

Modelos teóricos y elementos funcionales para el análisis de contenido documental: definición y tendencias

JOSÉ ANTONIO MOREIRO GONZÁLEZ

Departamento de Biblioteconomía y Documentación
Universidad Carlos III, Madrid, España
E-mail: jamore@bib.uc3m.es

MIGUEL ÁNGEL MARZAL GARCÍA-QUISMONDO

Departamento de Biblioteconomía y Documentación
Universidad Carlos III, Madrid, España
E-mail: mmarzal@bib.uc3m.es

Artículo



RESUMEN

Los autores ofrecen un panorama de las orientaciones actuales en la investigación para tratar el contenido informativo documental, labor analítica muy afectada por el impacto de las nuevas tecnologías. En consecuencia, el método expositivo adoptado busca abordar cuatro campos de investigación analítica: a) la evolución y las más recientes vías investigadoras en los tres grandes modelos de análisis documental de contenido, con especial atención tanto a la representación probabilística y vectorial en el modelo estadístico, como a la inteligencia artificial y las redes neuronales en el cognitivo; b) la representación de la información a través de las transformaciones experimentadas por los tesauros, hasta derivar en los mapas conceptuales; c) las alteraciones impuestas por la Sociedad de la Información a la organización del conocimiento y, que por ende se proyectan en los lenguajes de clasificación, en plena evolución, y d) el análisis de las imágenes, en auge en la Era de la Información, como un nuevo discurso textual que exige un análisis diferente.

Palabras Clave: Análisis Documental, Recuperación de Información, Organización del Conocimiento, Lingüística Documental, Lenguajes de Clasificación.

MODELOS TEÓRICOS Y ELEMENTOS FUNCIONALES PARA EL ANÁLISIS DE CONTENIDO DOCUMENTAL: DEFINICIÓN Y TENDENCIAS

**JOSÉ ANTONIO MOREIRO-GONZÁLEZ
MIGUEL ÁNGEL MARZAL-GARCÍA-QUISMONDO**

ABSTRACT

The authors offer a view of the current orientations in the investigation dealing with documental informative content, an analytic work greatly influenced by the impact of new technologies. Consequently, the expositive method adopted tries to approach four fields of analytic investigation: a) The evolution and the

most recent lines of investigation in the three big models of documental analysis, with special attention to the probabilistic and vectorial representation in the statistical pattern, and to A.I. and neural nets in the cognitive. b) The representation of information, from the transformations experienced in thesauri to conceptual maps. c) The changes imposed by the Information Society in the organisation of knowledge and their influence in the languages of classification, in full evolution nowadays. d) The analysis of images, a new textual discourse of growing importance in the Age of Information, that demands a new approach.

Key Words: Document Analysis, Information Retrieval, Knowledge Organization, Documental Linguistics, Classification Languages.

INTRODUCCIÓN

Desde que las computadoras se aplicaron a la gestión y tratamiento de la información se observó que se abrían numerosas alternativas para la construcción y manejo de las herramientas técnicas que sirven para ordenar, clasificar y representar la información. Precisamente las investigaciones computacionales impulsaron desde los años setenta la búsqueda de modelos formales, y como consecuencia la aparición de unos modelos teóricos que han buscado convertirse en paradigmas de los sistemas de recuperación de información automáticos. Los modelos conjuntan, pues, teoría y práctica y pueden enfocarse desde una perspectiva múltiple,¹ cuyos principios y tendencias vamos a revisar aquí.

MODELOS LINGÜÍSTICOS

En el contexto de la comunicación humana, los discursos en los que emitimos los mensajes presentan la estructura propia del lenguaje natural en el que se realizan. Por ello el objetivo de los modelos lingüísticos consiste en eliminar la ambigüedad de las palabras filtrándolas a través de un procesamiento a distintos niveles para tratar de comprender el significado de los documentos. Esos niveles son el morfológico-léxico; el sintáctico; el semántico, y el pragmático. Estos modelos han tenido un largo desarrollo y han generado corrientes teóricas y aplicativas de conocida aceptación:

- ❖ *Lingüística o Ciencia del texto*: Propuesta por Van Dijk, considera que los textos presentan un triple nivel estructural: microestructura, macroestructura, y superestructura. Operativamente, el texto es sustituido por los enunciados de las macroproposiciones que expresan su macroestructura global y las macroestructuras parciales. Sirve para reconocer (comprender), analizar (identificar las

1 Los modelos proporcionan mecanismos de recuperación, pero atienden también a la forma en que la información debe formularse para que se adapte a los mecanismos de búsqueda, así, K. Sparck Jones.- Some thoughts on classification for retrieval, en *Journal of Documentation*, 1970, 26: 89-101. D. Ellis.- A behavioral model for Information Retrieval system design, en *Journal of Information Science*, 1989, 15: 237-247. , P. Ingwersen. - *Information Retrieval Interaction*. London: Taylor Graham: 1992.

estructuras significativas) y representar la información de los discursos.² Para hacer esto se precisa del funcionamiento concomitante de un modelo estratégico o situacional, que aborda el análisis desde los aspectos pragmáticos e incluye información acerca de las situaciones, intenciones, objetivos, fines y contexto en que los documentos se generan y se tratan.³

- ❖ *Modelo de esquemas*: propuesto a partir de la psicología cognitiva por Minsky y Rumelhart. Para entender una idea hay que comprender cada uno de los elementos que la integran. Los conceptos se consideran según sea su posición en los esquemas del conocimiento. Son de una complejidad enorme:⁴

1. Esquemas de conocimiento

- 1.1. Experiencias anteriores (actos, acontecimientos, causas, fines) = Enlaces pragmáticos.

- 1.2. Conocimiento de los discursos = Enlaces sintácticos y superestructurales.

- 1.3. Conocimiento especializado = Enlaces semánticos.

- 1.4. Además, la información nueva o remática.

2. Contexto sociocultural.

- ❖ *Procesamiento holístico*: el lector interpreta desde su experiencia sintáctica, semántica y pragmática la situación que el texto le ofrece. Efectúa saltos continuos desde su memoria al modelo de situación y al texto (*bottom-up y top-down*). Sus esquemas de conocimiento se activan ante la información aportada por el texto.⁵
- ❖ *Gramática de casos*, de Fillmore: los discursos tienen una situación concreta que da significado a los términos que los componen. Permite obtener esquematizaciones para construir vocabularios científicos y para estudiar las interrogaciones de los usuarios. Es útil como procedimiento aplicable tras la reducción sintética de las macroestructuras para identificar los componentes terminológicos relevantes de un texto.⁶

2 T. A. van Dijk.- *La ciencia del texto. Un enfoque interdisciplinario*. Barcelona: Paidós, 1983. Sus teorías se describen relacionadas con el análisis de contenido documental en José A. Moreira González. - *Aplicación de las Ciencias del texto al resumen documental*. Madrid: Universidad Carlos III de Madrid - Boletín Oficial del Estado, 1993.

3 T. A. van Dijk.- *Texto y Contexto. (Semántica y pragmática del discurso)*. Madrid: Cátedra, 1982.

4 D. Rumelhart. - Schemata: the building blocks of cognition, en R. Spiro(ed).- *Theoretical issues in reading comprehension: perspectives from cognitive psychology, linguistics, artificial intelligence and education*. Hillsdale: Erlbaum, 1985: 75-92. M. Minsky. - *A framework for representing knowledge*, en P. Winston (ed).- *The psychology of computer vision*. New York: McGraw-Hill, 1975: 217-228.

5 A. L. Brown.- *Metacognitive development and reading*, en R. Spiro (ed).- *Theoretical issues in reading comprehension: perspectives from cognitive psychology, linguistics, artificial intelligence and education*. Hillsdale: Erlbaum, 1985: 455-474.

6 C.J. Fillmore.- *Em favor do caso*, en , L.P. Lobato (ed.).- *A semântica da Lingüística Moderna: o léxico*. Rio de Janeiro: Livraria Alves, ...: 277-365. Buscando la producción de analizadores automáticos lo utiliza A. Garsía Gutiérrez.- *Procedimientos de Análisis Documental Automático. Estudio de caso*. Sevilla. Junta de Andalucía, Consejería de Cultura, 1996.

❖ *Procesamiento del Lenguaje Natural (PLN)*. Propone aproximaciones modulares al procesamiento del lenguaje natural (PLN) centradas en la integración de los diferentes subcomponentes que afectan a los niveles fonológico, morfológico, sintáctico, semántico y pragmático del discurso. A partir del texto de entrada, el proceso de análisis de texto produce un documento que lo representa en una forma que pueda ser interpretada por el ordenador para poder ser reutilizada en un proceso de recuperación de información. En la práctica, este procesamiento funciona cooperativamente con métodos probabilísticos. El estudio informatizado de cada uno de estos niveles tiene como fin originar una serie de analizadores o *parsers*:

a) *Análisis morfológico*: cubre desde el reconocimiento de los accidentes de cada palabra, hasta su análisis léxico, para alcanzar la lematización o reducción de las palabras a su término canónico.

Curiosamente, el análisis lingüístico se hace utilizando también modelos estadísticos. El análisis léxico, que tiene como objetivo transformar una cadena de caracteres en un conjunto de palabras o *tokens*, y que es la primera parte dentro del proceso automático de adquisición de componentes, se encarga de proporcionar los términos (posibles descriptores) para que sean posteriormente examinados por otros procesos (filtrados, palabras compuestas, etcétera). Lo primero que se ha previsto al desarrollar un analizador léxico es decidir cuáles son los caracteres o símbolos que no son interesantes y que en muchos casos sirven para delimitar una palabra: los espacios en blanco. Pero además se deben considerar otros caracteres especiales como los guiones, signos de puntuación, dígitos y letras. Generalmente los signos de puntuación se eliminan en el proceso de análisis léxico. Sin embargo, en ciertas ocasiones la eliminación de un signo de puntuación desvirtúa el resultado de una búsqueda.

Corresponde también a este nivel la eliminación de las palabras vacías. Los procesos de filtrado, con sus correspondientes cálculos estadísticos, actúan eliminando los términos vacíos mediante su confrontación con una lista de palabras vacías (antidiccionario). Las palabras que son demasiado frecuentes en los documentos de una determinada colección, junto a las palabras desprovistas de significado sustancial, no aportan información. Todas estas *palabras vacías* deben eliminarse durante el proceso de análisis automático de texto, para evitar que puedan ser consideradas como potenciales términos de representación o de indización.

La eliminación de palabras vacías reduce considerablemente el tamaño de la estructura de indización. A pesar de los beneficios que supone la eliminación de palabras vacías, es posible que esta eliminación reduzca la respuesta a una consulta. En efecto, en el caso de que se pregunte por una frase que contenga una palabra vacía, la recuperación será imposible porque dicha palabra vacía no será un término de indización. El proceso de eliminación

de palabras vacías suele realizarse comparando el texto de entrada con una lista de palabras vacías.⁷

Hay que considerar también el tratamiento de los términos flexionados, aquellos relacionados morfológicamente entre sí y con significado común, como “pato”, “pata”, “patos”, “patas”, pero que presentan entre ellos variaciones de género, número o las producidas por la conjugación verbal. En este caso el tratamiento consiste en reducirlos a su término canónico, con objeto de unificar su tratamiento estadístico y de facilitar las operaciones de filtrado, creación de relaciones, e incluso redes semánticas de representación de los conceptos. La lematización se logra mediante los algoritmos de *stemming*: programas que buscan variantes sintácticas de una palabra dada y eligen una única palabra representante del concepto expresado por todas ellas. Su función consiste en mejorar la efectividad de la recuperación y reducir el tamaño de los ficheros de indización.⁸

- b) *Análisis sintáctico*: mediante la utilización de gramáticas se busca determinar la función que cumple cada una de las palabras en una oración, así como entender la combinación que se da entre las palabras para obtener oraciones gramaticalmente correctas. De su intervención se obtiene el reconocimiento de los homógrafos, y la identificación de las expresiones compuestas como base de la discriminación lingüística de las palabras. Su interés estriba en la necesidad del conocimiento especializado de representarse mediante palabras compuestas. Para lograrlo es necesario diseñar algoritmos que permitan incluir palabras compuestas como componentes del dominio, partiendo de la función sintáctica idéntica que tienen las palabras que conforman un término compuesto.
- c) *Análisis semántico*: Se busca que el ordenador pueda hacer inferencias para la extracción del conocimiento y su representación en forma de palabras o términos. El significado intenta alcanzarse usando la lógica extensional de las palabras del discurso, y la lógica intensional de las palabras de la lengua. Por lo general se aplican redes semánticas de la lengua y del discurso especializado al reconocimiento de la significación macroestructural en un discurso dado.

7 Se puede consultar una lista de palabras vacías para el idioma inglés y apreciar la categoría gramatical que suelen tener en Frakes, W., Prieto-Díaz, R. y Fox, C. - DARE: Domain Analysis and Reuse Environment, en *Annals of Software Engineering*, 1998, 5: 125-141.

8 Frakes, W y Baeza-Yates, R. - *Information Retrieval: Data structures and algorithms*. Upper Saddle River: Prentice-Hall: 1992. Existen diferentes algoritmos de *stemming* dependiendo del modo en el que trate los afijos. Los algoritmos de eliminación de afijos extraen los sufijos y prefijos de términos conservando sólo la raíz. Los *stemmers* utilizan la frecuencia de secuencias de letras en el cuerpo de un texto. Por ejemplo, el método *n*-grams une términos basándose en el número de *n*-grams (grupos de letras) que comparten dichos términos.

