la norma ISO 704 (2007:6-19) hace más énfasis en la construcción de un sistema de conceptos que se basa en las relaciones establecidas entre los conceptos del campo el conocimiento en cues-

tión. Las terminologías que se elaboren siguiendo esta norma deberán reflejar estas relaciones aunque no siempre están explícitamente señaladas.

Para construir un sistema de conceptos basado en una relación genérica, todos los conceptos subordinados están en una relación  $x$  es un tipo de  $y$ . De manera gráfica se pueden representar de la siguiente manera<sup>2</sup>:

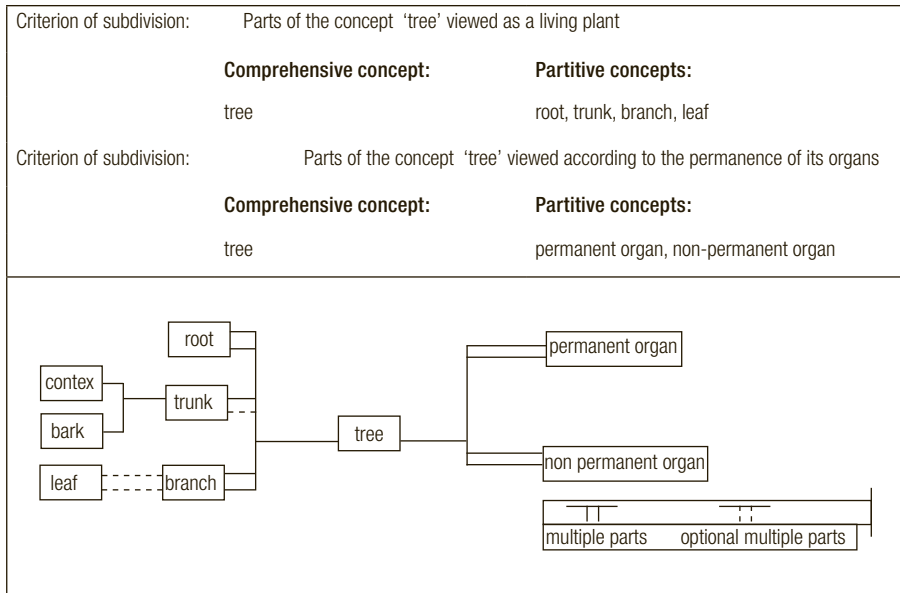


2 Ejemplo tomado de ISO 704(2007:10)

De otra manera, el sistema de conceptos puede representarse de la siguiente manera<sup>3</sup>:

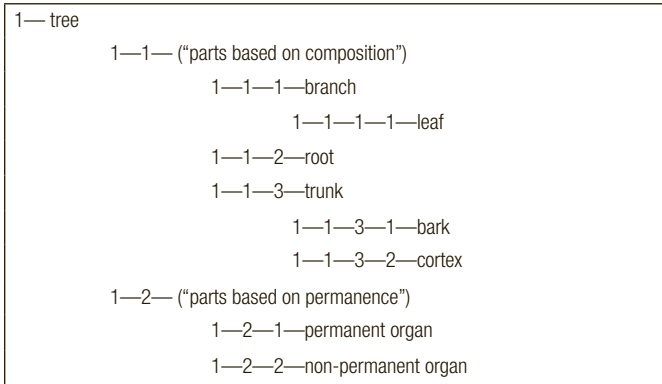
1	pointing device			
	1.1	touch pad		
	1.2	computer mouse		
			1.2.1	
				1.2.1.1
				1.2.1.2
				1.2.1.3
			1.2.2	
				1.2.2.1
				1.2.2.2
1.3	light pen ...			

Las relaciones partitivas (*a* es una parte de *b* / *a* es un elemento de *b*) se representan de manera similar, como en el siguiente ejemplo<sup>4</sup>:



3 Ejemplo tomado de ISO 704(2007:10)

4 Ejemplo tomado de ISO 704(2007:14)



### 3.2 Relaciones asociativas

De acuerdo con la norma ANSI/NISO Z39.19, estas relaciones asocian términos que no son sinónimos ni forman parte de una jerarquía pero que están semántica o conceptualmente relacionados de tal manera que la relación debe hacerse explícita en un vocabulario controlado ya que pueden sugerir términos que se pueden utilizar para la indización o para la recuperación de información. Existen varios tipos de relaciones asociativas en los tesauros, pero las más comunes son las simétricas, que se representan con el código RT. Algunos ejemplos de relaciones asociativas son:

1. células  
    RT citología  
    citología  
    RT células
  
2. burros  
    BT equinos  
    RT mulas  
    equinos  
    NT burros  
    NT caballos

NT mulas  
 caballos  
 BT equinos  
 RT mulas  
 mulas  
 BT equinos  
 RT burros  
 RT caballos

3. control de temperatura  
 RT termostatos  
 termostatos  
 RT control de temperatura

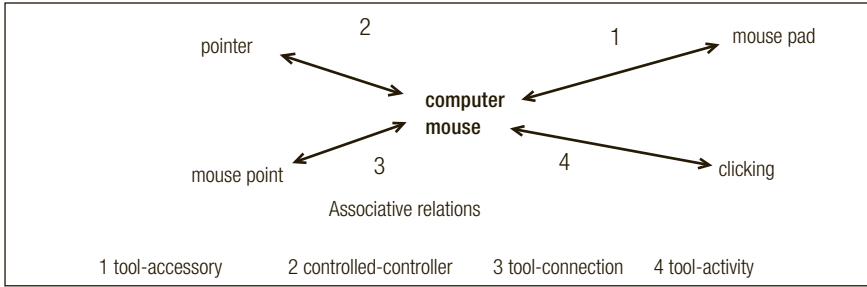
Por su parte, en la norma ISO 704, las relaciones asociativas existen cuando se puede establecer un nexo temático entre dos o más conceptos en virtud de la experiencia. En consecuencia hay muchos tipos de relaciones asociativas que se pueden establecer entre los conceptos de un campo del conocimiento. A continuación se muestran unos cuantos ejemplos<sup>5</sup>:

Concepts		Associative relation
milk carton	↔	milk container – contained
clicking	↔	computer mouse activity – tool
gametes	↔	zygote ↔ zygospore steps of a cycle
humidity	↔	corrosion cause – effect
baker	↔	bread producer – product
time	↔	clock duration – measuring device
painter	↔	brush profession – typical tool
screw	↔	screwdriver object – associated tool
noon	↔	day point in time – time unit

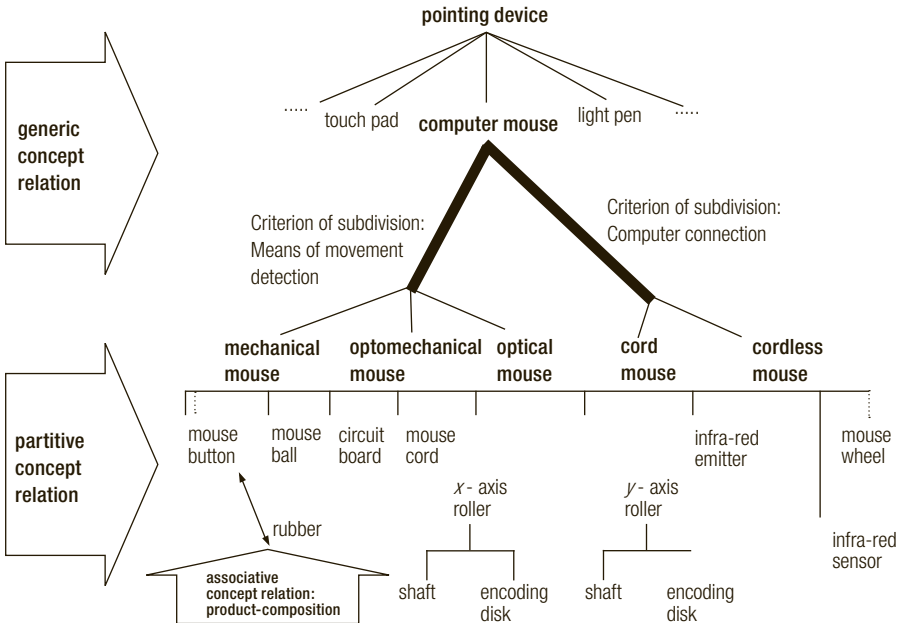
---

5 Ejemplos tomados de ISO 704(2007:16)

En un sistema de conceptos las relaciones asociativas se representan de la siguiente manera<sup>6</sup>:



La mayoría de los campos del conocimiento se estructuran mediante sistemas de conceptos mixtos, como en el del siguiente ejemplo<sup>7</sup>:



6 Ejemplo tomado de ISO 704(2007:17)

7 Ejemplo tomado de ISO 704(2007:18)



### 3.3 Relaciones de equivalencia

Una relación de equivalencia existe entre dos o más términos que representan el mismo concepto. En este caso, uno de ellos se identifica como el *término preferido* y los demás términos son referidos a él. Este tipo de relación se representa explícitamente en los vocabularios controlados mediante los códigos USE y UF (use for). Por ejemplo:

1. aspirina  
    USE ácido acetilsalicílico  
    ácido acetilsalicílico  
    UF aspirina
2. poliomielitis  
    UF parálisis infantil  
    parálisis infantil  
    USE poliomielitis
3. agua de mar  
    UF agua salada  
    agua salada  
    USE agua de mar

En el campo de la terminología, en el contexto de la norma ISO 704 se trabaja con una orientación al concepto, lo que significa que se recogen todos los sinónimos que designan al concepto en cuestión y se selecciona el término preferido, los sinónimos aceptables y aquellos cuyo uso debe evitarse.

### 3.4 Relaciones de homonimia

Para evitar ambigüedad en la representación del conocimiento mediante vocabularios controlados, según la norma ANSI/NISO Z39.19, los homónimos requieren ser clarificados mediante una indicación del campo al que pertenece el término. Por ejemplo:

1. mercurio (metal)  
    Mercurio (planeta)  
    Mercurio (dios romano)

2. gato (animal)  
gato (hidráulica)  
gato (juego de mesa)
3. banco (institución financiera)  
banco (mueble)  
banco (ictiología)

En terminología, por otro lado, el problema de los homónimos se soluciona de manera similar al especificar el campo al que pertenece dicho concepto.

#### 4. PRODUCTOS

Cada uno de estos documentos normativos está diseñado para producir diferentes tipos de productos: la norma ISO 704 se aplica sobre todo para la de recursos lingüísticos tales como vocabularios monolingües o multilingües que contienen terminología normalizada o no y, en el caso del la norma ANSI/NISO Z39.19, tesauros monolingües o multilingües.

#### 5. CONCLUSIONES

- a) Las dos disciplinas tienen necesidad de organizar y estructurar el conocimiento, cada una para sus fines específicos.
- b) La documentación necesita optimizar el manejo y recuperación de información. Para ello establece relaciones entre los términos semánticamente cercanos y minimiza la ambigüedad mediante la selección de un término preferido y la desambiguación de homónimos.
- c) La terminología requiere comprender los diferentes campos del conocimiento, para lo cual construye sistemas de conceptos mediante el establecimiento de relaciones interconceptuales. Esto garantiza la consistencia y la pertinencia de un concepto en el campo del conocimiento en cuestión.
- d) Cada disciplina presenta los resultados de su trabajo de forma diferente pero complementaria.

e) Tanto la terminología como para la documentación, en especial en el ámbito de la normalización, el ideal es lograr que un concepto sea denominado por un solo término y que cada término sea la representación de un solo concepto.

