

Metodología para el diseño de taxonomías corporativas

Sílvia Argudo
Miquel Centelles *

*Artículo recibido:
17 de noviembre de 2004.*

*Artículo aceptado:
1 de abril de 2005.*

RESUMEN

Se define la taxonomía corporativa como un tipo de vocabulario controlado que refleja el contexto, la audiencia y los contenidos de una organización determinada, y que permite la representación de todos sus objetos informativos para desarrollar diferentes funciones de los sitios corporativos: la organización de contenidos, la búsqueda, la navegación, la investigación competitiva, etcétera. Se propone una metodología para diseñar taxonomías en el contexto de la gestión de proyectos, con especial acento en las etapas que necesariamente deben llevarse a cabo: la planificación estratégica, la construcción de la taxonomía, la categorización de los objetos informativos, la aplicación y presentación de la taxonomía, y el ciclo de evaluación y mejora continua. Se presenta el mercado actual de soluciones informáticas que permiten automatizar estos procesos, y se caracterizan y ejemplifican los diferentes tipos de aplicaciones existentes.

Palabras clave: Taxonomías corporativas; Intranets corporativas; Organización de contenidos;

* Los dos autores pertenecen a la Facultat de Biblioteconomia i Documentació de la Universitat de Barcelona, España. (silvia.argudo@ub.edu); (miquel.centelles@ub.edu).

Información en la empresa; Estructuras conceptuales; Vocabularios controlados.

ABSTRACT

Methodology for the design of corporative taxonomies

Silvia-Argudo and Miguel-Centelles

A corporative taxonomy is defined as a type of controlled vocabulary that reflects the context, the audience and the contents of a given organization, and which allows the representation of all types of informational items for developing different functions in corporative sites: content organization, searching navigation, competitive intelligence, etc. A methodology for the design of taxonomies is set out in the context of project management, with special attention paid to the stages that necessarily must be carried out: strategic planning, construction of taxonomy, categorization of informational objects, application and presentation of the taxonomy, and the cycle of evaluation and continuous improvement. The present market of computer solutions for these processes is presented, including features and examples of different types of existing applications.

Keywords: Corporative taxonomies; Corporative intranets; Content organization; Information in the enterprise; Conceptual structures; Index vocabulary.

CONCEPTOS CLAVE A DESTACAR

- La combinación del exceso de información con el analfabetismo informacional da como resultado una baja productividad en las organizaciones.
- La creación de una taxonomía corporativa debe afrontarse como un proyecto planificado, gestionado y controlado adecuadamente.
- Es importante que todos los criterios establecidos durante el proceso de construcción de la taxonomía sean recogidos en un manual de directrices y concreciones...
- Debido al dinamismo del trinomio básico –contexto, audiencia y contenidos–, una taxonomía debe ser una herramienta en continuo cambio.
- Debe incorporarse en el ciclo evaluativo el análisis de costos y beneficios que comporta para la organización el diseño e implementación de la taxonomía.

INTRODUCCIÓN

Y sonó la señal de alarma: ¡Las empresas pierden dinero por falta de una organización adecuada de la información! Y la respuesta fue el término mágico: lo que necesitamos son *taxonomías*. Y como el contexto era la organización se añadió el adjetivo: *taxonomías corporativas*.

La información es un capital fundamental de la empresa; es imprescindible para tomar decisiones y para llevar a cabo los objetivos y funciones de la organización. Se cree que un ejecutivo puede recibir entre 200 y 300 mensajes de correo electrónico diarios; que la cantidad de información generada por un trabajador puede duplicarse cada 18 meses; que en dos años de ahora se genera más información que en toda la historia de la humanidad... Por otro lado, existe lo que se ha dado en llamar analfabetismo informacional, que se traduce en que un 85% de personas que no saben usar la estrategia booleana; que la media de términos usados para recuperar información es de 1,4; que quien realiza las búsquedas no conoce el vocabulario adecuado relacionado con la información que quiere recuperar... La combinación del exceso de información y del analfabetismo informacional da como resultado una baja productividad en las organizaciones. Un estudio de costos demostraba que, en el año 2002, el conjunto de compañías de una gran empresa perdió 12.000 millones de dólares debido a la incapacidad de localizar recursos informativos adecuados en el momento necesario y a la velocidad requerida. Y existen otros tantos intentos de traducir a cifras los costos que supone una organización ineficaz de la información. La solución, comúnmente aceptada, pasa por elaborar estructuras conceptuales formadas por términos que representan conceptos y las relaciones entre éstos.