- d) *Análisis pragmático*: la significación se sitúa dentro de un conocimiento de mundo, que relaciona al lenguaje con sus usuarios y con las circunstancias en que se produce la comunicación. Sin duda es el nivel de análisis más difícil, pues implica factores subjetivos, ocasionales y situacionales en la comprensión del mensaje de un texto.

Las últimas tendencias en indización automatizada, aplicadas a la recuperación en texto completo dentro del impulso que ha supuesto Internet a la información circulante, obliga a profundizar en los modelos lingüísticos.⁹ Sin embargo presentan como limitación mayor no alcanzar a representar las unidades del discurso analizadas globalmente, conjuntando todos los niveles que se implican.

MODELOS ESTADÍSTICOS

La representación probabilística

Dado que la mayoría de los sistemas de recuperación de información trabajan con lenguaje natural, producen cierta imprecisión a la hora de acertar con aquellos documentos que son relevantes para una petición concreta. Para evitarlo Robertson y Sparck Jones se convencieron en 1976 de que era necesario establecer un orden decreciente de importancia en los documentos recuperados tras efectuarse esa demanda, e iniciarse así los métodos de representación probabilística.¹⁰

Su formulación parte de la idea básica de calcular la distribución y el peso de los términos que intervienen en una consulta documental. Cuando un usuario plantea una demanda pretende obtener el conjunto de documentos de respuesta ideal, aquel formado exclusivamente por los documentos que son relevantes a su petición. De tal modo que el proceso de recuperación se centra en especificar las propiedades que debe satisfacer ese conjunto de respuesta ideal, por lo que hay que tantearlas. Este es el objeto de un primer acercamiento efectuado por vía probabilística. El usuario valoraría entonces cuáles de los documentos de la primera selección son relevantes y cuáles no tienen interés para él. La información es pasada al sistema, que en repeticiones sucesivas del proceso refina la búsqueda y se aproxima al conjunto ideal. Se intenta estimar si hecha una consulta q y habiendo un documento candidato d , existe la probabilidad de que el usuario juzgue que el documento d es relevante.

Estas son las técnicas probabilísticas más frecuentes que se aplican a la extracción de términos en la indización automática, así como a la organización y asociación

9 No vamos a profundizar ahora en este aspecto, ya que sería un tema de desarrollo monográfico. Ya que tendríamos que partir de otra introducción teórica.

10 S. E. Robertson y K. Sparck Jones. - Relevance Weighting of Search Terms, en *Journal of the American Society for Information Science*, 1976, 27, 3.

automáticas de conceptos para la construcción de tesauros y de otras representaciones terminológicas:¹¹

Sistemas de ponderación para el filtrado de términos

Estos sistemas buscan la distribución de la frecuencia de las palabras para calcular sus pesos, ya que no todas tienen la misma importancia en un texto. Es muy utilizada la Indización Estadística de Términos por Frecuencias o IDF,¹² sistema de filtrado basado en la ley de Zipf que establece que las palabras de mayor frecuencia no suelen ser significativas, ya que la importancia de un término es proporcional a la frecuencia de ocurrencia en el documento e inversamente proporcional al número de documentos en que aparece.¹³ La técnica IDF dispone un sistema de pesos en función de la frecuencia relativa de cada término en cada documento. Dado que la ley de Zipf sólo estaba pensada para filtrar términos simples, debe complementarse con el Método N-grams que se aplica al tratamiento de palabras compuestas. Este algoritmo trabaja de modo parecido al anterior, pero calcula la frecuencia no sobre cada término o palabra compuesta, sino con cadenas de caracteres de longitud predeterminada y fija.

Métodos de obtención de relaciones entre componentes

Para poder representar adecuadamente la información y utilizarla de un modo óptimo e inteligente es necesario clasificarla, de tal modo que se establezcan relaciones entre los componentes que la definen y describen. Existen numerosos y variados enfoques para realizar este proceso:

- ❖ Clasificadores cuantitativos o de coocurrencias: El más conocido es el Método de Chen, que analiza el uso de grupos de palabras que aparecen simultáneamente en varios documentos. Su acepción más corriente es la de método de coocurrencias (*Co-wording*), y por ser capaz de evaluar la relación entre dos términos, debemos considerarlo un método de clasificación. Su propósito es establecer un peso para la relación que existe entre pares de términos. Ese peso se calcula para cada término basado en el modelo de espacio vectorial y en una función de semejanza asimétrica.¹⁴
- ❖ Algoritmos Estadísticos de Agrupación en Clases: es un proceso de clasificación no supervisada de objetos. Se dispone de un conjunto de vectores $\{x_1, \dots, x_p\}$, que representan a los objetos y a partir de él se desea obtener el conjunto de clases $\{(1, \dots, n)\}$ que los engloban. El problema es que a priori no se sabe cómo

11 Cuanto se refiere a los métodos probabilísticos refleja parte del trabajo de J. Moreira González, I. Díaz, J. Llorens, J. Morato y M. Velasco - Generación automática de tesauros: Propuesta de un método lingüístico-estadístico, en *Ciencias de la Información*, 1999, 30, 4: 51-59.

12 G. Salton. - *Automatic Text Processing: the Transformation, Analysis, and Retrieval of Information by Computer*. New York: Addison-Wesley, 1989.

13 G. K. Zipf. - *Human Behaviour and the Principle of Least Effort: An Introduction to Human Ecology*. New York: Haffner, 1972.

14 H. Chen y K. J. Lynch. - Automatic Construction of Networks of Concepts Characterizing Document Databases, en *IEEE Transactions on Systems, Man and Cybernetics*, 1992, 22, p. 885-902.

se distribuyen los vectores en las clases, ni siquiera cuántas clases habrá.¹⁵ A partir del conjunto de vectores de características dado se trata de conseguir realizar agrupaciones de estos vectores en clases, de acuerdo con las similitudes encontradas. Se presentan a continuación a modo de ejemplo, dos de los clasificadores estadísticos que más se manejan actualmente:

- ☞ Algoritmo K-vecinos. Busca minimizar un índice de rendimiento, basado en la suma de distancias euclídeas cuadráticas de todos los miembros de un *cluster* a su centroide. Exige conocer el número de *clusters* *k* en los que se desea clasificar la muestra de vectores de la población. Si el número de clases no se conoce por adelantado, se puede dejar que el algoritmo determine el número de *clusters* utilizando parámetros definidos por el usuario. El modo de funcionamiento del algoritmo consiste en mover cada vector al *cluster* cuyo centroide esté más cercano al mismo, y actualizar después los centroides de los *clusters*. Su convergencia depende mucho del número de clases.
- ☞ Algoritmo K-vecinos axial o incremental:¹⁶ calcula los *clusters* de forma incremental. Pertenece a la familia de algoritmos de clasificación por centros móviles. Es una variante del algoritmo *k*-vecinos en su versión adaptativa, y del algoritmo de Forgy, en el caso iterativo. Dado un patrón de entrada, el algoritmo debe actualizar la representación de los *clusters* y devolver el índice del *cluster* actual al cual pertenece el patrón, sin necesitar tener presentes los demás patrones. De este modo puede tratarse una sucesión arbitrariamente grande de patrones en tiempo real.

Los modelos de representación vectorial

La red de asociación de un término con aquellos otros términos con los que tiene establecidas co-ocurrencias se puede representar de forma esquemática mediante un mapa lexicográfico. El término se muestra en relación a los términos que le son próximos, los que forman su contexto, y que precisan su significación y uso sin necesidad de requerir de ninguna categoría semántica. Los mapas lexicográficos surgen de la continuación de los métodos estadísticos y sirven en la recuperación, pues muestran la representación de las preguntas junto a las de los documentos que las responden actuando de forma similar a los hiperenlaces en los documentos digitales.

Precisamente porque el uso de pesos binarios es demasiado limitado, se propuso un modo de representación en el que los términos de indización conforman las coordenadas de un espacio de información multidimensional. Los documentos y las consultas se representan por unos vectores que reflejan el valor que se le asigna a cada término según el sistema de ponderación empleado. Cada vector refleja la posición del

15 A. Carpenter y S. Grossberg. - ART-2: Stable Self-organization of Pattern Recognition Code for Analog Input Patterns, en *Applied Optics*, 1987, 26: 4919-4930.

16 Propuesto por M. R. Anderberg. - *Cluster Analysis for Applications*. New York: Academic Press, 1973.

documento (o la consulta) al que representa en el espacio. La homogeneidad entre una consulta hecha y los documentos que responden a ella se calcula comparando sus vectores asociados, de acuerdo con una medida de similitud.¹⁷ De acuerdo con estos requisitos se diseñó el sistema SMART,¹⁸ que realmente exigía consultas muy costosas para discriminar por varios términos, cuando con unos pocos operadores de unión quedaría solucionado en un modelo de representación lógico.

LOS MODELOS COGNITIVOS

Proponen la utilización de procedimientos provenientes de distintas disciplinas cognitivas, empleados para aunar estructuras textuales, estructuras de conocimiento, objetivos y metas de los servicios de información, intenciones y expectativas de analistas y usuarios, contextos de análisis y de recepción de información, y estrategias, procesos y técnicas de representación y recuperación.

La representación del conocimiento mediante técnicas de inteligencia

Inteligencia Artificial: las ontologías

La representación de información mediante la inteligencia artificial (IA) se vale de mecanismos de inferencia para organizar el conocimiento. El conocimiento es básicamente información “rica” en significado (información acerca de información) que se agrupa en conjuntos representados formalmente en torno a una conceptualización.¹⁹ Mediante ella somos capaces de conformar una vista abstracta de un dominio de interés, en la que se simplifican los objetos y otras entidades que lo componen así como las relaciones que se producen entre ellos.²⁰ Todos esos elementos propios de un dominio se categorizan mediante las ontologías. Una ontología es una especificación explícita de una conceptualización que contiene un vocabulario de términos y su correspondiente especificación.²¹ Así pues, para que una computadora pueda interpretar el conocimiento debe contar:

- ❖ Con una sintaxis: lo que requiere de unos símbolos y de un conjunto de reglas para combinarlos.
- ❖ Y con una semántica que explique el significado de las expresiones construidas.

17 G. Salton y M.E. Lesk.- Computer Evaluation of Indexing and Text Processing, en *Journal of the ACM*, 1968, 15, 1: 23-42.

18 G. Salton.- *The SMART Retrieval System – Experiment in Automatic Document Processing*. Englewood Cliffs: Prentice-Hall, 1971.

19 Irene Díaz Rodríguez.- *Esquemas de representación de información basados en relaciones. Aplicación a la generación automática de representaciones de dominios*. Leganés (Madrid): Universidad Carlos III de Madrid, 2001. (Tesis doctoral inédita).

20 Genesereth y J. Nilsson.- *Logical Foundation of Artificial Intelligence*. Los Altos (Ca.): Morgan Kaufmann Publishers, 1987.

21 El término ontología proviene de la filosofía, donde se refiere a existencia. En Inteligencia Artificial se parte del principio de que aquello que existe puede ser representado.

Se muestra así la evidencia de un lenguaje de representación con capacidad para realizar inferencias y conocimiento acerca del dominio al que se aplica.

Una ontología general presenta estos elementos:²²

- ❖ Categorías, que incluyen objetos con propiedades comunes. Se disponen en taxonomías jerárquicas. Las categorías permiten organizar y simplificar el conocimiento de acuerdo con la herencia: clases, subclasses, etcétera.
- ❖ Medidas, por las que se relacionan objetos con cantidades relativas de tipos particulares, así su masa, edad, precios, ... Lo principal de una medida no es su valor numérico particular sino el hecho de que puede ordenarse.
- ❖ Objetos compuestos, que se integran en una categoría jerárquica por ser elementos constitutivos de otro (Ej. : coches están compuestos de llantas, motor, ...).
- ❖ Tiempo, espacio y cambio, que fijan las circunstancias en las que se producen las acciones y los eventos.
- ❖ Hechos y procesos: sucesos individuales que ocurren en un tiempo y lugar concretos. Se llaman procesos los hechos que nos muestran continuidad y homogeneidad.
- ❖ Sustancias, que tienen presencia temporal y espacial (Ej.: aceite). Tienen propiedades intrínsecas (color, temperatura de evaporación, contenido de grasa, ...), y propiedades extrínsecas (peso, forma, ...).

Las ontologías se pueden especificar mediante un inventario clasificado de sustantivos y de relaciones entre ellos, marcadas desde un ordenamiento parcial tipo-subtipo, pudiendo aparecer como:²³

- ❖ Ontología axiomática, que clasifica los subtipos mediante el uso de axiomas y definiciones enunciados en un lenguaje formal, como sucede en las teorías formales en matemáticas y en las colecciones de reglas de un sistema experto.
- ❖ Ontología para un prototipo, que distingue los subtipos comparándolos con un prototipo de cada subtipo. Debe existir un prototipo p y una medida de la *distancia semántica* $d(x, y, c)$ que calcule la disimilitud entre dos entidades x e y cuando ambas se consideran instancias de c . De este modo una entidad x se puede clasificar utilizando el siguiente algoritmo recursivo:
 - ☞ Suponer que x ha sido clasificado como una instancia de alguna categoría c , que tiene s_1, \dots, s_n subcategorías.
 - ☞ Calcular la distancia semántica $d(x, p_i, c)$ para cada subcategoría s_i con prototipo p_i .
 - ☞ Si $d(x, p_i, c)$ tiene un único mínimo para alguna subcategoría s_i , entonces se clasifica x como una instancia de s_i , y se debe llamar al procedimiento de

22 J. F. Sowa.- *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Pacific Grove, CA: Brooks Cole Publishing Co., 1999.