## REFERENCIAS BIBLIOGRÁFICAS

ANSI/NISO Z39.19 (2005)

*Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*, NISO Press, Bethesda, USA.

ISO 2788 (1986)

*Documentation - Guidelines for the establishment and development of monolingual thesauri*, Second edition, ISO, Ginebra.

ISO 5964 (1985)

*Documentation - Guidelines for the establishment and development of multilingual thesauri*, ISO, Ginebra.

ISO/DIS 704 (2007)

*Terminology work - Principles and methods*, ISO, Ginebra.



# Indización y recuperación de la información en Internet. El usuario ante la interpretación y traducción de la demanda

MARTA FUENTES AGUSTÍ  
*Universitat Autònoma de Barcelona*  
JOSÉ HERNÁNDEZ SERRANO  
*Universidad de Salamanca, España*

**L**os ciudadanos de hoy deberán hacer frente a un mundo donde dominar un conjunto de habilidades informacionales básicas resulta imprescindible, al mismo tiempo que deberán convivir en un entorno social y cultural basado en nuevas formas de participación, creación y comunicación caracterizadas por la rápida y constante expansión de las tecnologías de la información y la comunicación (TIC). Esta situación amplía notablemente el campo de investigación en torno a los sistemas de estructuración, indización, recuperación, gestión y organización de la información, al mismo tiempo que convierte la Web en un poderoso instrumento de acceso a la información.

Las autoras (Fuentes, 2006; Hernández, 2006) han observado y analizado cómo principiantes y expertos buscan la información en Internet y la relevancia de ser estratégico en el proceso de búsqueda y recuperación de información. En este proceso una de las fases principales es la interrogación al sistema de información a través de la utilización de la terminología adecuada. La eficacia de esta fase viene determinada por la compleja combinación resultante entre lo que el *usuario* desea encontrar, cómo lo traduce en forma de *pregunta* al

medio, y cómo éste *interpreta* la demanda propuesta por el usuario. Estas tres variables son las que derivan en una respuesta ajustada a la necesidad informativa del momento.

El presente artículo pretende abordar en primer lugar, la comprensión sobre el tipo de información que se halla en Internet; su organización y posibilidades de acceso; para profundizar después en los sistemas prototípicos de recuperación de la información utilizados por los estudiantes de educación secundaria y superior, de acuerdo a los estudios realizados. Por último, concluiremos con algunas consideraciones sobre la formación en procesos alternativos de búsqueda de información, en base a los objetivos de búsqueda, los conocimientos del usuario y la indización del sistema; el uso de términos adecuados y los criterios para su validación; así como la representación e interpretación de la demanda por parte del usuario y del sistema y su incidencia en los resultados obtenidos.

## LA INFORMACIÓN DE INTERNET. INDIZACIÓN Y RECUPERACIÓN.

Uno de los formatos que más se usa en la actualidad para transmitir información es el formato online, basado en las posibilidades que ofrece Internet. En la Red podemos encontrar información de temáticas y índole muy diversa. Por ejemplo, podemos fijarnos en su origen y distinguir diferentes procedencias (institucional, de comunidades locales concretas, gubernamental, de catálogos de bibliotecas, textos impresos digitalizados, de fuentes empresariales y comerciales, de los medios de comunicación, de ciudadanos individuales, etc); en su finalidad y hablar de información comercial, publicitaria, educativa, de ocio, comunicativa, administrativa, de autoprotección, autopromoción, etc.; en su creación y distinguir entre si se desarrolló directamente para la Web, si fue adaptada de una fuente primaria, si es una transcripción literal de una información impresa, etc.

Si nos centramos en la temporalización o vigencia de la información, ésta puede ser histórica, reciente, antigua o actualizada. Por otra parte, puede presentarse en diversos formatos de edición (.doc, .ppt, .pdf, .exe, .html, .zip, .jpg, .gif, etc.); de presentación (textual, hiper-

textual, numérica, visual, sonora, audiovisual, multimedia, etc.) o siguiendo diferentes esquemas, según la forma que ésta puede adoptar (monográficos, ensayos, informes, artículos, documentos académicos, libros electrónicos, comentarios, ponencias, juegos, ficción, sumarios, guías...)

Incluso podemos distinguir la información según su facilidad de localización (organizada y clasificada manualmente, accesible a través de foros, etc.), accesibilidad (de acceso libre o restringido), coste de acceso (gratuita o de pago), etc.

Esta diversidad de contenidos y formatos informativos viene propiciada fundamentalmente por la libertad y simplicidad de publicación y edición que permite Internet. Cualquier usuario de la Red, sin límites ni condiciones aparentes, puede convertirse en autor de la información generando una “realidad” internáutica que puede o no coincidir con la realidad cotidiana, física y tangible, de la información que estamos acostumbrados a transmitir y consumir.

Por un lado, la libertad de publicación propicia un aumento constante de la información a consultar y un amplio abanico de autores procedentes de diferentes organismos o instituciones, geográficamente dispersos con características heterogéneas en cuanto a edad, género o nivel sociocultural. Y por otro lado, la facilidad de publicación que se permite en Internet propicia una rápida edición y reduce la cadena informativa entre la fuente y los usuarios facilitando, a su vez, la actualización instantánea de la información.

Esta situación deja atrás la seguridad de fidelidad y fiabilidad de la información que garantizaban otros medios de información anteriores, permitiendo la circulación simultánea de diversas versiones de un mismo documento, y la desaparición de jerarquías situando en un mismo nivel las producciones de un escritor premio novel con las elaboradas por un estudiante.

Todo esto ofrece mayor visibilidad de la información, desde el punto de vista de su creación, y promueve una nueva identidad y visión de los contenidos, personas, grupos, textos, situaciones, etc. Al mismo tiempo, que permite una mayor circulación de información académica y científica, llegándose a visualizar los *colegios invisibles* de científicos.

Internet es el espacio de difusión de la información más abierto que existe en la actualidad, de fácil acceso, a-temporal y a-espacial, transversal, fácil de usar, atractivo, etc. pero, al mismo tiempo, se caracteriza por su *desorden documental*. Nació en el Laboratorio Europeo de Física de Partículas del Centro Europeo de Investigación Nuclear (CERN) con una finalidad muy distinta a la de transmitir información a escala mundial (Vea, 2005). Esto ha propiciado que no exista un control de organización de la información.

El incremento y falta de organización de la información, su diversificación y su localización distribuida, genera una serie de circunstancias y consecuencias que deben ser valoradas. Entre ellas tomamos en consideración la relación entre la facilidad de publicación y una posible sobreinformación —que dificulta la selección— e infoxicación —que dificulta la valoración de la credibilidad y fiabilidad de la información— junto con la inminente caducidad de misma (mucha información esta sin actualizar o no se sabe cuándo se actualizó por última vez).

Ante esta estructura informacional resulta imprescindible el uso de herramientas para poder localizar la información precisa en el momento oportuno. La Red nos ofrece diferentes posibilidades de acceso a la información que van de las más tradicionales (consulta de una base de datos recopilados y clasificados manualmente) a los más sofisticados motores de búsqueda nacidos en la era de Internet.

## HABILIDADES INFORMACIONALES PARA RECUPERAR LA INFORMACIÓN PUBLICADA EN LA RED

Venimos hablando de un formato para transmitir la información remoto, dinámico y en constante evolución que requerirá unos usuarios con habilidades informacionales generales y específicas (Fuentes, 2001).

El creciente intercambio de información económica, jurídica, deportiva, científica, artística, cultural, etc. sin una indización aparente, ha creado la necesidad de contar con sistemas de recuperación de información cada vez más sofisticados y diversos de índole muy diversa. Podemos hablar de tres posibilidades para la recuperación de la información publicada en la Red.

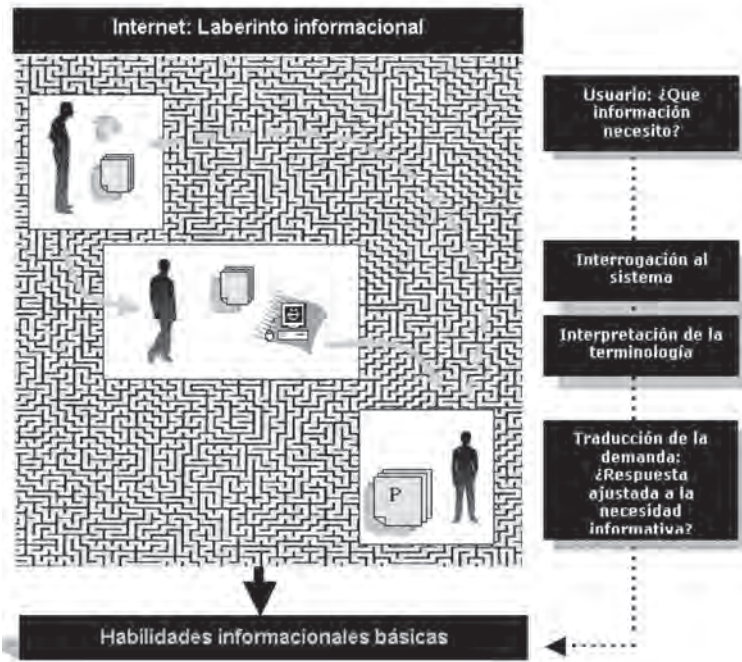


1. Mediante búsquedas realizadas en bases de datos tradicionales perfeccionadas con el avance de las Tecnologías de la Información y la Comunicación (TIC): bases de datos de las bibliotecas con acceso remoto, bases de datos referenciales, hemerotecas digitales, bases de datos a texto completo, especializadas, integrales, etc.
2. Mediante herramientas que permiten búsquedas directas en la red de Internet generalmente a través de robots: directorios o índices temáticos, motores de búsqueda genéricos o específicos, metabuscadores, portales.
3. Mediante herramientas y sistemas de comunicación interpersonal y trabajo colaborativo que nos ofrece la red: correo electrónico, chat, videoconferencias, foros, debates, listas de distribución, etc.

Las características de estos medios de búsqueda incrementan la necesidad de contar con estructuras terminológicas precisas para recuperar adecuadamente los contenidos deseados. No basta con seguir un proceso automatizado de búsqueda caracterizado por un *zapping* compulsivo sin partir de un criterio previo de selección sino que se precisa de un proceso deliberativo (Monereo y Fuentes, 2005).

Los expertos en la temática identifican varias fases en el proceso de recuperación de información en Internet (Fuentes, 2006, 309-327). A pesar de las peculiaridades de cada uno de sus modelos coinciden en resaltar la importancia de la definición de la demanda, la elección del medio y lenguaje de búsqueda, la selección de la información hallada y la interpretación y utilización de esta información. Centrándonos en el proceso de recuperación de la información debemos insistir en el laberinto informacional que supone Internet en cuanto a la falta de indización de la información y la exigencia de ciertas habilidades informacionales básicas. Habilidades que deberían poseer todos los usuarios de la red.

Figura 1. Recuperación de la información en Internet. De la necesidad informativa a los contenidos.