En bibliotecas y centros de documentación, dichas estructuras conceptuales se conocen con el nombre de tesauros y sistemas de clasificación. Sin embargo, al plantearse la cuestión de organización y recuperación de información en el entorno empresarial, aparece una nueva denominación: *taxonomía corporativa*. Este nuevo término ha provocado cierta confusión y debate relacionados ambos con la respuesta a esta pregunta: ¿estamos realmente ante una nueva herramienta o sólo se trata de un nombre diferente?

En este artículo se pretende aportar alguna luz al significado del término *taxonomía corporativa*, así como realizar una aproximación sencilla a la metodología de creación de esta herramienta como base para organizar contenidos digitales en entornos corporativos. Para conseguir este objetivo se inicia el artículo con una breve contextualización del concepto de *taxonomía corporativa*, para pasar seguidamente al proceso de elaboración de la misma. El proceso se estructura en las diferentes fases o grupos de actividades que deben

estar necesariamente presentes en cualquier proyecto de elaboración de una taxonomía corporativa para que ésta tenga una mínima garantía de éxito, y se complementa con una breve descripción de soluciones informáticas existentes en el mercado. Finalmente, se concluye con un breve repaso de los aspectos más fundamentales del desarrollo de una taxonomía corporativa.

CONTEXTO DE ELABORACIÓN DE TAXONOMÍAS CORPORATIVAS

¿Qué es exactamente una taxonomía corporativa? A pesar de los intentos de algunos autores,¹ no resulta sencillo definir el término taxonomía. La mayoría de las definiciones aportadas hasta ahora son aplicables a lenguajes documentales como tesauros y sistemas de clasificación, así como a otros tipos de vocabularios controlados; desde anillos de sinónimos hasta ontologías, pasando por catálogos de autoridades.

Rosenfeld y Morville (2002, p. 383-412) exponen la experiencia del equipo de trabajo de MSWeb al construir la estructura organizativa de la información para la intranet de Microsoft. Después de analizar la bibliografía existente y comprobar la ambigüedad y confusión que existía al respecto, el equipo de MSWeb adoptó como definición de taxonomía: “cualquier conjunto de términos que comparten algún principio de organización.” Esta definición tan amplia los llevó a crear una estructura compleja a partir de tres taxonomías de carácter diverso: vocabularios descriptivos controlados, esquema de metadatos y etiquetas de categorías.

La mejor forma de comprender el concepto de taxonomía corporativa es mediante la comparación entre ésta y los lenguajes documentales tradicionales. Esta comparación se centra en lo que se denominará el trinomio básico, conformado por el contexto, la audiencia y los contenidos. Una taxonomía corporativa se define a partir del análisis de estos tres elementos y se construye alrededor de ellos; la combinación de los tres es diferente en cada organización y, por lo tanto, las estructuras conceptuales resultantes son igualmente diferentes.

Gilchrist, Kibby y Mahon (2000, p. 6) consideran que una taxonomía corporativa aspira a ser:

...una correlación de los diferentes lenguajes funcionales, regionales y nacionales utilizados por la organización para dar soporte a un mecanismo de navegación y acceso al capital intelectual de la organización...