23 K. Reichgelt.- *Knowledge Representation: An AI Perspective*. New York: Ablex Publishing, 1991.

forma recursiva para comprobar si x se puede clasificar por alguna subcategoría de s_i .

- ☞ Si c no tiene subcategorías o si $d(x, p_i, c)$ no tiene un único mínimo entonces el procedimiento de clasificación se detiene porque no es posible realizar una clasificación más fina con los prototipos dados.

Los sistemas expertos

Los sistemas expertos resuelven problemas utilizando una representación simbólica del conocimiento humano. Esta representación ha de ser explícita, y generar una capacidad de razonamiento independiente de la aplicación específica. Los sistemas basados en conocimiento utilizan sistemas expertos para resolver problemas. Basan su rendimiento en la cantidad y calidad del conocimiento de un dominio específico. Se les exige además poder explicar sus conclusiones y el proceso de razonamiento, y tener un alto rendimiento en un dominio específico. Deben usar la heurística frente a los modelos matemáticos y funcionar desde inferencias simbólicas, y no desde algoritmos numéricos. Los componentes básicos de estos sistemas son:

- ❖ Base de conocimiento: representación del conocimiento de un dominio.
- ❖ Máquina de inferencia: proceso de razonamiento a partir de los datos aprovechando el conocimiento representado en la base.
- ❖ Memoria de trabajo: lugar de almacenamiento de los datos de entrada y de las conclusiones intermedias a las que se llega durante el proceso de razonamiento.
- ❖ Interfaz de usuario: entrada y salida del usuario al sistema, incorporando mecanismos interrogativos y explicativos.
- ❖ Interfaz de adquisición: mecanismos para automatizar el aprendizaje y para facilitar la adquisición y el depuramiento interactivo.

Métodos basados en las redes neuronales

Se utilizan para resolver problemas relacionados con el conocimiento humano como el reconocimiento de patrones, el reconocimiento del lenguaje hablado, el reconocimiento de imágenes, o los procesos de control adaptativo. El aprendizaje de una red neuronal está relacionado con los pesos de las conexiones entre sus nodos.²⁴ Cuando se le presenta un patrón a la red, ésta produce una respuesta. La señal que se recibe en la capa de neuronas de entrada, cuando se le presenta el patrón, se mueve a través de los enlaces o conexiones entre capas hacia las neuronas de la capa de salida. Estos enlaces modulan la señal a su paso con los pesos que los caracterizan. Por lo tanto, si se quiere modificar la señal que llega al final a la capa de salida, habrá que actuar sobre dichos pesos.

Un clasificador es un sistema que va a permitir determinar cuál de las M clases es la más representativa para un patrón de entrada no estático que contiene N elementos.

24 D. Hebb.- *Organization of Behaviour*. New York: Wiley & Sons, 1949.

El clasificador neuronal actúa en dos etapas, en la primera se contabiliza el número de elementos que pertenecen a cada clase y en la segunda se selecciona el máximo. La primera etapa se alimenta con los N elementos del patrón de entrada en paralelo, cuando se produce la comparación del patrón de entrada con los prototipos de las distintas clases y pasan los resultados intermedios a la siguiente etapa en paralelo. En la segunda pasada se selecciona el máximo. Todas las clases tendrán salida, pero al acabar la clasificación sólo se apreciará la salida para la clase con mayor probabilidad. El resto, con valores muy bajos, se consideran inapreciables. Se pueden utilizar las salidas como realimentación de la primera etapa adaptando los pesos iniciales según un determinado algoritmo de aprendizaje (principio de realimentación negativa). Redes neuronales muy utilizadas en análisis de contenido son las de Kohonen, ART1 y ART2.

El modelo de Kohonen se diseñó en adaptación a las características biológicas, el mapa de características, dispuesto mediante una red neuronal de dos capas, una de entrada y otra de competición. Los mapas autorganizativos establecen que ciertas redes neuronales pueden adaptar sus respuestas, de tal forma que la posición de la célula que produce la respuesta pasa a ser específica de una determinada característica de la señal de entrada.²⁵

Cada neurona está conectada con otras de su entorno de manera tal que produce una excitación en las más próximas y una inhibición en las más alejadas, y causa una interacción lateral. Tanto la excitación como la inhibición lateral son gradualmente más débiles a medida que nos alejamos de la neurona en cuestión. Este mecanismo hace que cuando un estímulo produce una reacción en una célula, las células de su entorno inmediato se vean influidas por dicha reacción, de modo positivo las más cercanas y de forma negativa las más alejadas.

Por su parte la teoría de la resonancia adaptativa o ART se originó en la necesidad de superar el dilema de la Estabilidad-Plasticidad de Grossberg. Su objetivo era diseñar un sistema que reaccionase de forma plástica a nuevos estímulos, pero de forma estable a aquellos que sean irrelevantes. Se agregaba la posibilidad de conmutar estos dos estados plástico y estable, cuando fuese necesario, para evitar la degradación de lo ya aprendido. El sistema debe tener incorporada alguna estimación para decidir la relevancia o no de un estímulo si lo quiere realizar sin supervisión. Grossberg ha diseñado dos modelos en los que se aplica la teoría de la resonancia adaptativa:²⁶ el ART-1 que sólo permite clasificar patrones binarios, y su variación, el ART-2, para trabajar con patrones analógicos.

25 Kohonen, T.- Self-Organized Formation of Topologically Correct Feature Maps, en *Biological Cybernetics*, 1982, 43: 82-95.

26 Grossberg, S.- How Does the Brain Build a Cognitive Code?, en *Psychological Review*, 1980, 87:19-32.

LA REPRESENTACIÓN DE INFORMACIÓN MEDIANTE TESAUROS.

Debido a la presencia de las polisemias, de los homógrafos y de las homonimias muchos de los sistemas de tratamiento de información clasifican la información textual en lenguaje natural mediante la utilización de lenguajes controlados. Estos lenguajes restringen la amplitud de los naturales mediante la utilización de un conjunto limitado de expresiones y de las relaciones semánticas entre ellas.²⁷ De este modo, la clasificación de información se ve favorecida por la utilización de una gramática más reducida. Estos lenguajes controlados pueden establecerse como:

- ❖ Lenguajes de clasificación o conjunto de codificaciones con las que se representa cada uno de los conceptos de una materia.
- ❖ Lenguajes combinatorios formados por un conjunto de descriptores o términos que determinan el conocimiento sobre un dominio específico.

Dentro de este último nivel se aprecian a su vez dos nuevos tipos:

- ☞ Listas de autoridades, confeccionadas a priori y de extensión finita que conforman una colección no ordenada de conceptos representados por palabras o por expresiones. Sólo los términos que figuran en esta lista pueden ser utilizados para clasificar los documentos. Las listas no están obligadas a contener relaciones semánticas entre los términos.
- ☞ Tesauros o listas de control terminológico donde se representan los conceptos de un área del conocimiento mediante términos. Es un lenguaje convenido y controlado para que los términos presenten una clara univocidad, en el que se ofrece la posibilidad de combinar sus términos tanto para analizar como para buscar el contenido concreto de los documentos. El esquema de representación de información basado en la estructura de tesauro establece relaciones entre los términos de indización para conseguir una recuperación más eficaz.²⁸

Los tesauros establecen relaciones semánticas entre sus términos, con el fin de inducir hacia la representación exacta de los conceptos y de facilitar el diálogo entre los contenidos documentales, los usuarios y los documentalistas. Dos términos de un lenguaje documental siempre que se relacionen lo hacen en reciprocidad. Esta relaciones pueden ser de equivalencia; definitorias para evitar la polisemia; jerárquicas o clasificatorias; y de asociación. Mientras que la clasificación categórica tiene como misión organizar los descriptores mediante una abstracción conceptual representativa de un sector del dominio que cubre el tesauro. El término que encabeza cada sector se ha denominado tradicionalmente macrodescriptor o *top term*, y su función en un lenguaje controlado es sólo la de organizar, sin valer, por su generalidad, para

27 G. Van Slype.- *Les Langages d'Indexation: Conception, Construction et Utilisation dans les Systèmes Documentaires*. Paris : Les Editions d'organisation, 1991.

28 G. Salton.- Experiments in automatic thesaurus construction for information retrieval, en *Proceedings IFIP Congress 1971*, TA-2, 43-49, 1971.

transferir hechos de la semántica de mundo y por ello sin poder ser utilizados para la indización.

Por su parte la asociación entre los términos pretende sugerir enlaces entre facetas de conceptos del tesoro. Los términos asociados establecen una relación difícil de fijar, pues depende mucho de la creatividad de los autores del tesoro, lo que se aprecia como un contrasentido, pues esta parte es la que establece las relaciones más interpretativas y evocadoras de un lenguaje controlado. La navegación entre los términos permite establecer enlaces horizontales y oblicuos por los que pasamos a relacionar unas categorías con otras, e incluso destacamos similitudes entre términos pertenecientes a la misma categoría pero que muestran una relación especial. Mediante esta navegación saltamos de una categoría a otra y, a la vez, en cada una de ellas estamos capacitados para ascender o descender por su disposición jerárquica.

Se muestra una propensión a construir actualmente Tesoros Conceptuales: aquellos que son compilados de forma postcoordinada a partir del sublenguaje de un área específica del conocimiento, en cuya utilización prima la recuperación.²⁹ La denominación de tesoro conceptual proviene de su fundamentación teórica alrededor de la noción de materia, es decir, del concepto del que tratan los textos. Así, Maniez propuso un modelo de tesoro de relaciones asociativas, que aunase términos y conceptos reales por su similitud de sentido en el contexto específico del usuario.³⁰ Tales tesoros presentan estas características novedosas:

- a) Aportan notas definitorias que aclaran las posibles dudas de uso, e incluso dan algunas explicaciones extensas.
- b) Razonan las equivalencias existentes entre términos
- c) Consideran incluso los términos coloquiales, también las variaciones y truncamientos de los términos reconocidos.
- d) Glosan todas las palabras “no vacías” existentes en las bases de datos.
- e) Contienen numerosas relaciones asociativas entre descriptores, e incluso con los no descriptores.

La integración que suponen estos rasgos demuestra la progresiva desaparición funcional de las diferencias entre los lenguajes controlados de los tesoros tradicionales y el lenguaje natural mediante el que se manifiestan las necesidades de los usuarios. La causa que lo permite es la presencia de ambos lenguajes en las interfaces de los Sistemas de Gestión Documental como herramientas de ayuda para la recuperación. Al funcionar integrados ambos lenguajes en los Sistemas de Gestión de la Información, evitan la indeterminación de las búsquedas que plantean en lenguaje natural usuarios no expertos.³¹

29 W. Schmitz-Esser.- *New approaches in Thesaurus Application*, en *International Classification*, 1991, 18, 3: 143-147.

30 J. Maniez.- *Los lenguajes documentales y de clasificación*. Madrid: Fundación Germán Sánchez Ruipérez, 1993: 213-215.

31 M. López Alonso.- *Los tesoros conceptuales como herramienta de precisión en los sistemas de organización científica*, en *Revista Interamericana de Bibliotecología*, 1999, 22, 1: 21-35.

Los términos se extraen de documentos de texto completo, y luego conforman las bases de conocimiento descentralizadas de Internet. Los enlaces en la red se establecen tras su adaptación al espacio conceptual hipertextual mediante el lenguaje HTML o el XML. Como resultado se obtiene un cuerpo terminológico cuya representación se establece como una red semántica neuronal: en cada nodo hay un concepto semántico con el que se asocian una serie de descriptores. Los enlaces entre los descriptores pueden también establecerse según las típicas relaciones de los tesauros: de equivalencia, facetadas o por asociación. El funcionamiento de la recuperación se establece desde la pregunta del usuario. Los conceptos solicitados se confrontan con la red terminológica, cuyos elementos están diseñados como mapas representativos de los textos, y de este modo actúan como lenguajes controlados que organizan la información de cualquier objeto contenido en la red.

Esta manera de construir los tesauros se ha consolidado precisamente porque ayuda al usuario a recuperar la información con soluciones provenientes del Modelo Cognitivo. La precisión se enriquece cuando el usuario puede comparar su conocimiento con la representación que refleja la terminología que se le ofrece. En su procesamiento se percibe que es en gran parte coincidente con el que se efectúa en los programas más avanzados de indización automática. Se comienza procesando el texto en lenguaje natural mediante lematizadores y analizadores morfosintácticos y semánticos, para concretar el análisis de las palabras y de sus relaciones. Primero identificando las relaciones simples, luego los sintagmas. Con los términos identificados se establecen bancos de datos terminológicos. Finalmente, mediante redes neuronales, se crean bases de conocimientos terminológicas mediante las que se establecen búsquedas automatizadas en Internet.

Estos tesauros han de aplicarse a la información de carácter enciclopédico distribuida por Internet. Para gestionar tal riqueza de conceptos es obligado desarrollar diferentes tesauros conceptuales, que se unan en una red semántica de estructuras neuronales. En su funcionamiento se combinan la tecnología hipertextual y las redes semánticas, las cuales proporcionan un mapa tanto de las variables con que se efectúa la demanda como de las obtenidas en la recuperación. El usuario tiene acceso a los sistemas de información a través de una representación gráfica de la red (tabla de contenidos gráficos con las conexiones visibles mediante enlaces hipertextuales), por la que puede rastrear los documentos siguiendo el recorrido de su interés. El desarrollo de asociaciones de redes neuronales que vuelven dinámicos los nodos hipertextuales permite que los conceptos situados en cada uno de ellos puedan ser consultados simultánea y no sólo secuencialmente, de acuerdo con las necesidades de cada usuario.