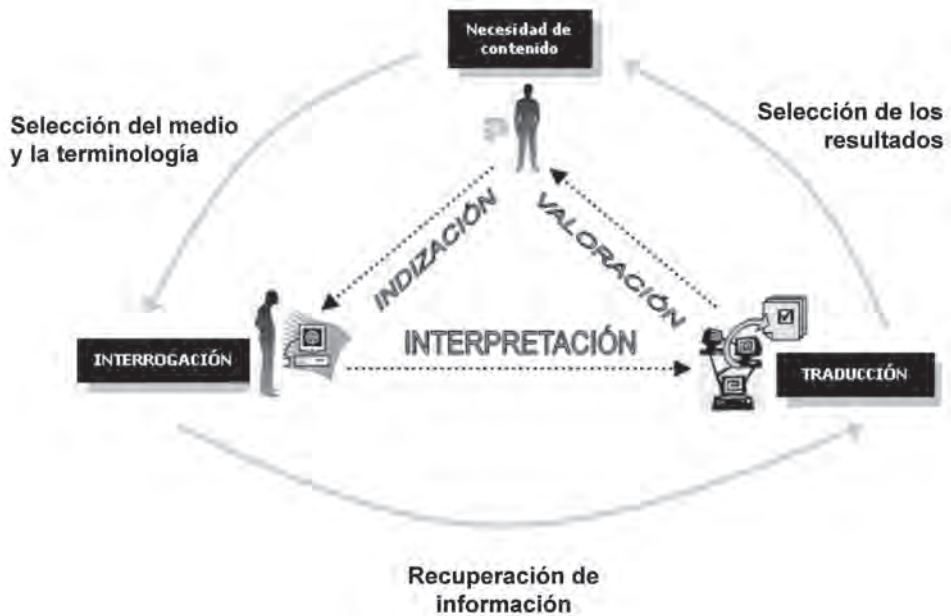


En la figura precedente hemos sintetizado el proceso de recuperación de la información haciendo hincapié en la necesaria *interpretación y traducción de la demanda por parte del usuario* el cual recibirá como respuesta la *interpretación y traducción del sistema* en un listado de posibles recursos, fuentes informativas, etc. que pueden o no dar respuesta a su necesidad.

Para que el sistema proporcione la respuesta deseada se debe interrogar correctamente al sistema de búsqueda elegido. Por un lado, se debe tener presente las características y modo de búsqueda del sistema y por otro las características de la información que se desea hallar. Con todo ello, se debe definir muy bien la demanda usando la terminología adecuada, ampliando la elección de las palabras clave a la elección de los conectores entre ellas y/o las opciones limitadoras de la búsqueda que nos ofrece el sistema.

Podemos hablar de tres grandes variables que inciden directamente en esta parte del proceso de búsqueda de información en la red de Internet: lo que el usuario desea encontrar, la pregunta que se formula al sistema y la interpretación que hace el sistema a dicha pregunta.

Figura 2: Contenido, pregunta y respuesta en la recuperación de información en Internet.



Cuando el usuario decide realizar la búsqueda en la red de Internet debe plantearse que tipo de recurso necesita (un documento de texto, una imagen, una hoja de cálculo, un video, etc. divulgativo, académico, etc. actual, etc.), cual será el sistema de búsqueda más apropiado (base de datos estructurada, motor de búsqueda genérico o especializado, directorio, metabuscador, lista de distribución, consulta a un experto mediante el correo electrónico, portal, etc.) y de entre estos el más adecuado (Google, Altavista, Yahoo, la web del ministerio, de

una escuela, etc.) para llegar al contenido deseado.

Esta premisa nos servirá de punto de partida para la exploración, pero una vez elegido el sistema de búsqueda (base de datos estándar, motor de búsqueda, directorio, etc.) la eficacia viene determinada por la compleja combinación resultante entre lo que el usuario desea encontrar, cómo lo traduce en forma de pregunta al sistema elegido, es decir, la traducción de su demanda al lenguaje del sistema para que este lo interprete adecuadamente y lo traduzca en un listado limitado de posibles documentos o enlaces que satisfagan la necesidad informativa del usuario (figura 2).

Si se usa el lenguaje de interrogación adecuado el sistema interpretará correctamente la demanda y este recuperará la información de manera eficaz. En caso contrario se deberá desestimar el resultado y repetir la búsqueda. Aconsejando no iniciarla de nuevo, sin más, sino actuar de manera estratégica (consciente e intencionalmente) y autorregular el proceso.

Los enlaces recuperados por el sistema deben de ser un listado limitado. Recuperar información en Internet no consiste en localizar información sobre la temática deseada sino que esta pueda dar respuesta a nuestros objetivos. Es decir, cumpla con los requisitos mínimos para que el usuario pueda asumir la selección del o los recursos idóneos teniendo en cuenta el número de resultados, el ajuste al tópico de la búsqueda, la credibilidad y fiabilidad de la información hallada, etc.

Elegir el camino correcto no es una tarea fácil y requiere de ciertos conocimientos y habilidades informaciones que desde la educación obligatoria u otras instituciones (por ej. las bibliotecas) se debe tratar de ayudar a adquirir mediante tareas auténticas.

Nuestros estudios (Fuentes, 2006; Hernández, 2006) han identificado entre los estudiantes seis modalidades de búsqueda de información. De estas modalidades las más usadas se alejan bastante de las usadas por los expertos. Este hecho se debe principalmente al tipo de tareas planteadas habitualmente en la escuela obligatoria o en la educación superior.

Generalmente los estudiantes han recuperado información de Internet para dar respuesta a demandas simples con lo que un proce-

so automatizado puede llegar a satisfacer su necesidad informativa. Mientras que ante una demanda más compleja se ha evidenciado la necesidad de ser estratégico en la búsqueda y las modalidades más habituales de búsqueda les han resultado infructuosas.

Con todo queremos insistir una vez más en la importancia de la interpretación y traducción de la demanda, en la necesidad de ser consciente de la relevancia de las decisiones tomadas (elección del sistema de búsqueda y de la terminología usada) y de todo el proceso que implica recuperar información de la red. Información que debe de ser seleccionada en base a unos criterios para pasar a ser analizada, tratada y usada adecuadamente.

## A MODO DE CONCLUSIÓN

Conocedores de la organización del conocimiento, los procesos de recuperación de la información y la rápida y constante evolución de los formatos de transmisión se debe proceder, por un lado, a analizar como acercar a los ciudadanos las habilidades informacionales básicas argumentadas y, paralelamente, analizar los métodos utilizados para elaborar herramientas de indización y recuperación de la información, las técnicas utilizadas para el tratamiento terminológico, la representación del contenido temático de los documentos multimedia, las normas y opciones para organizar las estructuras temáticas y terminológicas en diferentes ámbitos del conocimiento, los programas para la gestión y clasificación de la información recuperada, los problemas no resueltos de acceso a formatos diversos de transmisión de la información (brecha digital), de compatibilidad entre los sistemas de almacenamiento, indización y recuperación de la información, etc.

Sin pretender ser más exhaustos en la temática, concluimos este artículo insistiendo en que ante un banco creciente y sin control aparente de información se requiere, más que en cualquier otro formato de recuperación de información, adquirir ciertas habilidades informacionales generales y específicas para poder ser estratégico en la búsqueda y selección de información.

## REFERENCIAS BIBLIOGRÁFICAS

- Fuentes, M. (2001). Naufragar en Internet. Estrategias de búsqueda de información en redes telemáticas. *Virtual Educa*, Madrid, 27-29 de junio de 2001. [Publicado por la UOC en <http://www.uoc.edu/web/esp/art/uoc/0109037/fuentes.html#bibliografia>]
- Fuentes, M. (2006). *Estratègies de cerca i selecció d'informació a Internet. Anàlisi de les modalitats de cerca i selecció d'informació a Internet dels estudiants de quart curs d'educació secundària obligatoria*. UAB: Tesis Doctoral.
- Hernández, M. J. (2006). *Estrategias de búsqueda de información en Internet en estudiantes universitarios*. USAL: Proyecto de investigación doctoral.
- Monereo, C. y Fuentes, M. (2005). Aprender a buscar y seleccionar en Internet. En Monereo, C. (Coord.). (2005). *Internet y competencias básicas*. Barcelona: Graó. 27-50.
- Vea, A. (2005). *Who is Who in the Internet World?*. Barcelona: CIDEM.

# Indización y recuperación de la información: términos que representan las interrelaciones en la zona fronteriza norte

LIL BIDART

*Universidad de la República de Uruguay*

## RESUMEN

**E**l punto de partida de esta ponencia es la zona fronteriza, frontera, región fronteriza, etc. o las múltiples expresiones que se utilizan en los discursos de sus habitantes.

La frontera México-Estados Unidos plantea un desafío en la utilización del español y en la determinación de los términos que permitan la recuperación de la información generada en la región.

Se encuentran y se desencuentran países con un grado de desarrollo desigual, con tradiciones históricas, culturales y lingüísticas diferentes.

La línea fronteriza además de ser una línea real, territorial, adquiere una dimensión simbólica que se delimita por una línea imaginaria compuesta por las múltiples relaciones que se establecen diariamente entre las personas que habitan la región y sus relaciones con las otras regiones tanto de Estados Unidos de Norteamérica como de México.

La presencia de dos lenguas y probablemente la existencia de más de un dialecto, aumentan la complejidad del fenómeno.

Vamos a presentar un conjunto de términos que representan, en el imaginario colectivo, las complejas relaciones que establece la po-

blación permanente o los habitantes transitorios, con la región fronteriza.

La elaboración de un tesoro, el uso de lenguajes controlados, la recuperación de la información, deben reflejar la multidimensionalidad de la dinámica fronteriza.

## INTRODUCCIÓN

Quiero presentar aquí algunas reflexiones sobre la investigación de los términos que representan la compleja realidad de la frontera norte de México.

Mi trabajo como profesional de la información ha sido un proceso de comprensión paulatina de las complejidades de los diferentes aspectos de la bibliotecología y en especial, de las relaciones de la lingüística, ahora de la terminología, con el procesamiento y la recuperación de la información.

Cuando uno comienza a trabajar en una biblioteca, no puede ver toda la dimensión del problema. Para mí ha sido interesante este descubrimiento en la medida en que relaciona dos aspectos esenciales: las infinitas problemáticas del uso del lenguaje y el ejercicio de la profesión de bibliotecólogo teniendo como centro la comunicación con el usuario. No puedo dejar de recordar siempre las palabras de Barreto (2000) que resumen poéticamente esta relación: *Assim é nossa crença que o destino final, o objetivo do trabalho com a informação é promover o desenvolvimento do indivíduo de seu grupo e da sociedade. Entendemos por desenvolvimento de uma forma ampla, como um acréscimo de bem estar, um novo estágio de qualidade de convivência, alcançado através da informação. A ação social maior é fazer a luz brilhar para cada ser humano através da informação como mediadora do conhecimento.*

Quiero tomar como punto de partida de esta presentación, aun siendo una reiteración, la importancia del procesamiento técnico de la información para la recuperación de la misma.