La autogeneración de un tesoro conceptual exige el desarrollo de potentes algoritmos capaces de procesar los diferentes documentos de la base de datos y la obtención de sus índices de covarianza correspondientes. Se producirán así :³²

32 H. Chen, T. Yim y D. Fye.- Automatic thesaurus generation for an electronic community system, en *Journal of the ASIS*, (1995), 46, 3: 178-180.

- ❖ Listas de términos a partir de las cuales se identifiquen los conceptos más importantes.
- ❖ Indizadores automáticos que identifiquen las materias.
- ❖ Analizadores semánticos que agrupen los conceptos relacionados entre sí.

En esta línea se sitúa la propuesta de Paice, quien propone que tanto las preguntas de los usuarios como los textos que responden a ellas se representen mediante fragmentos de documentos extraídos, según el contexto, de los conceptos que representa.³³ Para construir las entidades desde las cuales se representen los conceptos contenidos en las preguntas y en los textos es necesario contar con una amplia base de conocimientos terminológicos de un campo concreto de referencia, en la que se incluirían:

1. Los descriptores del tesauro que constasen en los subtextos analizados.
2. Los conceptos implícitos pero que tengan relación en el tesauro con los primeros.
3. Los términos que pertenezcan, junto a los anteriores, a la misma red semántica.
4. Las relaciones de todo tipo que se den entre todos los términos anteriores.

Las recuperaciones se establecen por comparación con la red neuronal de cada nueva entidad que represente una pregunta. Así se localizarán por coincidencia los documentos situados en mayor proximidad a los conceptos pedidos. Podemos considerar como los métodos más utilizados de clasificación en el ámbito que nos afecta a los siguientes (de ellos hemos dado noción a lo largo de este artículo):

1. Métodos de agrupación basados en Grafos teóricos
2. Métodos de agrupación basados en Centros de masas
 - 2.1. Algoritmo K-vecinos o K-medias
 - 2.2. Algoritmo ISODATA
3. Algoritmos de clasificación basados en Redes neuronales
 - 3.1. Mapas autoorganizativos de Kohonen
 - 3.2. La teoría de resonancia adaptativa (ART-1 y ART-2)

La representación de información mediante mapas conceptuales

Un modelo, derivado tanto de los modelos estadísticos (representaciones vectoriales) como de los lingüísticos, busca hacer representaciones diagramáticas del conocimiento mediante diferentes lenguajes visuales (representaciones gráficas de los conceptos y de las relaciones entre ellos). Los *conceptos* son los *nodos* (cada uno con su tipo, nombre y contenido), y las *relaciones* son los *enlaces*. Son eficaces para organizar nueva información e integrarla en el conocimiento existente, ya que su construcción ayuda a reconocer nuevas relaciones entre conceptos y a refinar la comprensión de las relaciones existentes. El proceso de construcción de un mapa conceptual se compone de cinco etapas principales:

- 1) selección de los conceptos que formarán parte del mapa;
- 2) listado de los conceptos desde los más abstractos a los más concretos;

33 C. Paice.- A thesaural mode of Information Retrieval, en *Information Processing and Management*, 1991, 27, 5: 434-435.

- 3) agrupación de los conceptos relacionados;
- 4) ordenación de los conceptos;
- 5) enlace de cada par de conceptos mediante líneas etiquetadas.

Precisamente para evitar la limitación que imponen los tesauros al uso respecto de la aplicación de relaciones semánticas entre sus términos, y luego de corroborar las tendencias que venimos explicando, se ha elaborado la Norma ISO/IEC 13250 *Topic Maps* que define el modelo y la sintaxis de intercambio para formalizar mapas conceptuales.³⁴ La motivación para estudiar los mapas de temas tiene su origen en la necesidad de crear índices y de proporcionar así acceso a la información con base en un modelo de conocimiento cuyo corazón es el concepto de *tópico*. El estándar internacional proporciona una notación estandarizada para representar información referente a las fuentes de información utilizadas para definir tópicos y sus relaciones. Así, un conjunto de uno o más documentos que utiliza la notación definida por este estándar se llama “*topic map*” (mapa de temas).

Las ideas fundamentales del estándar son los conceptos, los diferentes casos que ofrecen, y las relaciones (asociaciones) que se dan entre ellos. Un mapa conceptual en su formato de intercambio es un documento o grupo de documentos SGML (o XML) que usan los diferentes tipos de elementos para representar los conceptos, los casos que presentan, y las asociaciones entre ellos. Precisamente por su estructuración semántica de los enlaces en la red se denomina a esta norma “el GPS del universo de la información”, en cuanto que es una solución para organizar y navegar por los extensos y continuamente crecientes recursos de información (publicaciones de editoriales y corporativas, así como sitios web), y establecer un puente entre los campos del conocimiento y la gestión de la información.

Un tópico, en sentido general, representa cualquier cosa –una persona, una entidad, un concepto. El representante abstracto de un tópico se llama materia. En cierto sentido, una materia se corresponde con lo que Platón llamaba “*idea*”. Formalmente el término tópico hace referencia a un elemento perteneciente a un “*topic map*” que representa una materia. Sin embargo habitualmente ambos conceptos se fusionan y nos referimos a ambos mediante el uso del término tópico.

Los tópicos (y sus respectivos tipos) varían según varíe la información que se esté tratando. En efecto, si se está hablando de documentación relativa a *software*, los tópicos representarán funciones, variables, objetos y métodos; si la información es relativa a botánica, los tópicos se referirán a plantas, algas y sus funciones.

Esta norma internacional proporciona una notación normalizada para representar información intercambiable en torno a la estructura de los recursos de información utilizados para determinar los conceptos, y las relaciones que se dan entre los conceptos. Un grupo de uno o más documentos interrelacionados que emplea la

34 ISO/IEC FCD 13250: 1999 – *Topic Maps*. Elaborada por ISO/IEC JTC1/SC34 – Descripción documental y Lenguajes de Procesamiento.

notación definida por esta norma internacional es denominado “mapa conceptual”. En él se define un espacio en el que las localizaciones son conceptos, y en el que las distancias entre conceptos son apreciables a partir del número de conceptos intermedios que deben ser visitados si se quiere ir de un concepto a otro y también el tipo de relación que define la trayectoria de un concepto a otro, si la hay, a través de los conceptos intermediarios, si estos existen.³⁵ La arquitectura de los mapas conceptuales está ideada para facilitar la unión de mapas conceptuales sin necesitar los mapas tópicos combinados para ser copiados o modificados. A causa de su carácter extrínseco, los mapas conceptuales pueden ser considerados como “incrustaciones”, o extensiones hacia los objetos informativos.

SGML se establece como la notación base de los mapas conceptuales. Un mapa conceptual intercambiable contiene al menos un documento SGML, pero puede incluir y/o referir a otros tipos de recursos informativos. Un grupo de recursos de información que comprende un mapa completo intercambiable de conceptos puede especificarse utilizando la ayuda del grupo de objetos enlazados *Bounded Object Set* (BOS).³⁶ Dado que XML es un subgrupo de SGML, puede también ser usado como notación base para los mapas conceptuales.

La notación de mapa conceptual se define como una “Arquitectura SGML”, y esta norma internacional toma la forma de un “*architecture definition document*” expresado de conformidad con el Anexo Normativo A.3 de ISO/IEC 10744:1997, el “*SGML Architectural Form Definition Requirements*” (AFDR). La definición formal de la notación de un mapa conceptual se expresa como un meta-DTD.

Un tópico posee dos características fundamentales: debe poseer un nombre y al menos una ocurrencia. Una ocurrencia es un puntero para la fuente de información en la que el tópico es relevante. Las ocurrencias pueden ser de varios tipos, cada uno de los cuales se denomina *rol de la ocurrencia*. Los conceptos de *tópico*, *tipo de tópico*, *nombre del tópico*, *ocurrencia* y *rol de la ocurrencia*, permiten organizar las fuentes de información de acuerdo con los *tópicos* así como crear índices, pero realmente no permiten mucho más.

Lo realmente importante de este modelo es la posibilidad de crear asociaciones entre tópicos. Cada tópico participante en una asociación posee un *rol de asociación* que establece el rol que juega el *tópico* en la asociación.

La estructura capítulo-sección de un texto de documentación científica se fracciona en cientos (o quizás miles) de módulos de texto separados que son autocontenidos respecto a un tema concreto. Del mismo modo, la asignación de códigos jerárquicos permite un acceso rápido a la información contenida en los módulos. Obviamente

35 Dos conceptos pueden conectarse mediante una asociación, y pueden también conectarse en virtud de participar en un acontecimiento común. Además, los objetos informativos pueden tener propiedades, así como valores de esas propiedades, asignados a ellos externamente. Estas propiedades son denominadas “*facet types*”: La palabra “*faceta*” se utiliza aquí de forma metafórica: una *faceta* es una propiedad de un grupo de documentos que puede ser usada para crear una perspectiva suya.

36 Definida por la arquitectura HyTime en ISO/IEC 10744:1997.

estas dos características hacen pensar en los mapas conceptuales como candidato ideal para ponerlo en práctica.

TENDENCIAS DE INVESTIGACIÓN EN CLASIFICACIÓN

Es perfectamente constatable la proyección que la sociedad de la información está realizando sobre la representación y organización del conocimiento y, por ende, en la clasificación de la información.

Las nuevas tecnologías, en efecto, han permitido el advenimiento de los **hipermedia**, un nuevo tipo documental que unifica en sí los diferentes modos de expresión informativa,³⁷ hasta desembocar en el concepto de **hiperdocumento** como producto simbiótico de los contenidos multimedia, los soportes informáticos y las redes de comunicación. Ha sido C. Durieux quien ha caracterizado estos hiperdocumentos dentro de la red, nuevo espacio que tiene como rasgos distintivos una constante metamorfosis (construcción permanente), la heterogeneidad, el isomorfismo, la exterioridad, la topología y la movilidad de los centros.³⁸ Los contenidos que se comunican por la red, lo hacen siguiendo unos parámetros no jerárquicos, heterogéneos, públicos y en constante cambio.³⁹ En consecuencia, existe una ingente creación e inserción de información en la red, con una diseminación escasamente estructurada.⁴⁰ Asistimos a la enunciación de una disyuntiva que puede hacer derivar la sociedad de la información, bien sólo a una sociedad de la comunicación, o a una sociedad del conocimiento.

La disyuntiva no es baladí dados sus extraordinarios efectos sobre el conocimiento humano. En la década de los 80 del pasado siglo, J. McHale observaba que en la nueva sociedad postmoderna la información presentaba como rasgos reseñables su simultaneidad, transdisciplinariedad, dinamismo e interacción multimedia,⁴¹ y sin embargo esta ubérrima abundancia de información en movimiento no derivaba necesariamente en una comprensión, asimilación y conocimiento. Entre otros muchos efectos, actualmente en investigación, las características anotadas alteran sus-

37 Cit. J. Feather. *The Information Society*. 2nd edition. London: Library Association Publishing Ltd., 1998.

38 La Recherche Terminologique en Traduction: Pour une approche Hypertextuelle. *Meta*, 1997, XIII.4. Cit. M. Pinto Tratamiento de los Contenidos en la Sociedad de la Información. En: *La Sociedad de la Información*. Madrid: Editorial Centro de Estudios Ramón Areces SA, 1999, 267-288.

39 G. Charton. Reperage de l'information sur Internet: Nouveaux utils, approches bibliothéconomiques et microstructures. En: *CDI virtuel: utilisation des réseaux pour l'accès à la communication. Stage national DISTNB/Sousdirection des Technologies Nouvelles CRDP de Reims, 29 Avril 1997*. Paris: Centre de Calcul, Recherche et Réseaux Jussieu, URFIST, 20 de junio 1999 [Consulta 22 de marzo de 2001].

40 P. Fayard. ¿Exceso de información o ceguera estratégica? Consideraciones sobre la Inteligencia y el Conocimiento. *Telos*, (1996), 44 : 11-14.

41 *El entorno cambiante de la información*. Madrid: Tecnos, 1981.

tancialmente la educación, que se ha visto obligada a adaptarse a una nueva *Era del Aprendizaje*⁴² en la que las formas de aprender deben fundamentarse en:

- ❖ Una estrategia secuencial y asociativa, en relación directa con el pensamiento relacional, convertido en el instrumento idóneo para un aprendizaje adecuado.
- ❖ El desarrollo de una habilidad de asimilación sobre documentos dinámicos, simultáneos e interactivos, en una clara referencia a los lenguajes URML.⁴³
- ❖ La formación continua conducente a una constante actualización a través de las comunidades educativas virtuales⁴⁴.
- ❖ El autoaprendizaje, que implica una mayor insistencia sobre los métodos de manipulación de las fuentes y recursos de información que sobre los contenidos.⁴⁵
- ❖ La *Escuela* comienza a ver relativizada su función como única institución de formación, por cuanto ha de compartir el espacio educativo con el *Aula sin Muros*,⁴⁶ mucho más efectiva en Internet.⁴⁷

La sociedad informacional le impone todavía otro condicionante al conocimiento humano: los sofisticados medios tecnológicos facultan, a tenor de lo arriba expuesto, para la masiva producción de mensajes multimedia, pero esta misma tecnología le permite al individuo diseñar su propio sistema de acceso a la información.⁴⁸ La información, pues, se produce masivamente pero su consumo está muy fragmentado e individualizado; y su diseño se orienta hacia una franja concreta de consumidores y una absoluta atemporalidad.⁴⁹ Se produce así la paradoja de que si bien las nuevas tecnologías fomentan la integración y organización de las actividades humanas, al mismo tiempo rompen el concepto del individuo como medida del entorno comunicativo.⁵⁰ Este condicionante tiene una notoria proyección

42 L. Joyanes Aguilar. *Cibersociedad: los retos sociales ante un nuevo mundo digital*. Madrid: McGraw-Hill, 1997.

43 Realiza una buena perspectiva hacia sus usos didácticos J.R. Hilera & S. Otón & J. Martínez. *Aplicación de la Realidad Virtual en la enseñanza a través de Internet*. Disponible en Internet <http://www.ucm.es/info/multidoc/multidoc/revista/num8/hilera-oton.html> . [Consulta 22 de marzo de 2001].