El segundo punto de partida que quiero mencionar es la responsabilidad del profesional bibliotecario en este proceso. Su aporte consiste en la incorporación de las técnicas de catalogación y clasi-



ficación, de la realización de análisis de contenidos adecuadas a las necesidades de los usuarios, al acervo, a los servicios de las unidades de información. Esta incorporación debe estar siempre enriquecida con la utilización de las nuevas tecnologías, de formatos y técnicas normalizados a nivel nacional e internacional y sobre todo, el profesional bibliotecario debe incorporar en forma permanente el desarrollo teórico realizado desde la bibliotecología y desde las disciplinas de relación esenciales como la lingüística, la terminología o la comunicación.

Por lo tanto, el proceso de análisis del contenido temático de la información, de la documentación, del acervo en su conjunto, implica una serie de dificultades que muchas veces se asumen en forma burocrática y se resuelven en forma mecánica tomando una palabra genérica de una lista de encabezamiento de materias o del lenguaje natural.

Esta forma de resolver el problema no contribuye a la recuperación de la información y menos a brindar al usuario la documentación pertinente, en el momento justo. Además, en este momento, debemos agregar los efectos de la aplicación de las nuevas tecnologías sobre la información que agregan un nivel mas de complejidad para que el bibliotecólogo pueda cumplir el papel esencial que señalamos anteriormente, de recuperador de la información, de mediador entre la información y el usuario para que esta pueda ser utilizada en su entorno social, laboral, de especialidad, cultural o de difusión.

El análisis temático del contenido de la información es el proceso que sintetiza la relación entre el creador de conocimiento, de información y el usuario. La mediación que realiza el profesional de la información contiene el llamado valor agregado que ahora se hace presente a través de los instrumentos electrónicos.

El lenguaje sigue siendo la esencia de este proceso. Es necesario conocer el tema pero sin duda que la representación del mismo y las reglas que rigen esta representación pasan a primer plano en el proceso de análisis.

No se pueden esperar resultados de calidad si no se adecua el uso de las nuevas tecnologías con los contenidos actuales de la información. Y no podemos pensar en una recuperación ajustada a las necesidades del usuario si no realizamos el proceso con un conjunto de habilidades específicas, propias del momento actual.

El bagaje cultural del profesional de la información necesariamente debe estar integrado por el conocimiento de la temática del acervo que debe manejar y del uso cabal de las nuevas tecnologías. A esto es necesario agregar, el manejo del lenguaje como instrumento fundamental de la comunicación con el usuario y en forma esencial, como portador de las representaciones temáticas de los contenidos.

Tal vez se haya dicho esto muchas veces, pero comprender que el lenguaje de especialidad va a ser usado a través de un tesoro, de un lenguaje documental que representa los contenidos temáticos, es un concepto complejo.

Desde el ángulo terminológico, no se ha incorporado fácilmente en los programas de formación del bibliotecario, en los primeros niveles.

Si nos referimos a los profesionales que han realizado el desarrollo teórico principal, naturalmente que han sido los primeros en vincular la bibliotecología al conjunto de disciplinas que contribuyen a la comprensión de un fenómeno tan complejo como es la información. Durante mucho tiempo, la formación del bibliotecario no se nutría de las teorías de la información, de las teorías de la comunicación social, de la terminología, de la lingüística o de la teoría del discurso.

En este momento del desarrollo de la sociedad de la información y de Internet, nuestros usuarios manejan las problemáticas del lenguaje natural y las problemáticas del lenguaje controlado. Sin embargo, no siempre pueden hacer uso en toda su potencialidad de esta situación de privilegio para el acceso, por no contar con las herramientas conceptuales necesarias. Por lo tanto, se acrecienta el papel del profesional de la información en el acceso y la recuperación de la información.

Es de todos conocido el hecho de que el lenguaje natural, nuestro lenguaje de todos los actos de comunicación, es un sistema de signos. Estamos hablando de signos que representan a los conceptos y estos a su vez se interrelacionan entre sí. Por lo tanto, los lenguajes documentales deben permitirnos establecer estas relaciones en un segundo nivel de representación.

Entonces me pregunto si el término frontera, ubicado en los tesoros en el nivel de nación o territorio, representa la complejidad del concepto frontera.

## LA RECUPERACIÓN DE LA INFORMACIÓN SOBRE LA FRONTERA

Este trabajo se origina en la necesidad de realizar el análisis temático de una colección de documentos sobre Baja California y sobre Tijuana en particular.

Nos encontramos con la problemática de la no existencia de un tesoro o de un lenguaje documental que permitiera la recuperación de la información sobre la frontera norte de México.

Casi todas las bibliotecas de Tijuana cuentan con una colección que muestra la producción cultural de la región. Sin embargo, no encontramos términos que representen esos contenido que no provengan de tesauros o listas de encabezamiento de materias de las áreas más generales del conocimiento.

Como dice Barité (2000:6)“... en los documentos donde se expresa el conocimiento, los conceptos se encuentran combinados, enfrentados, comparados o debatidos, lleva a la necesidad de que los instrumentos de clasificación de esos documentos invoquen y contemplen la compleja trama involucrada”.

En el caso de las regiones fronterizas, esta compleja trama esta integrada por dos aspectos esenciales, relacionados entre si. Por un lado, efectivamente, la ubicación geográfica es el centro del problema. El territorio binacional que une o separa dos naciones determina las particularidades de las relaciones. Pero, a la vez, plantea el conjunto del pensamiento relacionado con el limite, con la alteridad, con la concepción filosófica del mundo.

Esta polisemia del termino frontera se establece en el plano teórico y crudamente en la región.

En general, hemos encontrado que los términos seleccionados para representar los contenidos de la problemática fronteriza son genéricos. Hemos detectado muy pocos casos, en los lenguajes controlados, del uso de expresiones que identifiquen las particularidades, que diferencien los problemas fronterizos.

La investigación del termino frontera la he realizado en las bibliotecas públicas y en las bibliotecas de enseñanza de Tijuana.

En esta primera etapa del trabajo, he analizado el término en los catálogos en línea de las bibliotecas de América Latina, de los estados

norteamericanos fronterizos con México; he realizado búsquedas en Internet; he consultado algunas bases de datos especializadas; he analizado las palabras claves de los artículos de las publicaciones seriadas del Colegio de la Frontera Norte, etc. y en especial he consultado los tesauros en línea y en papel.

Fue necesario también una comparación con los catálogos en línea de las bibliotecas universitarias de España, realizando la misma búsqueda con el termino frontera.

En especial, he realizado una comparación con los términos utilizados en el análisis de los artículos que se refieren a la frontera de Uruguay con Brasil por ser una frontera bastante estudiada.

En primera instancia nos enfrentamos a dos situaciones diferentes. Por un lado, el análisis de los contenidos de una colección que abarca el tema frontera pero que no es especializada. Y por otro, el análisis de las colecciones especializadas en la temática fronteriza.

Naturalmente que si tenemos una colección especializada en frontera, el termino perderá eficacia. Sin embargo, la participación de las instituciones fronterizas en Internet hace necesario repensar los términos seleccionados para uso interno o de los especialistas de la institución que genera la información. La difusión de esa información especializada debería condicionar la elección de los términos que representen una temática tan particular que es consultada no solamente por los usuarios que la generan sino también por otros y por lo tanto exige que vuelva a aparecer el termino o la expresión que identifique el concepto frontera para asegurar que esa información pueda ser recuperada.

Debemos agregar a toda la problemática planteada la necesidad de incorporar al análisis temático las imágenes, el sonido, toda la información multimedia ya que en este caso, tiene una relevancia esencial en la recuperación de la información.

A continuación mostraré, a modo de ejemplo, el termino frontera y sus casi sinónimos, encontrados con más frecuencia.

Frontera (s)

Línea fronteriza

Línea

Límite (s)

Límites territoriales  
Representaciones fronteras  
Región(es) frontera(s)  
Frontera física  
Demarcación frontera  
Término no usado zona frontera  
Franja frontera  
Frontera nortea

De esta lista debemos diferenciar el caso de “Representaciones fronteras” ya que esta referido a una situación que va mas allá de la geográfica y contempla aspectos ideológicos.

En los casos estudiados, nos encontramos que una misma colección puede tener uno o varios de estos términos para indicar documentos de la misma índole. La falta de coherencia en el uso de estos términos es una constante en la documentación analizada.

La recuperación utilizando estos términos es de muy baja especialización. Se recuperan diversos temas y es necesario repetir la búsqueda con los términos específicos que hayamos detectado. Tampoco se establecen relaciones que nos remitan a documentos que puedan ampliar o completar la información.

De esta forma, debido a que el término frontera esta referido a los aspectos geográficos en la mayoría de los análisis, llegaremos a través de una maraña de información, a descubrir los documentos que se refieren al término frontera en cuanto problema teórico. La polisemia del término en el lenguaje natural, se traslada a su representación en los lenguajes controlados.

Cuando el término frontera es usado en forma mas abiertamente simbólica, la recuperación de la información se hace aun más compleja.

En todas las búsquedas realizadas, hemos recuperado documentación referida a este uso del término. Sin duda que existe un uso muy extendido con este sentido del término.

Podemos mencionar como ejemplo algunos títulos: fronteras de la epistemología, la fertilidad asistida: una senda sin fronteras, la frontera olvidada: la ideología.

Otra forma de presentación del tema es a través del nombre de las ciudades. Aparecen directamente sin la palabra frontera lo que impli-

ca que el usuario realice su búsqueda a través de nombres de ciudades que debe conocer de antemano. Otra curiosa forma es la relación establecida por un guión que une los nombres de las dos ciudades. Ese guión representa la compleja trama de relaciones que se establecen entre los habitantes de una y otra ciudad.

Como ejemplo mencionaremos:

Mexicali-Calexico, Tijuana-Chula Vista

También se utilizan los nombres de los países fronterizos. Y así encontramos: frontera norte de México, frontera México-Estados Unidos.

Hay que tomar en cuenta también que la mayoría de los programas para la elaboración de catálogos o las bases de datos, recuperan las palabras significativas del título.

Por lo tanto, los campos de descriptores están disponibles para profundizar en el análisis temático. No siempre se toma en cuenta esta situación y aparece el termino frontera en los dos niveles.

Ahora bien, los habitantes de la frontera y los especialistas que estudian la temática han generado un lenguaje de especialidad, una terminología que forma parte de su lenguaje natural de comunicación y a su vez, permite la denominación de los objetos estudiados. En este caso, frontera o las expresiones mencionadas que se usan para designar el espacio geográfico es el único nivel que alcanza el acuerdo conceptual para la designación de los fenómenos que caracterizan la región.

Provenientes del lenguaje natural, de las denominaciones que utilizan los miembros de esta comunidad para comunicarse, se incorporan a los lenguajes controlados facilitando la comunicación en el nivel de los productores intelectuales de la información.

El segundo nivel de acuerdo conceptual que he registrado esta referido al adjetivo fronterizo. En este caso, casi no aparece en los lenguajes controlados. El Macrothesaurus de la OECD es casi la excepción y relaciona frontera con migración fronteriza.

El éxito de este término que acompaña los más diversos sustantivos, se encuentra en la recuperación de la información en lenguaje natural.

En este nivel, podemos encontrar que este adjetivo puede representar tanto la problemática sociocultural de la región fronteriza como la vinculación simbólica con algunos temas teóricos contribuyendo a la búsqueda de más conocimiento sobre las realidades de la frontera.

## CONCLUSIONES

Esta investigación no está concluida en la medida en que estos elementos que he analizado permiten llegar solamente al diagnóstico de la situación terminológica referida a la frontera. La organización de los conceptos de especialidad se hace imprescindible.

El paso siguiente sería la elaboración de un tesoro que permita el análisis de la información fronteriza, la asignación de representaciones conceptuales con acuerdo de todos los participantes en el proceso y que facilite el intercambio y la difusión de la información.

Sería muy interesante que este trabajo pudiera realizarse en forma conjunta con los generadores de la información, los especialistas en el estudio de la problemática fronteriza, los profesionales de la información, los lingüistas y terminólogos y también los comunicólogos que difunden la información hacia el conjunto de la sociedad.

## BIBLIOGRAFÍA CONSULTADA

Barité, Mario (2000) La terminología: implicaciones y aplicaciones respecto a los lenguajes documentales. Consultado en: <http://docdigital.upf.es/digital/aula2000/aula.htm>

Tejeda González, José Luis. *La línea, la frontera y la modernidad*. Estudios Fronterizos, Vol.5, no.10, 2004.

Cueva Perus, Marcos. Fronteras y representaciones fronterizas. *Estudios Fronterizos*, Vol.6, no.11, 2005.

***Primer Simposio Internacional sobre Organización...***

Melo, José Luiz Bica de. *Fronteiras: da linba imaginária ao campo de conflitos*. *Sociologias*, Porto Alegre, ano 6, nº 11, 2004.

Macrothesaurus OECD consultado en línea:  
<http://168.96.200.17/ar/oeed-macroth/es/>

Eurovoc consultado en línea:<http://europa.eu/eurovoc/> Tes-  
auro OIT consultado en línea: [http://www.ilo.org/public/  
libdoc/ILOthesaurus/spanish/](http://www.ilo.org/public/libdoc/ILOthesaurus/spanish/)

Tesauro de la UNESCO consultado en línea: [http://databases.  
unesco.org/thessp/](http://databases.unesco.org/thessp/)



## El vocabulario controlado de la base de datos IRESIE: Estrategias para su construcción

MARÍA ÁNGELA TORRES VERDUGO  
ROSA MARÍA ORIOL MUÑOZ  
*Universidad Nacional Autónoma de México*

### RESUMEN

Se reseña la evolución del Vocabulario Controlado (VC) que ha servido para que, por un lado, los documentalistas indiquen cada uno de los 59 000 registros de artículos que actualmente contiene la base de datos “Índice de Revistas de Educación Superior e Investigación Educativa” (IRESIE) y por otro, para que el usuario lo utilice en el proceso de búsqueda y selección de la información. En el documento se refleja el desarrollo del VC, el cual ha estado ligado a los vaivenes del IRESIE; no sólo en relación con los cambios de adscripción institucional sino también con la disposición de recursos económicos, la cobertura geográfica e idiomática de la base.

Se revisan su integración, a partir de la selección de los descriptores de los siguientes tesauros: el de la Educación de la Unesco, el del ERIC (Educational Resources Information Center), y el Tesauro Colombiano de la Educación, el impacto en el VC por los cambios en la temática de los artículos, relacionados con los debates propios de los países en vías de desarrollo, nuevos enfoques educativos y psicológicos y el surgimiento de nuevas subdisciplinas cuando se incrementaron el número de revistas procedentes de México, América Latina y

España. También se analiza el reflejo del desarrollo de las tecnologías de información, la telemática, comunicación satelital y la red de Internet que se constituyó en la vía de difusión de las base de datos, marco de referencia que obligó a encontrar otras estrategias para la interacción directa usuario-base, por lo que el VC se publicó en forma electrónica en la página web del IRESIE ([www.iisue.unam.mx/iresie/](http://www.iisue.unam.mx/iresie/)).

Finalmente se reflexiona sobre los procesos cognitivos inherentes al trabajo de análisis documental como son: percepción, atención, retención, concentración, asociación de ideas, generalización y discriminación de conceptos.

## EL VOCABULARIO CONTROLADO DEL ÍNDICE DE REVISTAS DE EDUCACIÓN SUPERIOR E INVESTIGACIÓN EDUCATIVA (IRESIE)

### Marco referencial.

Los autores que han trabajado sobre esta temática, dan distintas denominaciones a los lenguajes utilizados en el proceso de indización, Emilia Currás los denomina lenguaje documentario, María Texia Iglesias los llama lenguajes de indización, Blanca Gil, los denomina lenguajes documentales. Sin embargo hay coincidencias al definir los “lenguajes documentales” como lenguajes artificiales, extraídos de los lenguajes naturales pero que aplican ciertas normas convencionales.

Gil Urdiciain, B. (1996) define lenguaje documental como “...todo sistema artificial de signos normalizados, que facilitan la representación formalizada del contenido de los documentos para permitir la recuperación, manual o automática, de información solicitada por los usuarios” (p.18)

Estos lenguajes forman un subconjunto de términos especializados o lenguaje técnico extraídos de comunicaciones científicas de una rama del conocimiento, son dinámicos y se utilizan para indizar documentos en bases de datos especializadas, más aún, como afirma Naumis Peña, C (2007) “la construcción de lenguajes documentales garantiza mayores niveles de relevancia en la recuperación de información porque se restringe el uso de las unidades léxicas definidas

o redefinidas dentro de un sistema de información en particular”. (p.88)

Hay diversos enfoques para clasificar a los lenguajes documentales; Slype (1991) citado por Naumis Peña, C. (2007, p.102) los tipifica en dos categorías: lenguajes de clasificación y lenguajes de indización. Los primeros constituyen las clasificaciones bibliográficas utilizadas generalmente para catalogar los libros y los segundos corresponden a los lenguajes de indización, que a su vez se categorizan como: libres y controlados.

Los *lenguajes libres*, están integrados por términos no pre-definidos, se van generando a partir de la realización del proceso de indización, son los llamados descriptores libres y las listas de palabras clave, en este caso no son propiamente lenguajes documentales, pues no están controlados, sirven para la comunicación de una comunidad científica.

Los *lenguajes controlados* son listados de vocablos normalizados cuyo objetivo es potenciar los procesos de indización, consulta y recuperación de la información contenida en diversos tipos de documentos, así como facilitar el proceso de transformar la información en una verdadera fuente de conocimiento.

El lenguaje libre tiene como ventaja que es ilimitado, más específico, pero propicia abundancia de vocablos, imprecisiones, polisemias, su uso lleva a inconsistencias en la indización, favorece la dispersión y aumenta la ambigüedad semántica.

El lenguaje controlado en cambio, disminuye la ambigüedad semántica, es más consistente en la representación de los contenidos, facilita la búsqueda más amplia, controla sinónimos, homógrafos, establece una estructura de términos relacionados semánticamente, proporciona al usuario puntos de acceso más precisos, pero en las búsquedas se puede perder alguna información si la indización no fue adecuada.

En los sistemas que utilizan *lenguaje controlado* el esfuerzo mayor está en la fase de análisis e indización, en cambio, en los sistemas con lenguaje libre está en la búsqueda y recuperación de la información. En general, en las bases de datos se combinan la utilización de los dos tipos de vocabularios en el tratamiento documental.

Los lenguajes documentales pueden ser pre-coordinados o post-coordinados, los primeros combinan los términos en el momento de la catalogación: es el caso de las clasificaciones y las listas de encabezamientos de materia. En cambio, los post-coordinados combinan los términos en el momento de la consulta; en este caso los conceptos que se extraen en la indización para expresar el contenido temático del documento tienen todos la misma categoría, no se expresan en un orden determinado y en la fase de la recuperación, los usuarios hacen las combinaciones que responden a su necesidad de información.

Respecto a su estructura, los lenguajes documentales pueden ser jerárquicos, combinatorios y sintácticos, en los primeros los conceptos se distribuyen en clases o categorías, dependen unos de otros, van de lo general a lo particular, los lenguajes de estructura combinatoria asocian o combinan libremente los conceptos para expresar el tema o los temas, a la estructura sintáctica pertenecen los lenguajes que recurren a una sintaxis, para representar los contenidos de los documentos.

### La base de datos IRESIE: sus antecedentes

El IRESIE se desarrolla en el Instituto de Investigaciones sobre la Universidad y la Educación (IISUE), antes Centro de Estudios sobre la Universidad (CESU). Durante los primeros veinte años se desarrolló como una actividad más de la biblioteca y para el análisis documental, se tenía que contratar por honorarios a documentalistas externas con el apoyo económico que durante los años 1979-1984, 1993-1995, 1999-2001 se obtuvo del CONACyT y del Programa Interinstitucional de Investigaciones sobre Educación Superior (PIES), por lo que en las temporadas que no se tenía financiamiento externo, se suspendía.

El objetivo del IRESIE es apoyar a la comunidad académica nacional e internacional en las actividades de investigación, docencia y planificación de la educación, facilitando el acceso oportuno y a los artículos publicados en revistas académicas sobre educación, publicadas en idioma español y portugués que se reciben en las principales bibliotecas de las instituciones educativas de la ciudad de México.

La base de datos **IRESIE** ofrece el acceso a las referencias y resúmenes de más de 59 000 artículos generados en el campo de la

educación y publicados en alrededor de 690 títulos de revistas que proceden de México y otros países Iberoamericanos. La base de datos histórica, que contiene registros en inglés (hasta 1987) y español incluye 93000 registros.

El IRESIE aborda todos los aspectos relacionados con la educación: currículo, evaluación, investigación, enseñanza, aprendizaje, formación de profesores, administración, política educativa, economía de la educación, etc. en sus diferentes niveles educativos: superior, medio-superior, medio, básico, preescolar y en las distintas modalidades y tipos de educación: formal, no formal, continua, abierta, a distancia, especial, de adultos, etcétera.

### Vocabulario Controlado: características

El *Vocabulario Controlado del IRESIE* consiste en una lista normalizada de descriptores e identificadores, es un lenguaje de indización, su estructura es combinatoria y es postcoordinado. Cumple con un doble objetivo, por un lado sirve como un instrumento de trabajo para los documentalistas en el proceso de análisis e indización de los artículos que ingresan a la base de datos IRESIE y por otro, sirve a los usuarios como una herramienta para la identificación de los temas precisos a utilizar en el proceso de consulta y recuperación de la información.

EL VC incluye actualmente alrededor de 1600 *descriptores o temas* que cubren variados aspectos del campo de la educación y disciplinas complementarias, además de *identificadores*: integrados por nombres de países, regiones, niveles educativos, instituciones, nombres de grandes pensadores, teóricos, filósofos de la educación y pedagogos, niveles educativos etc.

Está organizado en forma alfabética, todos los temas o descriptores tienen la misma categoría. Si un tema no es descriptor preferido se envía con la palabra “véase” al término normalizado. Para orientar visualmente a los documentalistas y a los usuarios de la base de datos, se diferencia tipográficamente, escribiendo el tema o descriptor preferente con letras mayúsculas y el descriptor no utilizado con letras minúsculas,

*Primer Simposio Internacional sobre Organización...*

Ejemplo: Educación primaria *véase* EDUCACIÓN BÁSICA  
Grupos marginados *véase* MARGINACION SOCIAL  
Para el caso de descriptores relacionados se utiliza la expresión  
*véase también* PROBLEMAS AUDITIVOS *véase también* SORDOS  
EXÁMENES *véase también* PRUEBAS DE APROVECHAMIENTO

Los descriptores pueden ser:

- **Simples**, constituídos por una sola palabra que expresen un concepto, los cuales se pueden combinar para la búsqueda de información:

ADMINISTRACIÓN      UNIVERSIDADES  
CURRÍCULO              ESCUELAS

- **Compuestos** por dos o más palabras en orden natural  
ENSEÑANZA DE LA HISTORIA  
ESTRATEGIAS DE ENSEÑANZA

El uso del singular o plural se basa en los siguientes criterios:

- **El singular** se aplica a los nombres de conceptos abstractos como: disciplinas, propiedades, fenómenos, procesos etc.