44 Analiza E. Lorente Bilbao. *Las Comunidades virtuales de enseñanza - aprendizaje*. Disponible en Internet <http://www.ucm.es/info/multidoc/multidoc/revista/num8/eneko.html>. [Consulta 22 de marzo de 2001].

45 J. C. Marcos Recio. *Fuentes de información, educación y proyectos de investigación multimedia*. Disponible en Internet <http://www.ucm.es/info/multidoc/multidoc/revista/num8/jmarcos.html>. [Consulta 22 de marzo 2001].

46 Definida por E. Carpenter & M. McLuhan. *El Aula sin muros*. Barcelona: Laia, 1974.

47 Según apuntan M. Gisbert Cervera et al. Entornos Virtuales de Enseñanza – Aprendizaje: Proyecto GET. *Cuadernos de Documentación Multimedia* (1998) 6-7. Disponible en Internet <http://www.ucm.es/info/multidoc/multidoc/revista/cuad6-7/evea.htm>. [Consulta 25 de marzo de 2001].

48 Cifr. F.L. Miksa. *The DDC, the Universe of Knowledge and the Post- Modern Library*. New York: Forest Press, 1998.

49 Dibuja con suma habilidad este nuevo entorno informativo M. Castells. *La Era de la Información: Economía, Sociedad y Cultura*. Madrid: Alianza Editorial, 2000, vol. 1, p. 359-408.

50 Cifr. R. Barglow. *The Crisis of the Self in the Age of Information: Computers, Dolphins and Dreams*. London: Routledge, 1994.

epistemológica en la clasificación del conocimiento: los sistemas de clasificación construyen una realidad útil en un momento y para una finalidad determinados, pero no son una referencia absoluta, sino una *conveniencia* en la organización del conocimiento.⁵¹

En estas circunstancias, los sistemas de clasificación han debido iniciar una estrecha relación con la organización del conocimiento, entendida como una ciencia tri-dimensional, que mediante el establecimiento de unos principios, métodos e instrumentos, sea capaz de gestionar el conocimiento humano desde la representación, organización y comunicación del mensaje documental.⁵² Desde esta perspectiva, los sistemas de clasificación han comenzado a incorporar la necesidad de dar respuesta a dos interrogantes:

1. La búsqueda de información es, sin duda, una manifestación de la conducta humana, que ha de ser considerada en documentación desde su óptica psicológica. Es preciso el concurso de la antropología y la psicología sociológica para determinar cómo la *cultura* condiciona el conocimiento humano y las necesidades informativas.⁵³ Es este conocimiento el que ha animado al desarrollo de los *estudios de usuarios*, cuyos objetivos son el análisis de las necesidades y comportamientos en la información, junto con el análisis de las motivaciones y actitudes de los usuarios.⁵⁴ Esta vía de investigación sienta sus cimientos en el *paradigma cognitivo*, que recientemente presenta como carácter más novedoso la consideración nodal del comportamiento del usuario en los sistemas de recuperación de información,⁵⁵ tendencia que para algunos autores debe ser reconocida como un nuevo paradigma.⁵⁶
2. El planteamiento de qué es la *materia*, asunto o tema del mensaje informativo de un documento. Si la respuesta podía ser abordada mediante una metodología propia del análisis documental sobre el *documento*, bien asentada por lo demás, no sucede lo mismo por lo que respecta al *hiperdocumento*. La materia del hiperdocumento está en relación directa con los conocimientos previos del usuario, sus prejuicios, y sus expectativas de respuesta, determinantes indudables de la interrogación. El tema o materia del hiperdocumento debe referirse a su *potencialidad*

51 Defiende F.L. Miksa, obr. cit. p. 86.

52 Interesante perspectiva de M.A. Esteban Navarro & F.J. García Marco. Primeras Jornadas sobre Organización del Conocimiento: Organización del Conocimiento e Información Científica. *Scire*, (enero – junio)1, 1 : 149-157.

53 Tesis que defienden H. Albrechtsen & B. Hjørland. Information Seeking and Knowledge Organization. *Knowledge Organization*, 1997, 24, 3 : 135-6.

54 Sumamente aclaratoria la comunicación de M. Izquierdo Alonso & J. Ruiz Abellán & J.T. Piñera Lucas. Los estudios de usuarios en los programas de gestión de calidad. Propuesta de un marco teórico integrador para el estudio del usuario de información. En: *Los sistemas de información al servicio de la sociedad: actas de las jornadas. VI Jornadas Españolas de Documentación*. Valencia: FESABID, 1998: 433-444.

55 Cifr. C. Kuhltham. Inside the search process: information seeking from the user's perspective. *Journal of the American Society for Information Science*, 1995, 42, 5 : 361-371.

56 Como, por ejemplo, W. Sugar. User centered perspective of information retrieval research and analysis methods. *Annual Review of Information Science and Technology*, 1995, 30 : 77-109.

epistemológica,⁵⁷ conforme a la transversalidad de los mensajes en la sociedad informacional.

De otro lado, los condicionantes de la sociedad informacional, antes expuestos, han tenido un efecto extraordinario en las formas de crear, producir, diseminar y usar el conocimiento, lo que ha perturbado profundamente la estructura del saber y los campos académicos normalizados en la tradicional disposición del conocimiento. Ciencias, técnicas y disciplinas han comenzado a fundirse en interrelaciones hasta ahora desconocidas, por efecto de la aparición de nuevas comunidades científicas y discursivas que producen documentos y no se adaptan con plenitud a las estructuras académicas existentes,⁵⁸ en tanto que son cada vez más frecuentes las búsquedas de información al margen de la estructuración del conocimiento en disciplinas.⁵⁹ La consecuencia de todo ello es que ni la estructura epistemológica académica del conocimiento, ni tampoco los sistemas de recuperación de información, engloban y pueden aplicarse a todo el conocimiento.

Cuando desde la óptica de la documentación y sus técnicas para organizar el conocimiento se ha abordado la cuestión, se ha diagnosticado como fenómeno del máximo interés la transversalidad del conocimiento en varias ciencias o disciplinas, entendida como la aplicación del método científico de un área de conocimiento a otras áreas. El carácter anfibológico del concepto de información ha obligado, empero, a depurar a la vez el concepto de *transversalidad* en referencia a otros conceptos semánticamente próximos, tarea en la que ha sido muy útil el trabajo definitorio de I. Dahlberg,⁶⁰ quien ha distinguido:

- ❖ **Interdisciplinariedad**, proceso en el que una disciplina es estudiada desde la perspectiva de otras afines.
- ❖ **Multidisciplinariedad**, una especialidad de la interdisciplinariedad. Consiste en aquel proceso que analiza sólo un hecho o fenómeno propio de un área de conocimiento, desde la óptica de otra área de conocimiento. Un caso específico es la **syndisciplinariedad**, que implica la colaboración de dos o más disciplinas para el análisis de un caso concreto.
- ❖ **Pluridisciplinariedad**, proceso que invierte los términos, al considerar ahora una *propiedad* en su comportamiento, cuando es referida a distintas disciplinas.

La automatización y el desarrollo de Internet han hecho que para los lenguajes de clasificación e indización los conceptos de multidisciplinariedad e interdisciplinariedad complementen a la transversalidad por aproximación en sus efectos, en cuanto

57 Cit. B. Hjörland. *Information Seeking and Subject Representation: An activity – theoretical approach to information science*. Westport; Connecticut; London: Greenwood, 1997.

58 Cifr. R.S. Clegg & G. Palmer (eds.). *The politics of management knowledge*. London [etc.]: Sage, 1996.

59 Según hace constar D.A. Klein. *Decision – analytic intelligent systems automated explanation and knowledge acquisition*. Hillsdale; New Jersey: Lawrence Erlbaum Associates, 1994.

60 Current trends in knowledge organization. En: *Organización del conocimiento en sistemas de información y documentación: actas del I Encuentro de ISKO – España. Madrid 4 y 5 de noviembre de 1993/ ed. F.J. García Marco*. Zaragoza: Librería General, 1995: 7-26.

a su eficacia en la representación del conocimiento y su recuperación: los tres procesos indicados fomentan la inestabilidad, así como también imposibilitan una predicción aceptable. La solidez de las predicciones en la programación de los sistemas de recuperación de información se basaba en la firmeza de las disciplinas, una firmeza que ni la indización automatizada ni los hiperenlaces les ha devuelto.⁶¹ Ha sido D. W. Langridge quien mejor ha explicado el desfase entre la nueva creación del conocimiento y su organización por un lado, y la clasificación por otro.⁶² En su opinión en el nuevo contexto intelectual las *clases* se basan en el *interés* de la investigación – conocimiento, y no en su acoplamiento a las *formas disciplinarias del conocimiento*. En consecuencia, distingue entre las formas de conocer o *disciplinas* y los objetos de conocimiento o *fenómenos*.

Por definición una disciplina discrimina del saber la parte que enuncia y agrupa, esto es, se comporta como un sistema discreto de organización del conocimiento. La clasificación, luego, jerarquiza las disciplinas, por lo que consecuentemente la información de un fenómeno estará subordinada en su ordenación a la disciplina o subdisciplina desde la que se haya abordado,⁶³ proporcionará una *clasificación de aspectos*, que dispersa la materia o información de un fenómeno en varias disciplinas. Este tipo de clasificación posee, empero, formas de reacción frente a la transversalidad: realizar una división de áreas de conocimiento dentro de los fenómenos, las facetas, y luego subordinar los fenómenos a las disciplinas desde las que pueden ser analizados.

Algunos lenguajes de clasificación han tendido, pues, hacia una teoría de la clasificación facetada – analítica, fundamentada en el método de clasificación del **Racionalismo**.⁶⁴ La tendencia, sin embargo, no ha obtenido los resultados apetecidos de eficacia y eficiencia en la recuperación de información, a tenor de la forma de creación de conocimiento propia de la sociedad de información, por cuanto la moderna epistemología se desplaza desde unos planteamientos positivistas a otros más interpretativos y neopragmáticos. Es, por tanto, mucho más útil en la clasificación el empleo de una combinación de métodos de clasificación, a saber:

- ❖ Racionalismo, basado en las divisiones lógicas.
- ❖ Empiricismo, a partir de análisis estadísticos.
- ❖ Historicismo, por clasificar considerando los desarrollos.

61 Cit. N. J. Williamson. An Interdisciplinary world and Discipline Based Classification. *Structures and relations in knowledge organization: Proceedings for the Fifth International ISKO Conference, 25-28 August*. Ergon – Verl, 1998, Lillie – France; ed. Hadi [et al.]: Würzburg. Advances in knowledge Organization, 1998, 6.

62 D. W. Langridge. *Classification, its kinds, systems, elements, and applications*. London [etc.]: Bowker Saur, 1992.

63 Sólida argumentación sobre los desfases de la clasificación tradicional por parte de J. Mills & V. Broughton. *Bliss Bibliographic Classification. Second Edition. Introduction and Auxiliary Schedules*. London: Butterworths, 1977. Pese a su antigüedad mantiene la lúcida vigencia de un análisis muy coherente.

64 Presenta un elaborado cuadro de los diferentes métodos de clasificación fundamentales B. Hjørland. Theory and Metatheory of Information Science. A New Interpretation. *Journal of Documentation*, 1998, 54, 5 : 606-621, en p. 612.

❖ Pragmatismo, basado en el análisis de las finalidades y las consecuencias,

En la sociedad informacional la multidisciplinariedad es la norma,⁶⁵ que genera nuevos campos de conocimiento y asociaciones interdisciplinaria entre ellos, incapaces de ser representados por la organización académica de disciplinas. Si antes importaba discriminar el saber, ahora es preferible hacer relevantes las similitudes entre las disciplinas y la interdisciplinariedad entre áreas de conocimiento. Se desencadena así un proceso de conocimiento muy dinámico por la misma interdisciplinariedad: un *interés* suscita un nuevo campo multidisciplinario de conocimiento, que luego puede continuar como un área de conocimiento interdisciplinario y derivar por fin hacia una nueva disciplina.

A partir de los nuevos fundamentos epistemológicos multidisciplinarios,⁶⁶ la clasificación debe asumir la *flexibilidad* como un instrumento imprescindible para responder con eficacia y rapidez a cualquier posible demanda en referencia directa al entorno intelectual y de conocimiento, tanto de los productores como de los usuarios. Los sistemas de recuperación de información deben adecuarse a cada uno de los proyectos y finalidades de los usuarios, así como acomodarse a las nuevas identidades disciplinarias y áreas multidisciplinarias, que no son ni estáticas ni homogéneas. La clasificación se orienta, entonces, en opinión de B. Hjørland,⁶⁷ hacia tres perspectivas analíticas:

- ❖ Adaptación a las necesidades informativas concretas de las colectividades *factual y usual*⁶⁸, en el marco del *constructivismo social*.
- ❖ Análisis del contenido, con especial hincapié en el estudio del comportamiento durante la búsqueda de información, según los diferentes dominios del conocimiento.
- ❖ Estudio de la funcionalidad de los sistemas de recuperación de información, con el fin de enmendar un funcionamiento inadecuado.