PSICOLOGIA  
PRÁCTICA PROFESIONAL

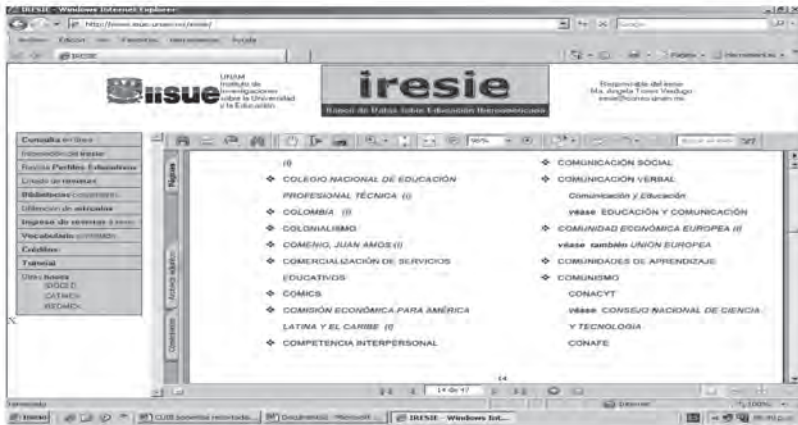
- **El plural** se utiliza para objetos que pueden contarse: equipos, sustancias, etc.

INSTITUCIONES EDUCATIVAS  
MATERIALES DIDACTICOS

Entre los criterios para la inclusión de nuevos descriptores están: la frecuencia de uso como lenguaje natural por parte de los autores de los artículos, las demandas de los usuarios detectadas a través del servicio de Búsquedas Bibliográficas y el surgimiento de temas emergentes: por ejemplo GÉNERO, ESTUDIOS DE GÉNERO.

Se han hecho tres versiones impresas: 1994, 1996 y 2007. También se puede consultar a través de la página WEB del IRESIE [www.iisue](http://www.iisue).

unam.mx/iresie en la liga de Vocabulario Controlado, en el cual se ofrece una breve introducción, datos conceptuales, ejemplos para su uso y el Vocabulario en orden alfabético.



Este VC controlado es utilizado también para la indización de otras bases de datos como las que se desarrollan en el Instituto Latinoamericano de la Comunicación Educativa y en el Centro de Estudios Educativos.

Su compilación ha estado determinada por varios factores, entre ellos los cambios de adscripción institucional, la modificación de la cobertura geográfica e idiomática de las revistas, la disposición de recursos económicos para la contratación de recursos humanos; así como el avance tecnológico.

En los años iniciales del proyecto, el VC fundamentalmente se integró con vocablos extraídos del Tesoro de la Educación de la UNESCO - OEI, edición 1977, el Tesoro Colombiano de la Educación y del Tesoro del ERIC (Educational Resources Information Center). A lo largo de los años se fue enriqueciendo con nuevos descriptores extraídos de los propios textos analizados, que respondían a las necesidades de la literatura publicada en México y Latinoamérica, en concordancia con la evolución de los conceptos en el campo educativo y el surgimiento de nuevas subdisciplinas. En algunas ocasiones, estos vocablos se adaptaron a la terminología utilizada en México.

Es importante anotar que de 1979 a 1988 la base de datos contenía un 70% de artículos en idioma inglés, un 26 % en español y el resto en portugués y francés, en vista de que las revistas que se analizaban procedían fundamentalmente de Estados Unidos y de los países europeos, en tanto que México y América Latina estaban poco representados. Sin embargo la consulta de la información por parte de los usuarios oscilaba entre un 60 a 65 % de artículos en español y de un 35 a 40 % en inglés. (Reportes técnicos enviados al CONACyT 1980-1985). Estos datos constituyeron un elemento importante en la toma de la decisión del cambio de cobertura idiomática de la base de datos del inglés a español.

Al desaparecer el Centro Universitario de Tecnología Educacional para la Salud (CEUTES) en febrero de 1989, el IRESIE fue transferido al Centro de Investigaciones y Servicios Educativos (CISE), cuya función principal era impartir formación docente a los maestros de las diferentes Facultades y Escuelas de la Universidad y de manera particular a los profesores del nivel bachillerato: de las Escuelas Preparatorias y de los Colegios de Ciencias y Humanidades.

En el CISE se desarrolló el IRESIE de 1989 a 1997, incrementándose su contenido principalmente con artículos que abordaban los diferentes aspectos de la educación superior y media-superior. A partir de 1989 se indizaron únicamente revistas en español y portugués, procedentes de México, América Latina, España y Portugal y algunos títulos publicados por organismos internacionales de Estados Unidos, Alemania y Francia. En este período se optimizó el proceso documental incluyendo un breve resumen, además de los datos de descripción bibliográfica.

La base de datos se transfirió a MICRO CDS ISIS y los campos correspondientes a temas e identificadores ya no estaban codificados, por lo que se tenían que capturar las palabras o frases completas, metodología que trajo consigo errores tipográficos, y ortográficos, sumados a problemas de atención y percepción por parte de los documentalistas, que ocasionaron duplicaciones conceptuales

Por ejemplo: CALIDAD EDUCATIVA en lugar de CALIDAD DE LA EDUCACION

ESTRATEGIAS DE APRENDIZAJE en lugar de ESTRATEGIAS DEL APRENDIZAJE



Al desaparecer el CISE en 1997, el IRESIE fue adscrito al anterior CESU ahora IISUE, en esta dependencia el proyecto se consolidó administrativamente, pasó de ser un programa desarrollado dentro de la biblioteca, sin personal académico propio, a constituirse en una Coordinación integrada por tres Técnicos Académicos, una secretaria y un bibliotecario dependiente de la Secretaría Académica.

### Factores que motivaron la revisión del VC

Los factores que motivaron la revisión del VC fueron diversos, entre ellos se pueden mencionar los siguientes:

El impacto de los temas abordados en las revistas en español y portugués que reemplazaron a las revistas en inglés que contenían un buen número de artículos de carácter técnico y teórico con enfoques positivistas, en tanto que aquellas que procedían de México y América Latina trataban problemáticas educativas propias de la región (marginación, pobreza, alfabetización, migración) con enfoques más reflexivos.

El cambio de políticas y contexto institucional, se pasó de un Centro que ofrecía formación didáctica para personal del área de la salud (CEUTES), a un Centro cuyas funciones principales eran la formación de docentes de educación media-superior y superior, en donde además se llevaban a cabo investigaciones sobre la didáctica crítica y se ofrecían servicios educativos a una población interdisciplinaria el IRESIE se fortaleció en educación media-superior y superior lo cual también se reflejó en el VC.

El no contar con personal académico adscrito al proyecto, obligó a incorporar alumnas de servicio social para el análisis documental, que si bien procedían de carreras del área de Pedagogía, requerían de capacitación en el proceso documental y de supervisión que no siempre se podía llevar a cabo; además de que este proceso era recurrente porque ya que estaban capacitadas terminaba su ciclo de servicio y se iban, y volvía a comenzar la capacitación con nuevas alumnas.

Otro factor determinante fue el hecho de que se procesaron en dos años 14 000 artículos, gracias al financiamiento que se recibió de parte del CONACYT entre 1999-2002, que permitió la contratación de

cuatro documentalistas por honorarios. Este procesamiento de gran cantidad de registros por parte de un equipo nuevo de documentalistas y la “novatez” de los estudiantes de servicio social, trajo muchas inconsistencias que precisaron a realizar una nueva revisión del VC.

## Metodología para la Revisión del Vocabulario Controlado

Se han llevado a cabo dos revisiones del VC, la primera durante los años 1994-1995, y la segunda entre los años 2004-2005. En cada una de ellas se integraron sendos equipos de trabajo interdisciplinarios formados por pedagogos, psicólogos educativos y una bibliotecaria. También se contó con herramientas de apoyo bibliográfico: el Tesauro de la Educación de la UNESCO, el del ERIC, diccionarios especializados, además del acervo de la biblioteca.

Se establecieron algunos criterios para realizar el proceso, uno de ellos fue el de preferir el vocablo de uso más frecuente por parte de los autores o de los documentalistas y que por consiguiente tenía más ocurrencias; que al revisar el vocabulario se fueran detectando los sinónimos susceptibles de eliminarse y direccionar con la palabras “véase” al término lingüístico más adecuado o los términos relacionados, (término genérico o término específico) para referirlos con la frase “véase también” en vista de que los usuarios a quienes está dirigido el IRESIE lo entienden con mayor facilidad que con las abreviaturas UF, TR, NT o BT, que únicamente lo manejan los documentalistas; como bien dice Catalina Naumis “la característica fundamental de la lengua es que tanto el emisor como el receptor deben compartir los marcos de referencia, porque si no poseen un conocimiento adquirido directamente de la realidad de referencia no puede haber entendimiento” (2007, p.90)

El proceso de investigación se llevó a cabo de la forma siguiente:

- Se elaboraron listados del VC a partir del “Índice temático” de la base de datos, los cuales se entregaron a cada uno de los participantes.
- Cada miembro del equipo revisó la pertinencia o no de los descriptores, cuando había dudas se imprimían las referencias bi-

bliográficas que habían sido indizadas con ese descriptor, para analizarlos a partir de los títulos de los artículos y los resúmenes, e incluso en algunas ocasiones fue necesario revisar las fuentes originales.

- Se programaron reuniones periódicas para exponer los avances y resolver los problemas conceptuales que se encontraban al revisar y valorar cada uno de los temas e identificadores.

Como producto de este trabajo de valoración:

- Se eliminaron los descriptores ambiguos, algunos sinónimos o cuasi sinónimos.
- Se incluyeron nuevos descriptores fundamentalmente referidos a las nuevas tecnologías de la información o sobre temas emergentes.
- Se unificaron bajo un mismo descriptor aquellos términos cuyo significado era el mismo, pero estaban con diferente grafía.
- Se identificaron “los descriptores” afines para referir al usuario con las palabras “véase también” a otros descriptores relacionados.
- Se identificaron los “no descriptores” para referir al usuario con la palabra “véase” a los descriptores usados en el VC.
- Se determinaron los términos que requerían definiciones y notas de alcance.
- Se eliminaron los términos poco utilizados (menos de cinco). En estos casos se revisaron cada uno de los registros y se indizaron bajo otros descriptores.
- Se inició la integración de un glosario con los vocablos que requerían definiciones.
- Se publicó la primera versión impresa del VC del IRESIE en el año de 1994, una segunda en 1996 y una tercera edición en el presente año.
- A partir del año 2003 se publicó el *Vocabulario Controlado* en la página web del iresie [www.iisue.unam.mx/iresie](http://www.iisue.unam.mx/iresie).
- Se corrigieron en la base de datos los errores tipográficos, ortográficos y en algunos casos también de asignación errónea de los descriptores.

Para realizar la segunda revisión y valoración de los años 2004-2005 se aprovechó la facilidad de consulta del IRESIE en internet, realizando *búsquedas libres* o búsquedas mediante el *operador de adyacencia \$*, y la búsqueda en el *índice de temas*. Las dos primeras modalidades permiten buscar un vocablo o concepto, en cualquier campo del registro, es decir en el título del artículo, temas o en el resumen, en tanto que la última modalidad sólo busca en el *Índice de Temas*

### Procesos cognitivos inherentes al proceso de análisis documental

El trabajo cotidiano de analizar la información publicada en las revistas indizadas en el IRESIE, implica hacer una lectura rápida del documento (artículo, reporte de investigación, entrevista, entre otros) para identificar las ideas principales, leyendo el título, subtítulos, introducción y conclusiones. Después hay que anotar los datos de identificación bibliográfica de la revista y del artículo en los formatos de captura de la base de datos diseñada *ex-profeso*, elaborar un resumen de su contenido y posteriormente anotar los descriptores e identificadores.

Con frecuencia ocurre que al buscar en el VC los términos que el autor expone, no se encuentran, así que en ese caso, se recurre a otras revistas especializadas, a otras bases de datos y al propio IRESIE para verificar la frecuencia con que este “nuevo” término es citado, mediante una búsqueda por operadores booleanos, utilizando el operador de adyacencia obligada \$ (signo de pesos para el IRESIE). También se examina en el mismo artículo la definición que el autor ofrece, para ver si existe un término sinónimo. En este caso, es preferible utilizar el que ya está aceptado y no crear uno nuevo, sobre todo si la frecuencia con la que se cita es baja.

Como este proceso implica el dominio de los conceptos, es decir, se tiene que tener claro el contenido semántico del término, se recurre también a la consulta con expertos, que puede ser a través de los libros o documentos que definen los términos o consultándolos personalmente.

Según el Diccionario de Pedagogía y Psicología (1999), *cognición* es “el conjunto de estructuras y actividades psicológicas cuya función es el conocimiento, por oposición a los dominios de la afectividad. Término general empleado por los psicólogos para designar cualquier actividad mental, que engloba el uso del lenguaje, el pensamiento, el razonamiento, la solución de problemas, la conceptualización, el recuerdo y la imaginación.” *Cognoscitivo* es el “adjetivo referido al proceso mental de comprensión, juicio, memoria y razonamiento, en contraste con los procesos emocionales y volitivos”.

En el proceso de tratamiento documental de los artículos de las revistas que se realiza para ingresar los registros a la base de datos IRE-SIE, existe una función muy importante que es la asignación de palabras claves, las cuales se dividen en dos categorías: *a) temas* en donde se anotan un máximo de cinco términos o descriptores, los cuales representan breve y objetivamente el tema central y los subtemas del artículo analizado; y *b) identificadores*, se asignan un máximo de tres términos y tienen la función de contextualizar la información contenida en el artículo, por ejemplo: país, tipo de trabajo, nivel educativo o población a la que se refiere el autor; también en esta categoría se incluyen los nombres de autores clásicos en el área educativa: Freire, Pablo; Piaget, Jean.

La precisión en la asignación de estos términos van a determinar la calidad de la base de datos e influirá en la eficiencia del funcionamiento de la base a la hora de la recuperación de la información. Para estandarizar los términos a usar como palabras claves, se elabora y usa el VC.

La actividad de seleccionar y asignar los descriptores e identificadores es un trabajo intelectual, que requiere de la activación de procesos cognitivos por parte del documentalista. Algunos de estos procesos son: atención, retención, percepción, concentración, asociación de ideas y generalización de conceptos. Lógicamente estos procesos están relacionados entre sí.

A continuación se señalan a que se refiere cada uno de éstos y como se vinculan a la actividad de asignar los descriptores e identificadores:

**Percepción.** Es el proceso en el cual el organismo recibe información del medio ambiente en forma de estímulos que son registrados por procesos sensoriales a través de los sentidos. En la actividad de asignar descriptores o temas, se vincula, porque el documentalista tiene que hacer una “igualación a la muestra”, que implica ver (percibir) los términos tal y como están en el VC, por ejemplo: CALIDAD EDUCATIVA por CALIDAD DE LA EDUCACIÓN. Aunque los dos descriptores signifiquen semánticamente lo mismo, para efectos de estandarización de los términos y para los sistemas lógicos de la computadora no es lo igual. Se deberá seleccionar *CALIDAD DE LA EDUCACIÓN*, que es el descriptor aceptado dentro del VC.

**Atención.** Es la percepción selectiva de estímulos que en su conjunto transmiten información. Este proceso tiene capital importancia pues además de hacer la igualación a la muestra, se tiene que discriminar a lo largo del texto de los artículos, donde aparecen estos términos para poder asignarlos en los campos de descriptores o identificadores.

**Retención o memoria,** es el proceso que se ocupa de lo que se ha almacenado y de la forma como lo recuperamos o lo recordamos. Este proceso también está vinculado con la formación de conceptos la cual se define como la representación mental de un conjunto de objetos, hechos o entidades, la capacidad de los sujetos para organizar el conocimiento en categorías, con base en sus características comunes, es una actividad de gran relevancia en percepción, memoria y pensamiento. La forma como haya almacenado los términos, le permitirá al documentalista ser preciso a la hora de asignarlos en forma de descriptores o identificadores

Esto también implica una formación previa en carreras relacionadas con el tema de la especialización del banco de datos, en nuestro caso: la educación. Además leer o conocer las definiciones, le permitirá tener el contenido semántico no de cada palabra que conforma el descriptor, sino del término en su conjunto. Por ejemplo:

Solución de problemas,	Ambientes virtuales,
Historia oral,	Historia de vida



Es por ello, que el análisis documental de una base de datos especializada como el IRESIE, debe ser realizado por personal profesional relacionado con el campo de la educación, que esté permanentemente actualizado y comprometido con su trabajo para que pueda desarrollar una actividad intelectual que redunde en la calidad de la base de datos.

## CONCLUSIONES

La compilación de un lenguaje especializado y su transferencia a una herramienta lingüística como lo es el Vocabulario Controlado constituye un aporte importante para las áreas disciplinarias que representa, en este caso el campo educativo, a través del análisis progresivo de los términos significativos, se puede incluso, estudiar el avance de una disciplina, sus enfoques, tendencias, corrientes, los temas emergentes.

La coparticipación de investigadores, docentes y usuarios en general con los documentalistas constituye un elemento clave en la compilación y revisión de cualquier VC. Igualmente la formación profesional y/o especialidad de los documentalistas y su actualización mediante la asistencia a seminarios, cursos, debates y otros eventos académicos del área educativa redundan positivamente en la calidad del VC y en el proceso de indización.

Finalmente es importante enfatizar que la interrelación con los usuarios mediante la impartición de talleres sobre la metodología para la elaboración de resúmenes o de desarrollo de habilidades para la consulta a bases de datos constituyen estrategias trascendentales para recibir la retroalimentación correspondiente de parte de los usuarios finales de un sistema de información.

## BIBLIOGRAFÍA

- Canda Moreno, F. coord. (1999). *Diccionario de pedagogía y psicología*. Madrid: Cultural, 376 p.



Ellis Henry, C. (1980). *Fundamentos del aprendizaje y procesos cognoscitivos del hombre*. México: Trillas, 309 p.

ERIC Clearinhose for Linguistics (1980). *Tesaurus of ERIC descriptors*. Phoenix: Oryx, 419 p.

Gil Urdiciani, B. (1996). *Manual de lenguajes documentales*. Madrid: Noesis, 264 p.

*Informe de la Coordinación del IRESIE 2005*. México: CESU, 12 p. Documento interno.

*IREsIE: reportes técnicos enviados al CONACyT: 1981-1985*. México: CEUTES, documentos internos.

Naumis Peña, C. (2007). *Los tesauros documentales y su aplicación en la información impresa, digital y multimedia*. México: Universidad Nacional Autónoma de México, Centro Universitario de Investigaciones Bibliotecológicas, 288 p.

Poveda Gómez, N. comp. (1978). *Tesaurus Colombiano de la Educación: versión preliminar*. Bogotá: Instituto Colombiano para el fomento de la Educación. Sección de Documentación en Educación, 173 p.

*Tesaurus de la Educación UNESCO: OIE*. (1991). 5 ed. París: UNESCO, 144p.

Torres Verdugo, Ma. A. (1995). *Control de calidad de la base de datos IRESIE*. México: Colima, 12 p. Mimeografiado.



# Directorio de autores

## CONFERENCISTAS

• *CARDERO, Ana María*

[acardero@servidor.unam.mx](mailto:acardero@servidor.unam.mx)

Seminario de Lexicología, Lexicografía  
y Terminología  
Facultad de Estudios Superiores Acatlán  
Universidad Nacional Autónoma de México  
MÉXICO

• *GELBUKH, Alexander*

[gelbukh@gelbukh.com](mailto:gelbukh@gelbukh.com)

Centro de Investigación en Computación  
Instituto Politécnico Nacional  
MÉXICO  
International Social Networks Research Center  
SLOVAKIA

• *LÓPEZ HUERTAS, María José*

[mjlopez@ugr.es](mailto:mjlopez@ugr.es)

Facultad de Biblioteconomía y Documentación  
Universidad de Granada  
ESPAÑA

• *LÓPEZ YEPES, José*

[yepes@ccinf.ucm.es](mailto:yepes@ccinf.ucm.es)

Departamento de Biblioteconomía y Documentación  
Facultad de Ciencias de la Información  
Universidad Complutense de Madrid  
ESPAÑA

*Primer Simposio Internacional sobre Organización...*

- *LUNA TRAILL, Elizabeth*  
[traill@servidor.unam.mx](mailto:traill@servidor.unam.mx)

Instituto de Investigaciones Filológicas  
Universidad Nacional Autónoma de México  
MÉXICO

- *ROSEMBLAT, Graciela*  
[grosemblat@mail.nih.gov](mailto:grosemblat@mail.nih.gov)

Biblioteca Nacional de Medicina,  
Institutos Nacionales de la Salud, Bethesda, MD  
UNITED STATES OF AMERICA

- *SIERRA MARTÍNEZ, Gerardo*  
[GSierraM@iingen.unam.mx](mailto:GSierraM@iingen.unam.mx)

Instituto de Ingeniería  
Universidad Nacional Autónoma de México  
MÉXICO

- *SMIRAGLIA, Richard*  
[Richard.Smiraglia@liu.edu](mailto:Richard.Smiraglia@liu.edu)

Long Island University, Brookville NY  
UNITED STATES OF AMERICA

- *VOUTSSÁS MÁRQUEZ, Juan*

Centro Universitario de Investigaciones Bibliotecológicas,  
Universidad Nacional Autónoma de México.  
MÉXICO

## PONENTES

- *ÁNGELES ESCOBAR, Noé*  
[nangeles@colmex.mx](mailto:nangeles@colmex.mx)

Biblioteca Daniel Cosío Villegas  
El Colegio de México.  
MÉXICO

- *AGUILAR-AMAT, Anna*  
[anna.aguilar-amat@uab.cat](mailto:anna.aguilar-amat@uab.cat)