El objetivo de la adopción de estas perspectivas es doble:

- ❖ Proporcionarle soporte a una nueva epistemología, en la que los fenómenos han de ser clasificados en el contexto de las actividades e intereses del usuario, en flagrante contradicción con el empirismo y el racionalismo clásicos.
- ❖ Diseñar la organización de una *nueva epistemología pragmática*, para optimizar el aprendizaje y el uso del conocimiento.

65 Documentada afirmación de C. Beghtol. General Classification Systems: Structural Principles for Multidisciplinary specification. *Structures and relations in knowledge organization: Proceedings for the Fifth International ISKO Conference, 25-29 August* Ergon – Verl, 1998, Lille – France; ed. By Hadi [et al.]: Würzburg, 1998.

66 Documentada afirmación de C. Beghtol. General Classification Systems: Structural Principles for Multidisciplinary specification. *Structures and relations in knowledge organization: Proceedings for the Fifth International ISKO Conference, 25-29 August* Ergon – Verl, 1998, Lille – France; ed. By Hadi [et al.]: Würzburg, 1998.

67 *Theory and Metatheory...op. cit.*

68 En feliz y estimulante definición que ha hecho B. Calenge. *Les politiques d'acquisition: constituer une collection dans une bibliothèque*. París: Circle de la Librairie, 1994.

En este nuevo contexto, siguiendo las consideraciones de C. Beghtol⁶⁹, la clasificación debe atender tres retos:

- ❖ Otorgarle una nueva función a las disciplinas, como principio prioritario estructural en la clasificación, dentro de la epistemología pragmática.
- ❖ Reconocer la distinción entre la *ficción* y *no – ficción*, como principio secundario estructural en la clasificación.
- ❖ Lograr que las técnicas de recuperación de información respondan de la forma más apropiada a una interrogación *online* sobre documentos electrónicos.

B. Hjørland y H. Albrechtsen⁷⁰ han realizado un profundo análisis sobre las implicaciones de estos tres retos, he aquí sus conclusiones:

- 1º En la nueva epistemología pragmática, la clasificación no es descriptiva y neutral respecto a la definición y organización de las disciplinas y las *comunidades discursivas*, sino que se transforma en un instrumento del constructivismo social en la generación del conocimiento. La ciencia no es un simple reflejo de la realidad, sino una institución social que manifiesta una ideología en la investigación del entorno. La clasificación consiste, entonces, en evaluar los valores, metas y resultados de hacer ciencia en un determinado sentido. La investigación en la clasificación tiende actualmente a adoptar una posición metodológica de *realismo pragmático*: realismo, cuando la clasificación le concede prioridad al *fundamento*, es decir a los procesos; pragmatismo, cuando la clasificación atiende a la aplicación: En ambas dimensiones de la clasificación, la fundamental y la aplicada, el referente común es el *orden natural*, definido por C. Beghtol.⁷¹
- 2º Epistemológicamente, tiene perfecto sentido la distinción entre *ficción*, para referirse a toda la información sobre planteamientos científicos, respecto de la *no – ficción*, en referencia a toda la información sobre técnica aplicada. En los sistemas electrónicos modernos de información, donde la polirrepresentación es la constante, esta división es muy útil.
- 3º En los actuales sistemas de recuperación *online* se plantea la disyuntiva de si la información debe ser extractada a partir de los documentos o si los documentos *online* deben permanecer como *unidades documentarias*. Los semiólogos J. Kristeva primero, M. M. Bakhtin después,⁷² concibieron el texto como un mosaico, un traspunto de otros varios textos a los que está referido. El texto se compone de unas *unidades* capaces de ser integradas en otros textos, premisa que ha sido amplia y profundamente estudiada en el análisis documental de contenido. La investigación en clasificación tiende a desarrollar una mejor comprensión de los diferentes

69 Knowledge Domains: Multidisciplinary and Bibliographic Classification Systems. *Knowledge Organization*, 1998, 25, 1-2 : 1-9.

70 An Analysis of Some Trends in Classification Research. *Knowledge Organization*, 1999, 26, 3 : 131-139.

71 Knowledge Domains... *op. cit.* p. 2.

72 *Speech genres and other late essays*. Ed. by C. Emerson & M. Holmquist. Austin: University of Texas Press, 1986.

puntos de acceso de materia en el campo de las anotaciones, para ser utilizados con una mayor pertinencia en la recuperación interactiva. Esta tendencia implica un profundo conocimiento en lo concerniente a la producción de documentos y la valoración de la información.

La constatación de los problemas expuestos y el planteamiento de los retos expresados, han hecho que la investigación en clasificación pergeñe tres vías resolutivas

1ª La utilización de los lenguajes de clasificación existentes (CDU, LCC, DDC) a los que, sin embargo, les hayan sido incorporadas ciertas innovaciones para su adaptación al nuevo entorno. Este es el caso del proyecto para transformar el DDC en una guía de materias en Internet, mediante:

- ❖ La incorporación de un glosario terminológico complementario al DDC, para un uso más eficaz en Internet.
- ❖ La revisión de los títulos del DDC, para obtener una mayor expresividad y capacidad de difusión.
- ❖ La elaboración de prototipos de la Guía Temática Dewey en la red, a partir de los resultados de las dos actividades anteriores.⁷³

Es ésta una tendencia que sigue los pasos de las innovaciones de los tesauros en Internet, con base en las cuales se prefiere incidir en una nebulosa de asociaciones y diluir la organización jerárquica, de tal modo que las remisiones RT, USE, UP, son sustituidas por *Considere también*, o, mucho más novedosamente, *Elegido desde*.⁷⁴

2ª La adaptación de los mayores sistemas de clasificación existentes, a través de tres proyectos de revisión:

- ❖ Bliss Bibliographic Classification 2, que incorpora la notación retroactiva, junto con una estructura facetada, emplazamientos opcionales y *clases de fenómenos* para tópicos sin disciplina.
- ❖ Reconfiguración de la CDU por McIlwaine y Williamson, que utilizan la estructura facetada del BBC2. El objetivo consiste en otorgarle a la CDU una estructura que le proporcione una mayor capacidad interdisciplinaria y desarrolle en ella un área similar a la *Clase de Fenómenos* del BBC2.⁷⁵
- ❖ Desarrollo de enlaces entre descriptores de tesauros y títulos del DDC, como el proyecto de Olson, iniciado a partir de la definición del campo de estudio de la mujer.

3ª La elaboración de nuevos sistemas de clasificación capaces de ser aplicados a las necesidades plurales de la creación del conocimiento multidisciplinario y la difusión electrónica, mediante la sustitución de las disciplinas académicas por tipos de textos y el desarrollo de los *tópicos análogos* o *tópicos homólogos*. El proyecto más

73 Realiza una sistemática descripción del proyecto D. Vizine-Goetz. Dewey as an Internet Subject Guide. *Advances in Knowledge Organization*, 1998, 6 : 191-197.

74 Sugiere M. Lykke Nielsen. Future Thesauri: GAT kind of Conceptual Knowledge do Searchers Needs? *Advances in Knowledge Organization*, 1998, 6 : 51-64.

75 Una explicación por parte de uno de los actores del proyecto nos la ofrece N.J. Williamson. An Interdisciplinary world... *op. cit.*

avanzado es el *Spectrum* de la Britannica Online (1998), que reconoce dos niveles, uno primero, facetado en diez grupos, y uno segundo específico, más académico. Es un orden *viewpoint* que enfatiza la interrogación desde las perspectivas.⁷⁶

Una correcta reflexión analítica sobre estas tres vías de investigación en clasificación, nos permite apuntar la doble dimensión de adaptación que la sociedad informacional está desarrollando como campo de trabajo.

- a) El diseño de sistemas de clasificación que organicen el conocimiento siguiendo el *modelo de pensamiento relacional hipertextual*. Si bien es útil, en efecto, la incorporación de estructuras facetadas, comienza a ser evidente la conveniencia de evolucionar desde los grandes sistemas de clasificación precoordinaados hacia otros de indización por descriptores virtuales, mediante una asignación automática.⁷⁷ La finalidad entrañaría desarrollar una indización automatizada semántica que fomentara los hiperenlaces y vínculos dinámicos⁷⁸ hasta generar una red de hipertesoros.⁷⁹ El objetivo de este diseño relacional en la clasificación no es otro sino el de asegurar relevancia en la recuperación de información, que ya no procede sólo de unidades físicas de información sino de un compendio innúmero e intangible de recursos electrónicos. Si el desarrollo de los metadatos se orienta a mejorar la recuperación de información, la investigación sobre lenguajes de marca, como el RDF,⁸⁰ busca hallar alternativas para una clasificación de los nuevos recursos.⁸¹
- b) La representación del usuario y sus necesidades informativas en los sistemas de clasificación, para hacer una recuperación pertinente de la información; esto es, convertir estos sistemas en una guía experta de investigación. En este sentido es sumamente útil analizar las oraciones sustantivas de información al interrogar al usuario, para extraer las *pequeñas piezas de información textual*⁸² susceptibles de precisar

76 Expone C. Beghtol. General Classification Systems... *op. cit.*

77 Una propuesta que diseñó con precisión A. Bayarri. *La CDU en el contexto de las bases de datos documentales: indización automática con descriptores virtuales, asignados a partir del contenido. Desarrollo de un prototipo en el sistema de micro – ISIS de UNESCO*. Madrid. CINDOC, 1996.

78 Es sumamente ilustrativa la comunicación de C.G. Figuerola & J.L. Alonso Berrocal & A.F. Zazo Rodríguez. Nuevos puntos de vista en la Recuperación de Información en la Web. *Los sistemas de información al servicio de la sociedad: actas de las jornadas. VI Jornadas Españolas de Documentación*. Valencia: FESABID, 1998: 273-280.

79 L. West & P. Murray-Rust. Steps towards the global linking of knowledge. *Managing Information*, 1998, p. 36-39.

80 Cifr. T. Bray. RDF and Metadata. *XML.co*. June, 1998. Disponible en Internet <http://www.xml.com/xml/pub/98/06/rdf.htm> [Consulta 19 abril 2001]. También en E. Miller An Introduction to the Resource Description Framework. *D – Lib Magazine*. May, 1998. Disponible en Internet <http://www.dlib.org/dlib/may98/miller/05miller.html>.

81 Analiza E. Méndez Rodríguez. RDF: un modelo de metadatos flexible para las bibliotecas digitales del próximo milenio. *7es Jornades Catalanes de Documentació. Les biblioteques i els centres de documentació segle XXI: peça clau de la societat de la informació*. Disponible en Internet: <http://www.bib.uc3m.es/~mendez/publicaciones/7jc99/rdf.htm> [Consultado 29 de abril 2001].

82 Cifr. J.A. Sillines. Argumentation-based indexing for information retrieval from learned articles. *Journal of Documentatio*, 1992, 48, 4 : 387-401. También C.D. Paice. A Thesaural Model of Information Retrieval. *Information Processing & Management*, 1991, 27, 5 : 433-447.

la recuperación por contenido contextual. Esta orientación adquiere su plena eficacia mediante el desarrollo de redes neuronales en los SARI, así como en los *topic maps*. Una vía de investigación muy atractiva será aquella que se dirija a la representación del usuario mediante la introducción de categorías definidas a partir del estudio de usuarios. La consecución de esta representación exige la definición de un paradigma en donde el usuario sea un elemento activo, así como una metodología que materialice la presencia del usuario a través de la consideración de aspectos subjetivos (intencionalidad en la búsqueda) y aspectos objetivos (tema de la necesidad informativa).

TENDENCIAS EN EL ANÁLISIS DE IMÁGENES

Es fácilmente perceptible la forma en que la imagen adquiere en la sociedad informacional una vigorosa y nueva dimensión en sus funciones y en su valor. La imagen ha redescubierto su efectividad semiótica en la comunicación y se orienta ya hacia la constitución de un nuevo código que precisará sus nuevos modos de lectura y escritura.⁸³ Es también perfectamente constatable cómo los notables progresos en la tecnología sobre imágenes digitales, así como el masivo almacenamiento microinformático, facultan para un correcto tratamiento de la imagen. Estas capacidades, semiótica y tecnológica, se suman a la necesidad de reutilizar imágenes, por lo que ha llegado a ser imperioso el desarrollo de sistemas de recuperación de información gráfica.

La recuperación de información gráfica ha seguido dos modelos estratégicos:

- 1º Técnicas de recuperación conceptual sobre los textos que identifican la imagen (título, pie de foto) o que describen la imagen, aplicando las técnicas de indización diseñadas para documentos texto–lingüísticos y que constituyen un proceso analítico–descriptivo inadecuado para la naturaleza de la imagen.⁸⁴
- 2º Técnicas de recuperación automática de imágenes en la órbita de las tecnologías de tratamiento de la imagen digital, que Enser agrupó en cuatro modelos, a saber: lingüístico–lingüístico, visual–visual, lingüístico–visual y modelo mixto,⁸⁵

83 Son sumamente interesantes a este respecto los trabajos de J.A. Ortega Carrillo. *Alfabetización visual y desarrollo de la inteligencia: programa de intervención didáctica basado en el entrenamiento de capacidades perceptivo – visuales, grafo – motrices y de lectura crítico – analítica de la imagen*, Granada: Fundación Educación y Futuro: Fundación Genesian, 1996. También el Proyecto Fin de Carrera presentado en la Facultad de Humanidades, Comunicación, Documentación de la Universidad Carlos III Madrid del D.F. Arranz Lago. *Internet y el Libro electrónico frente a la cultura tradicional*. Getafe: D. F. Arranz Lago, 2000.

84 Hacen especial hincapié en este desfase T. Kato & T. Kurita. *Visual interaction with Electronic Art Gallery*. En: M. Tjoa & R. Wagner (eds.). *Database and Expert Systems Applications. Proceedings of the International Conference: Viena, 1990*; O. Firschein. *Image Understanding Program at ARPA. IEEE Expert*, 5, 10 (October 1995) : 8-38; T. Kanade. *Inmersion into Visual Media. New Applications of Image Understanding. IEEE Expert*, 1: 11 (February 1996) : 73-80; H. Treat. *Searching Images Using Multimedia anager. Information Services & Use*, (1996) , 16 : 15-24.

85 Cifr. *Progress in Documentation Pictorial Information Retrieval. Journal of Documentation*, 51 : 2 (June 1995) : 126-171.

ampliamente analizados por Robledano Arillo.⁸⁶ Un repaso al estado de la cuestión acredita que en el momento presente y en un futuro próximo, la inteligencia artificial debe recurrir a las técnicas de representación normalizada de conceptos y al contenido icónico en los lenguajes documentales, pero con un claro sesgo hacia la semiología de la imagen.

Es evidente que el desarrollo de la recuperación de imágenes, dentro de la recuperación de información, sería más efectivo si se elaborasen investigaciones que integraran ambas tendencias, pero no existe una clara determinación en esta vía metodológica. Esto explica que la recuperación de imágenes siga dos sistemas:

Sistemas visuales automáticos

Pensados para una aplicación en dominios restringidos estos sistemas utilizan funciones de similitud por atributos de color, textura y estructura de la imagen y de los objetos de la imagen,⁸⁷ caracteres *primitivos* en opinión del profesor Eakins, en tanto que no se ha investigado con suficiencia el reconocimiento automático de objetos y contenido icónico, los que comportan un nivel de abstracción mucho mayor. Apenas se están esbozando proyectos para reconocer tipos de objetos y escenas concretas.

Estos sistemas encuentran en su desarrollo problemas matemáticos de tipo algorítmico (generar un vector de características gráficas para cada atributo visual de recuperación, así como un algoritmo de similitud entre estos vectores), pero su limitación más significativa está en su campo de aplicación. La tendencia más común es la aplicación en el *contenido intrínseco de la imagen*; esto es, el conjunto de atributos derivados de los caracteres materiales de la imagen, bien mediante la distribución en el espacio de las sustancias constitutivas de la imagen tradicional, bien en el análisis de distribución y relación entre los píxeles de la imagen electrónica. Una recuperación de imágenes sería tanto más eficaz cuanto considerase el *contenido extrínseco de la imagen*; esto es, el contenido conceptual de la imagen, cuya abstracción intelectual es producto de la acción humana en una fase de representación cognitiva, o en la alimentación de una base de conocimientos capaz de asociar conceptos y patrones gráficos.

El contenido extrínseco de la imagen tiene como elementos estructurales la percepción de la imagen, que relaciona la forma de la imagen y los patrones cognitivos del indizador⁸⁸ y se proyecta en el espacio icónico (permitiendo inferencias a partir de la apariencia) y el espacio plástico (orden de valores cromáticos y compositivos); y el texto anejo o referido a la imagen. La lectura de la imagen estará definida por la interrelación entre los espacios plástico, icónico y textual.

86 Tendencias actuales y perspectivas futuras en la recuperación de información gráfica. En: F. J. García Marco (ed.) *Actas del III Encuentro de ISKO – España, Getafé, 19-20 de noviembre de 1997*. Zaragoza, 1997, p. 119-132.

87 Cit. J. Eakins. *Automatic Image Content Retrieval: are we getting any where?* Abstract of paper presented at the 3rd ELVIRA conference: May 1996.

88 Analizan J. R. Hilerá & V. J. Martínez. *Redes Neuronales Artificiales: Fundamentos, Modelos, Aplicaciones*. Madrid: Rama, 1995, p. 23.

La recuperación en un sistema automático visual mediante atributos extrínsecos reclama el diseño de una base de conocimiento en la que se puedan asociar conceptos con íconos o atributos plásticos. Las bases de conocimiento almacenan de manera formalizada representaciones simbólicas de reglas o de hechos y siguen un modelo muy simple consistente en el empleo de reglas de producción silogística: IF (condición) y THEN (conclusión). Mediante este modelo la base de conocimiento relaciona atributos intrínsecos (color, textura, contorno) de fragmentos de imágenes aislados previamente y representados matemáticamente, con conceptos susceptibles de ser representados por imágenes fácilmente aislables.

Este sistema de recuperación mediante atributos extrínsecos sólo es factible en dominios restringidos, por cuanto que es inabarcable representar todo el conocimiento visual susceptible de manifestarse en cualquier imagen de modo tan simple y válido para cualquier usuario y uso. Es preciso, además, que la generación de la base de conocimiento sea humana, a la que luego seguirá un proceso automático de aprendizaje del propio sistema para la generación automática de nuevas reglas que mejoren la base. Es problemático también, por último, representar matemáticamente de modo tan simple los objetos, puesto que los objetos no formalizados presentan una morfología muy variada.

Sistemas visuales mixtos

Permiten búsquedas en las que se emplean conjuntamente las técnicas del modelo de recuperación visual – visual y las representaciones textuales por indización de imágenes en una base de datos textual, ya sea relacional (Multimedia Manager), ya documental (Corbis).

Sin duda son los sistemas que mejor se adecuan a la tendencia actual en los sistemas de recuperación de imagen, dada su capacidad para representar y recuperar contenidos con distintos grados de abstracción. Son sistemas muy útiles para recuperar imágenes sobre un tema cuyos atributos visuales son difíciles de definir; y además reducen el tiempo de recuperación al ampliar la capacidad de selección del usuario entre imágenes similares.

Los sistemas visuales mixtos presentan actualmente tres modalidades, a partir siempre de un sistema visual puro:

Integración con un sistema lingüístico

Permiten una recuperación mediante consultas visuales y lingüísticas, utilizando tres estrategias:⁸⁹ un ojeo de imágenes a partir de categorías temáticas; consultas

89 Han sido tipificadas y descritas por C. Olivia Frost *et al.* Research on Human – centered Design and Evaluation of Multimode Image Retrieval. En: Fourth Delos Workshop: *Image Indexing and Retrieval*, San Miniato, Pisa, Italy, 28-30 August, 1997, ERCIM Workshop Proceedings nº 97- W004. Disponible en Internet <http://www.ercim.org/publication/ws-proceedings/DELOS4/frost.pdf> [consulta 24 de abril 2001].

fundamentadas en el texto, mediante palabras clave, y consulta visual, a través de una *imagen-ejemplo*, como guía en la estrategia de búsqueda; sumamente útil cuando el usuario desconoce la terminología específica del área de interés solicitada en su interrogación.

Este modelo se utiliza en el sistema Retrieval Ware Image Server, de Excalibur Technologies, así como la aplicación Ultimedia Manager, con tecnología QBIC de IBM, que permite analizar y recuperar imágenes a partir del color, textura, forma y posición del objeto gráfico, en tanto que atributos. La consulta puede realizarse, bien cumplimentando campos relacionales de la base de datos textual asociada a imágenes, bien permitiéndole al usuario seleccionar atributos para recuperar por coincidencia, incluso permitiéndole dibujar formas con el ratón en la pantalla.

Integración con un modelo de representación visual y recuperación lingüística

Prototípico es el sistema ART MUSEUM, donde se encuentran integrados dos módulos:

- ❖ **QVE** o de la imagen. Recupera a partir de caracteres morfológicos de las imágenes digitalizadas empleando gráficos visuales aplicados sobre un índice de contornos de todas las imágenes almacenadas por el sistema, que el usuario puede dibujar esquemáticamente e introducir en el sistema mediante una pizarra electrónica.
- ❖ **QVD** o de usuario. Relaciona la descripción de sensaciones de un usuario ante la imagen con las imágenes que provocan esa misma sensación en el usuario a partir del color, el elemento más representativo de las sensaciones en las imágenes. El método consiste en el registro del usuario y de los adjetivos que utilizó. El sistema relaciona los conceptos abstractos sensitivos del usuario con los atributos formales y morfológicos que los han inferido.

Integración con un modelo de clasificación sobre el texto anejo a la imagen

Modelo presente en los sistemas WebSEEK, Visual SEEK y Web Chip,⁹⁰ dirigidos a imágenes tomadas de la red. El método analítico consiste en extraer caracteres visuales por histogramas y regiones de color, para luego analizar el texto anejo a la imagen mediante técnicas de indización automatizada contra categorías temáticas elaboradas por indización humana.

Todos estos modelos, sin embargo, adolecen de un mismo inconveniente: la escasa insistencia en una investigación más profunda de las necesidades de los usuarios para la consecución de unas estrategias de búsqueda más pertinentes, así como una escasa evaluación de los sistemas de recuperación de imágenes desde la perspectiva del usuario.

90 Según exponen Shih – Fu Chang *et al.* Finding Images/Video in Large Archives. Columbia's Content Based Visual Query Project. *D – Lib Magazine*, February, 1997. Disponible en Internet <http://mirrored.ukoln.ac.uk/lis-journals/dlib/dlib/dlib/february97/columbia/02chang.html>.

Tendencias en el análisis automático de imágenes de video.

Campo de investigación en notable desarrollo pese a no haber diseñado todavía técnicas automáticas de análisis de imágenes, capaces de recuperar por atributos semánticos de alto nivel de abstracción como los iconográficos e iconológicos. El actual estado de la tecnología informática no permite su uso en exclusividad, sino que la recuperación automática de imágenes de video ha de ser un complemento en la labor de análisis documental por parte de documentalistas, quienes identifican conceptos y realizan la transcodificación de la imagen al texto lingüístico, salvo en los casos en que las imágenes de video van acompañadas de texto ilustrativo de su contenido, en cuyo caso es posible usar sistemas de clasificación o indización automáticos.⁹¹

Empero, la investigación en torno al análisis automático de imágenes de video está captando un creciente interés científico,⁹² de tal modo que son reseñables seis áreas de investigación:

Reconocimiento automático de secuencias

Segmentación de la totalidad de un video en unidades temáticas menores. La fragmentación exige un complejo análisis de contenido de alto nivel semántico, por lo que este sistema sólo permite resultados cuando se representa un texto explicativo de la imagen o se acompaña el video de una transcripción del texto hablado.

Segmentación temporal de secuencias de video

Identifica los planos dentro de las secuencias de video mediante parámetros formales capaces de detectar cambios de plano.

Análisis de movimiento de cámara y objetos que componen la imagen

Permite segmentar planos de larga duración en unidades menores y más homogéneas, por lo que es muy útil para identificar los fotogramas más representativos.

Identificación de los valores de encuadre y del tipo de objetivo utilizado

Vía de investigación muy compleja por cuanto precisa técnicas específicas de análisis semántico de las imágenes.

91 Registro de resultados concretos, en este caso, hallamos en R. K. Srhari. Automatic Indexing and Content – Based Retrieval of Captioned Images. *Computer*, (September 1995) 28, 9 : 49-56; también en N. C. Rowe. Precise and Efficient Retrieval of Captioned Images: The MARIE Project. *Library Trends*, (Feb. 1999) 48, 2 : 475-495.

92 Ofrecen un estado de la cuestión M. J. Witbrock & A. G. Hamptmann. Speech Recognition for a Digital Video Library. *Journal of the American Society for Information Science* (May 1998) 49, 7 : 619-632; Ph. Aigrain & H. Zhang & P. Dragutin. Content – Based Representation and Retrieval of Visual Media: A State – of – the – Art Review, *Multimedia Tools and Applications*, 1996, 3 : 179-202; D. P. Huijmans & A. W. M. Semeulders (eds.). *Visual Information and information Systems. Third international conference VISUAL '99 Amsterdam, The Netherlands, June 2-4 1999 Proceedings*. Berlin [etc]: Springer, 1999.

Análisis del sonido

Campo de investigación que apunta hacia tres elementos analíticos:

- ☞ Detección, segmentación y representación de las distintas categorías de sonido en un video.
- ☞ Identificación de los hablantes.
- ☞ Transcripción del discurso oral, si bien este discurso debe ser suficientemente extenso y explícito del contenido. Es el elemento que más resultados está proporcionando en la investigación, como demuestra el proyecto Informedia,⁹³ donde el usuario puede utilizar consultas visuales y textuales.

CONCLUSIONES

La exposición realizada a lo largo del presente artículo nos permite pergeñar los siguientes como los hitos más reseñables de nuestra argumentación:

1. En la sociedad informacional el conocimiento se genera, comunica y representa de modo muy distinto a la jerarquización taxonómica que organizó el conocimiento y clasificó las ciencias desde el siglo XIX, modelo sobre el que se diseñaron los primeros lenguajes de clasificación.
2. Estos lenguajes de clasificación, pues, deben responder a dos interrogantes: la eficacia en incorporar la conducta y el contexto cultural del usuario en los sistemas de información; plantear y definir qué es *materia* en el mensaje informativo del hiperdocumento.
3. Las ciencias protagonizan una vigorosa y doble orientación evolutiva hacia la especialización, pero también hacia la transversalidad. Los nuevos fenómenos de interdisciplinariedad, multidisciplinariedad y pluridisciplinariedad afectan la capacidad organizativa de los sistemas de clasificación.
4. La clasificación debe adoptar una *flexibilidad* condicionada por el constructivismo social, la pujanza del análisis de contenido y los estudios de funcionalidad de los sistemas de recuperación de información.
5. La clasificación, pues, tiene planteados tres contenciosos en la sociedad informacional: debe transformarse en un instrumento del constructivismo social; debe insistir en la diferencia entre *ficción* (información científica) y *no-ficción* (técnica aplicada), y debe considerar el documento electrónico como *unidad documentaria*, esto es, un mosaico de mensajes.
6. Las vías resolutivas apuntadas se dirigen hacia: el uso de los lenguajes de clasificación existentes, con innovaciones de adaptación al nuevo entorno; el rediseño de los lenguajes de clasificación existentes, y la elaboración de nuevos lenguajes de clasificación.