Grupo Tradumática (Traducción y Nuevas Tecnologías  
Departamento de Traducción e Interpretación  
Universitat Autònoma de Barcelona  
ESPAÑA

- *ARENAS GRISALES, Sandra*  
[sarena@bibliotecologia.udea.edu.co](mailto:sarena@bibliotecologia.udea.edu.co)  
Escuela Interamericana de Bibliotecología  
Universidad de Antioquia, Medellín  
COLOMBIA
  
- *BARBERENA BLÁSQUEZ, Elsa*  
[elsa@servidor.unam.mx](mailto:elsa@servidor.unam.mx)  
Posgrado en Bibliotecología y Estudios de la Información  
Universidad Nacional Autónoma de México  
MÉXICO
  
- *BIDART ESCOBAR, Lil*  
[lilbidart@yahoo.com.mx](mailto:lilbidart@yahoo.com.mx)  
Coordinadora del Sistema de Bibliotecas  
Universidad de Tijuana, Baja California  
MÉXICO
  
- *BITTENCOURT DOS SANTOS, Fernando*  
[fernandobatuba@hotmail.com](mailto:fernandobatuba@hotmail.com)  
Departamento de Ciência da Informação  
Universidade Estadual Paulista “Júlio de Mesquita Filho”, Marília-SP  
BRASIL
  
- *BOCCATO, Vera Regina Casari*  
[ybocato@power.ufscar.br](mailto:ybocato@power.ufscar.br)  
Universidad Federal de San Carlos  
BRASIL
  
- *BOGOMILOVA LOZANOVA, Elena*  
[Ebogomil@gmx.net](mailto:Ebogomil@gmx.net)  
El Colegio de México  
MÉXICO
  
- *Do COUTTO, Mariza Leal de Meirelles*  
[matleme@carpa.ciagri.usp.br](mailto:matleme@carpa.ciagri.usp.br)  
División de Desarrollo e Innovación del  
Sistema Integrado de Bibliotecas  
Universidade de São Paulo - São Paulo, SP  
BRASIL

***Primer Simposio Internacional sobre Organización...***

• *EHNIS DUHNE, Érika*

[mariposa@servidor.unam.mx](mailto:mariposa@servidor.unam.mx)

Centro de Enseñanza de Lenguas Extranjeras  
y Centro de Enseñanza para Extranjeros,  
Universidad Nacional Autónoma de México  
MÉXICO

• *FOUNTAIN, Joanna F.*

[fountain@thegateway.net](mailto:fountain@thegateway.net)

Sam Houston State University, Huntsville, Texas  
UNITED STATES OF AMERICA

• *FUENTES AGUSTI, Marta*

[Marta.Fuentes@uab.cat](mailto:Marta.Fuentes@uab.cat)

Universitat Autònoma de Barcelona  
ESPAÑA

• *GARCÍA BARBABOSA, Ageo*

[garciaageo@yahoo.com](mailto:garciaageo@yahoo.com)

Biblioteca Howard-Tilton Memorial  
Tulane University, New Orleans, Louisiana  
UNITED STATES OF AMERICA

• *GARCÍA PEÑALVO, Francisco J*

[fgarcia@usal.es](mailto:fgarcia@usal.es)

Facultad de Informática y Automática  
Universidad de Salamanca  
ESPAÑA

• *GONCALVES MOREIRA TÁLAMO, María de Fátima*

[mfgmtala@usp.br](mailto:mfgmtala@usp.br)

Universidade de São Paulo - São Paulo, SP  
BRASIL

• *HERNÁNDEZ SERRANO, M. José*

[marijosehs@gmail.com](mailto:marijosehs@gmail.com)

Universidad de Salamanca  
ESPAÑA

• *IBARRA, Rafael*

[Ibarraraf@aol.com](mailto:Ibarraraf@aol.com)

Biblioteca Digital  
Dirección General de Bibliotecas  
Universidad Nacional Autónoma de México  
MÉXICO

- *KOBASHI, Nair Yumiko*

[nykobash@usp.br](mailto:nykobash@usp.br)

Escola de Comunicações e Artes  
Universidade de São Paulo - São Paulo, SP  
BRASIL

- *LEME, María Ángela de Toledo*

[matleme@carpa.ciagri.usp.br](mailto:matleme@carpa.ciagri.usp.br)

Division de Biblioteca y Documentacion  
Escola Superior de Agricultura “Luiz de Queiroz”  
Universidade de São Paulo - Piracicaba, SP  
BRASIL

- *LIMA, Vânia Mara Alves*

[vamal@usp.br](mailto:vamal@usp.br)

Escola de Comunicações e Artes  
Universidade de São Paulo - São Paulo, SP  
BRASIL

- *LÓPEZ GINEZ DE LARA, Marilda*

[larama@usp.br](mailto:larama@usp.br)

Universidade de São Paulo - São Paulo, SP  
BRASIL

- *LÓPEZ GUZMÁN, Clara.*

[clara@servidor.unam.mx](mailto:clara@servidor.unam.mx)

Dirección General de Servicios de Cómputo Académico  
Universidad Nacional Autónoma de México  
MÉXICO

- *MATAMOROS SEGURA, Lucía*

[imatamor@una.ac.cr](mailto:imatamor@una.ac.cr)

Centro de Información Documental de Ciencias Sociales (CIDCSO)  
Facultad de Ciencias Sociales  
Universidad Nacional (Costa Rica), Heredia  
COSTA RICA

- *MESA LAO, Bartolomé*

[barto.mesa@uab.cat](mailto:barto.mesa@uab.cat)

Grupo Tradumática (Traducción y Nuevas Tecnologías)  
Departamento de Traducción e Interpretación  
Universitat Autònoma de Barcelona, Barcelona  
ESPAÑA

*Primer Simposio Internacional sobre Organización...*

• *MONTOYA CORREA, Carlos Arturo*  
[carlosarturomontoya@bibliotecologia.udea.edu.co](mailto:carlosarturomontoya@bibliotecologia.udea.edu.co)  
Escuela Interamericana de Bibliotecología  
Universidad de Antioquia, Medellín.  
COLOMBIA

• *MORATO LARA, Jorge*  
[jmorato@ie.inf.uc3m.es](mailto:jmorato@ie.inf.uc3m.es)  
Departamento de Informática  
Universidad Carlos III de Madrid  
ESPAÑA

• *MOREIRO GONZÁLEZ, José Antonio*  
[jamore@bib.uc3m.es](mailto:jamore@bib.uc3m.es)  
Departamento de Biblioteconomía y Documentación  
Universidad Carlos III de Madrid  
ESPAÑA

• *MÚNERA TORRES, María Teresa*  
[mmuner@bibliotecologia.udea.edu.co](mailto:mmuner@bibliotecologia.udea.edu.co)  
Escuela Interamericana de Bibliotecología  
Universidad de Antioquia, Medellín  
COLOMBIA

• *ORIOI MUÑOZ, Rosa María*  
[iresie@correo.unam.mx](mailto:iresie@correo.unam.mx)  
Índice de Revistas de Educación Superior e Investigación Educativa  
Instituto de Investigaciones sobre la Universidad y la Educación  
Universidad Nacional Autónoma de México.  
MÉXICO

• *ORTIZ ANCONA, Dante*  
[dante@dgb.unam.mx](mailto:dante@dgb.unam.mx)  
Dirección General de Bibliotecas  
Universidad Nacional Autónoma de México.  
MÉXICO

• *PALACIOS, Vicente*  
[palacios@di.uc3m.es](mailto:palacios@di.uc3m.es)  
Departamento de Informática  
Universidad Carlos III de Madrid  
ESPAÑA



• *POZZI, María*

[pozzi@colmex.mx](mailto:pozzi@colmex.mx)

El Colegio de México

MEXICO

• *ROCHA PALMA, Cristina*

[rochacri2000@yahoo.com](mailto:rochacri2000@yahoo.com)

Escuela de Bibliotecología y Ciencias de la Información

Universidad de Costa Rica

COSTA RICA

• *ROJAS LEÓN, Gloria L.*

Consultores Nacionales Asociados

COLOMBIA

• *ROJAS MORALES, Nydia*

[nroj@una.ac.cr](mailto:nroj@una.ac.cr)

Escuela de Bibliotecología, Documentación e Información

Universidad Nacional de Costa Rica, Heredia

COSTA RICA

• *SABELLI, Martha*

[msabelli@adinet.com.uy](mailto:msabelli@adinet.com.uy)

Escuela Universitaria de Bibliotecología y Ciencias Afines,

Universidad de la República

Facultad de Administración y Ciencias Sociales

Facultad de Arquitectura

Universidad ORT

URUGUAY

• *SÁNCHEZ CUADRADO, Sonia*

[ssanhec@ie.inf.uc3m.es](mailto:ssanhec@ie.inf.uc3m.es)

Departamento de Informática

Universidad Carlos III de Madrid

ESPAÑA

• *SÁNCHEZ GIJÓN, Pilar*

[pilar.sanchez.gijon@uab.cat](mailto:pilar.sanchez.gijon@uab.cat)

Grupo Tradumática (Traducción y Nuevas Tecnologías)

Departamento de Traducción e Interpretación

Universitat Autònoma de Barcelona, Barcelona

ESPAÑA

*Primer Simposio Internacional sobre Organización...*

- *SILVA CASARÍN, Helen de Castro.*  
[helenc@marilia.unesp.br](mailto:helenc@marilia.unesp.br)

Departamento de Ciência da Informação  
Universidade Estadual Paulista  
"Júlio de Mesquita Filho" Marília -SP  
BRASIL

- *TORRES VERDUGO, Ma. Ángela*  
[angelat@servidor.unam.mx](mailto:angelat@servidor.unam.mx)

Índice de Revistas de Educación Superior e Investigación Educativa  
Instituto de Investigaciones sobre la Universidad y la Educación  
Universidad Nacional Autónoma de México.  
MÉXICO

- *URIBE TIRADO, Alejandro*  
[auribeti@eafit.edu.co](mailto:auribeti@eafit.edu.co)

Escuela Interamericana de Bibliotecología  
Universidad de Antioquia, Medellín  
COLOMBIA

- *VEGA MARTINEZ, Aurora*  
[aurora@tulane.edu](mailto:aurora@tulane.edu)

Biblioteca Howard-Tilton Memorial  
Tulane University  
7001 Freret St. New Orleans, Louisiana  
UNITED STATES OF AMERICA

- *VELÁZQUEZ-MONTES, Imelda*  
[ivm@servidor.unam.mx](mailto:ivm@servidor.unam.mx)

Servicios de Información Digital, Hemeroteca,  
Facultad de Química.  
Universidad Nacional Autónoma de México.  
MÉXICO

***Memoria del Primer Simposio Internacional sobre Organización del Conocimiento: Bibliotecología y Terminología.*** 27al 29 agosto de 2007. La edición consta de 300 ejemplares. Cuidado de la edición, Zindy Elizabeth Rodríguez Tamayo. Formación editorial, Christopher Barrueta Álvarez. Revisión especializada, Francisco Xavier González y Martha Castro López. Centro Universitario de Investigaciones Bibliotecológicas. Fue impreso en papel cultural ahuesado de 90 gr. en Producciones Editoriales Nueva Visión México, ubicados en Juan A. Mateos No. 20, Col. Obrera, D.F. Se terminó de imprimir en el mes de enero de 2009.