93 Describen M. J. Witbrock & A. G. Hamptmann, *op. cit.*

7. La recuperación de información gráfica ha seguido dos modelos estratégicos: técnicas de recuperación conceptual y recuperación automática.
8. Estos modelos, no integrados, originan dos sistemas de recuperación de imágenes: los sistemas visuales automáticos (similaridad por atributos de color, textura y estructura de la imagen), y los sistemas visuales mixtos que usan la imagen y el texto en la recuperación, y cuyas modalidades son: la integración con un sistema lingüístico; la integración con un módulo de representación visual y recuperación lingüística; y la integración con un modelo de clasificación sobre texto anejo a la imagen.
9. El análisis automático de imágenes de video aún debe utilizar la transcodificación de la imagen a texto, si bien se están desarrollando seis áreas de investigación: reconocimiento automático de imágenes; segmentación temporal de secuencias; análisis de movimiento de cámara; identificación de los valores de encuadre; análisis de sonido; y resumen automático.

BIBLIOGRAFÍA.

- Ph. Aigrain ; H. Zhang ; P. Dragutin. Content – Based Representation and Retrieval of Visual Media: A State – of – the – Art Review, *Multimedia Tools and Applications*, 3, 1996, p. 179-202.
- H. Albrechtsen ; B. Hjörland. Information Seeking and Knowledge organization. *Knowledge Organization*, 24, nº 3, 1997, p. 135-6.
- M. M. Bakhtin. *Speech genres and other late essays*. Ed. By C. Emerson & M. Holmquist. Austin: University of Texas Press, 1986.
- R. Barglow. *The Crisis of the Self in the Age of Information: Computers, Dolphins and Dreams*. London: Routledge, 1994.
- A. Bayarri. *La CDU en el contexto de las bases de datos documentales: indización automática con descriptores virtuales, asignados a partir del contenido. Desarrollo de un prototipo en el sistema de micro – ISIS de UNESCO*. Madrid. CINDOC, 1996.
- C. Beghtol. General Classification Systems: Structural Principles for Multidisciplinary Specification. *Structures and relations in knowledge organization: Proceedings for the Fifth International ISKO Conference, 25-29 August Ergon – Verl, 1998, Lille – France*; ed. By Hadi [et al.]: Würzburg, 1998.
- –. Knowledge Domains: Multidisciplinary and Bibliographic Classification Systems. *Knowledge Organization*, 25, nº 1-2, 1998, p. 1-9.
- T. Bray. RDF and Metadata. *XML.co*. June, 1998. Disponible en Internet <http://www.xml.com/xml/pub/98/06/rdf.htm> [Consulta 19 abril 2001].
- B. Calenge. *Les politiques d'acquisition: constituer une collection dans une bibliothèque*. París: Circle de la Librairie, 1994.

- M. Caridad (coord.). La Sociedad de la Información. Madrid: Editorial Centro de Estudios Ramón Areces SA, 1999.
- E. Carpenter & M. McLuhan. *El Aula sin muros*. Barcelona: Laia, 1974.
- M. Castells. *La Era de la Información. Economía, Sociedad y Cultura*. Madrid: Alianza Editorial, 2000, vol. 1, p. 359-408.
- G. Charton. Reperage de l'information sur Internet: Nouveaux utils, approches bibliothéconomiques et microstructures. En: *CDI virtuel: utilisation des réseaux pour l'accès à la communication. Stage national DISTNB/Sousdirection des Technologies Nouvelles CRDP de Reims, 29 Avril 1997*. Paris: Centre de Calcul, Recherche et Réseaux Jussieu, URFIST, 20 de junio 1999 [Consulta 22 de marzo de 2001.
- R.S. Clegg ; G. Palmer (eds.). *The politics of management knowledge*. London [etc.]: Sage, 1996.
- I. Dahlberg. Current trends in knowledge organization. En: *Organización del conocimiento en sistemas de información y documentación: actas del I Encuentro de ISKO – España. Madrid 4 y 5 de noviembre de 1993/ ed. F.J. García Marco*. Zaragoza: Librería General, 1995: 7-26.
- C. Durieux. La Recherche Terminologique en Traduction: Pour une approche Hipertextuelle. *Meta*, XIII, nº 4, 1997.
- J. Eakins. *Automatic Image Content Retrieval: are we getting any where?* Abstract of paper presented at the 3rd ELVIRA conference: May 1996.
- P. Enser. Progress in Documentation Pictorial Information Retrieval. *Journal of Documentation*, 51 : 2 (June 1995), p. 126-171.
- M.A. Esteban Navarro ; F.J. García Marco. Primeras Jornadas sobre Organización del Conocimiento: Organización del Conocimiento e Información Científica. *Scire*, 1, nº 1, (enero – junio 1995), p. 149-157.
- P. Fayard. ¿Exceso de información o ceguera estratégica? Consideraciones sobre la Inteligencia y el Conocimiento. *Telos*, vol. 44, 1996, p. 11-14.
- J. Feather. *The Information Society*. 2nd edition. London: Library Association Publishing Ltd. 1998.
- C.G. Figuerola ; J.L. Alonso Berrocal ; A.F. Zazo Rodríguez. Nuevos puntos de vista en la Recuperación de Información en la Web. *Los sistemas de información al servicio de la sociedad: actas de las jornadas. VI Jornadas Españolas de Documentación*. Valencia: FESABID, 1998: 273-280.
- O. Firschein. Image Understanding Program at ARPA. *IEEE Expert*, 5: 10 (October 1995), p. 8-38.

- M. Gisbert Cervera *et al.* Entornos Virtuales de Enseñanza – Aprendizaje: Proyecto GET. *Cuadernos de Documentación Multimedia*, nº 6-7, 1998. Disponible en Internet
<http://www.ucm.es/info/multidoc/multidoc/revistacuad6-7/evea.htm>.
- J. R. Hilera ; V. J. Martínez. *Redes Neuronales Artificiales: Fundamentos, Modelos, Aplicaciones*. Madrid: Rama, 1995, p. 23.
- J.R. Hilera ; S. Otón ; J. Martínez. *Aplicación de la Realidad Virtual en la enseñanza a través de Internet*. Disponible en Internet
<http://www.ucm.es/info/multidoc/multidoc/revista/num8/hilera-oton.html>. [Consulta 22 de marzo de 2001].
- B. Hjörland. *Information Seeking and Subject Representation: An activity – theoretical approach to information science*. Westport; Connecticut; London: Greenwood, 1997.
- B. Hjörland. Theory and Metatheory of Information Science. A New Interpretation. *Journal of Documentation*, 54, nº 5, (1998), p. 606-621.
- B. Hjörland. An Analysis of Some Trends in Classification Research. *Knowledge Organization*, 26, nº 3, 1999, p. 131-139.
- D. P. Huijmans ; A. W. M. Semeulders (eds.). *Visual Information and information Systems. Third international conference VISUAL '99 Amsterdam, The Netherlands, June 2-4 1999 Proceedings*. Berlin [etc]: Springer, 1999.
- J. M. Izquierdo Alonso ; J. Ruiz Abellán ; J.T. Piñera Lucas. Los estudios de usuarios en los programas de gestión de calidad. Propuesta de un marco teórico integrador para el estudio del usuario de información. En: *Los sistemas de información al servicio de la sociedad: actas de las jornadas. VI Jornadas Españolas de Documentación*. Valencia: FESABID, 1998: 433-444.
- L. Joyanes Aguilar. *Cibersociedad: los retos sociales ante un nuevo mundo digital*. Madrid: McGraw-Hill, 1997.
- T. Kanade. Inmersion into Visual Media. New Applications of Image Understanding. *IEEE Expert*, 1: 11(February 1996), p. 73-80.
- T. Kato ; T. Kurita. *Visual interaction with Electronic Art Gallery*. En: M. Tjoa ; R. Wagner (eds.). *Database and Expert Systems Applications. Proceedings of the International Conference*: Viena, 1990.
- D.A. Klein. *Decision – analytic intelligent systems automated explanation and knowledge acquisition*. Hillsdale; New Jersey: Lawrence Erlbaum Associates, 1994.
- C. Kuhltham. Inside the search process: information seeking from the user's perspective. *Journal of the American Society for Information Science*, 42, nº 5, 1995, p. 361-371.

- D. W. Langridge. *Classification, its kinds, systems, elements, and applications*. London [etc.]: Bowker Saur, 1992.
- E. Lorente Bilbao. *Las Comunidades virtuales de enseñanza - aprendizaje*. Disponible en Internet <http://www.ucm.es/info/multidoc/multidoc/revista/num8/eneko.html>. [Consulta 22 de marzo de 2001].
- M. Lykke Nielsen. Future Thesauri: GAT kind of Conceptual Knowledge do Searchers Needs? *Advances in Knowledge Organization*, vol. 6 1998, p 51-64.
- M. J. López – Huertas Pérez. La representación del usuario en la Recuperación de Información. *Los sistemas de información al servicio de la sociedad: actas de las jornadas. VI Jornadas Españolas de Documentación*. Valencia: FE-SABID, 1998: 521-528.
- J. C. Marcos Recio. *Fuentes de información, educación y proyectos de investigación multimedia*. Disponible en Internet <http://www.ucm.es/info/multidoc/multidoc/revista/num8/jmarcos.html>. [Consulta 22 de marzo 2001].
- F.L. Miksa. *The DDC, the Universe of Knowledge and the Post- Modern Library*. New York: Forest Press, 1998.
- E. Miller An Introduction to the Resource Description Framework. *D –Lib Magazine*. May, 1998. Disponible en Internet <http://www.dlib.org/dlib/may98/miller/05miller.html>. [Consulta 12 de abril de 2001].
- J. Mills ; V. Broughton. *Bliss Bibliographic Classification. Second Edition. Introduction and Auxiliary Schedules*. London: Butterworths, 1977.
- C. Olivia Frost *et al.* Research on Human – centered Design and Evaluation of Multimode Image Retrieval. En: Fourth Delos Workshop: *Image Indexing and Retrieval, San Miniato, Pisa, Italy, 28-30 August, 1997, ERCIM Workshop Proceedings n° 97- W004*. Disponible en Internet <http://www.ercim.org/publication/ws-proceedings/DELOS4/frost.pdf>[consulta 24 de abril 2001].
- J.A. Ortega Carrillo. *Alfabetización visual y desarrollo de la inteligencia: programa de intervención didáctica basado en el entrenamiento de capacidades perceptivo – visuales, grafo – motrices y de lectura crítico – analítica de la imagen*. Granada: Fundación Educación y Futuro: Fundación Genesian, 1996.
- C.D. Paice. A Thesaural Model of Information Retrieval. *Information Processing & Management*, 27, n° 5, 1991, p. 433-447.
- G. Qin Thana. *Psicología y lenguaje*. Madrid: ICE de la Universidad Complutense, 1994.

- J. Robledano Arillo. Tendencias actuales y perspectivas futuras en la recuperación de información gráfica. En: F. J. García Marco (ed.) *Actas del III Encuentro de ISKO – España, Getafe, 19-20 de noviembre de 1997*. Zaragoza, 1997, p. 119-132.
- N. C. Rowe. Precise and Efficient Retrieval of Captioned Images: The MARIE Project. *Library Trends*, 48:2 (Feb. 1999), p. 475-495.
- J.A. Sillines. Argumentation-based indexing for information retrieval from learned articles. *Journal of Documentation*, 48, n°4, (1992), p. 387-401.
- Shih – Fu Chang *et al.* Finding Images/Video in Large Archives. Columbia's Content Based Visual Query Project. *D – Lib Magazine*, February, 1997. Disponible en Internet <http://mirrored.ukoln.ac.uk/lis-journals/dlib/dlib/dlib/february97/columbia/02chang.html>.
- R. K. Srhari. Automatic Indexing and Content – Based Retrieval of Captioned Images. *Computer*, 28:9 (September 1995), p. 49-56.
- W. Sugar. User centered perspective of information retrieval research and analysis methods. *Annual Review of Information Science and Technology*, 30, 1995, p. 77-109.
- H. Treat. Searching Images Using Ultimeida anager. *Information Services & Use*, 16, 1996, p. 15-24.
- D. Vizine-Goetz. Dewey as an Internet Subject Guide. *Advances in Knowledge Organization*, vol. 6, 1998, p 191-197.
- L. West & P. Murray-Rust. Steps towards the global linking of knowledge. *Managing Information*, 1998, p. 36-39.
- N. J. Williamson. An Interdisciplinary world and Discipline Based Classification. *Structures and relations in knowledge organization: Proceedings for the Fifth International ISKO Conference, 25-28 August. Ergon – Verl, 1998, Lillie – France*; ed. Hadi [et al.]: Würzburg. *Advances in knowledge Organization*, vol 6, 1998.
- M. J. Witbrock ; A. G. Hamptmann. Speech Recognition for a Digital Video Library. *Journal of the American Society for Information Science*, 49 . 7, (May. 1998), p. 619-632.