1 Sirva como ejemplo Gilchrist (2003).

La expresión “lenguajes funcionales... utilizados por la organización” implica que la taxonomía corporativa es una herramienta personalizada, hecho que constituye una primera diferencia respecto a los tesauros y sistemas de clasificación. Los lenguajes documentales tradicionales se basan en la materia de los documentos, e intentan estructurar un campo del saber a partir del establecimiento de relaciones entre los conceptos que lo forman. Una vez elaborado, un tesoro o sistema de clasificación debe servir para indizar o clasificar cualquier documento que trate del tema estructurado, para encontrar ese documento en cualquier unidad de información y para que éste pueda ser consultado por cualquier usuario. Por su parte, una taxonomía corporativa intenta ser una estructura organizada de los contenidos precisos de una organización concreta que necesitan ser usados por una audiencia específica para realizar tareas y funciones puntuales. Es decir, se diseña de forma exclusiva para una organización en particular, con sus características y objetivos, para contener de forma integrada los recursos que son requeridos en el cumplimiento de esos objetivos, y para que esos recursos sean localizados y utilizados por los usuarios o audiencia de esa organización a partir de sus contenidos o de cualquier otro atributo.

Además de esa diferenciación general que afecta al trinomio básico, pueden destacarse otros matices que inciden en la divergencia de los dos tipos de herramienta:

- La taxonomía corporativa está pensada para organizar de forma integrada todo tipo de contenidos digitales: datos no estructurados, información elaborada, información sobre personas, servicios y productos, aplicaciones informáticas, mensajes de correo electrónico, etcétera. Una botella de vino se considera un contenido en el entorno corporativo, igual que un artículo de revista. Los dos objetos informativos tienen características que permiten describirlos e incluirlos en una estructura organizada para facilitar su recuperación. Este sentido amplio del concepto “contenido digital” pasa también por la integración de diversos formatos y morfologías de la información, asociados a aplicaciones diferentes y accesibles a usuarios diversos en función de permisos y otros asuntos de accesibilidad.
- La taxonomía corporativa está pensada no sólo para indizar o clasificar el recurso o contenido con miras a ser recuperado de forma exacta mediante una búsqueda en el sistema, sino que se considera una estructura que permite su explotación por parte de diferentes tipos de aplicaciones, aunque la tendencia actual concede más importancia a su capacidad para obtener estructuras de categorías que faciliten la

navegación. Sin embargo, estas estructuras pueden ser explotadas por sistemas de ayuda en la exploración y la navegación, por buscadores, por sistemas de filtraje y personalización de información, por aplicaciones de minería de datos, de investigación competitiva, de colaboración entre expertos, etcétera. Es decir, la taxonomía no es únicamente una estructura de conocimiento, sino que además permite el desarrollo de aplicaciones e incorpora contenidos concretos.

Aun cuando los principios que guían la estructura son los mismos que los que puedan darse en un tesoro –control del vocabulario y relaciones entre conceptos–, las diferencias comentadas hasta aquí, fueron motivo suficiente para que en el año 2000 se iniciara una revisión de la norma de construcción de tesauros monolingües norteamericana, ANSI/NISO Z39.19-1994.² El propósito de dicha revisión es adaptar la norma a la nueva situación, en la que se necesitan estructuras conceptuales que incorporen los requisitos expuestos por las organizaciones y por el nuevo entorno de acceso a la información (navegación en la WWW), así como directrices que faciliten el acceso a la información sin intermediarios más la construcción de estructuras conceptuales por parte de personas externas al ámbito de la biblioteconomía y documentación. Como se puede leer en los documentos públicos acerca de la revisión de la norma, se declara el objetivo de convertirla en una pauta para elaborar diversos tipos de estructuras conceptuales interoperables: tesauros, sistemas de clasificación, vocabularios controlados, anillos de sinónimos y taxonomías, entre otras.

METODOLOGÍA DE ELABORACIÓN DE TAXONOMÍAS CORPORATIVAS

La creación de una taxonomía corporativa debe afrontarse como un proyecto planificado, gestionado y controlado adecuadamente. Llevarlo a cabo requiere una importante inversión de recursos económicos y humanos y, por lo tanto, es imprescindible adoptar una actitud estratégica y plantearse el proyecto con el rigor necesario para conseguir los objetivos al menor precio.

El proyecto debería ser llevado a cabo por un equipo compuesto por individuos con distintas habilidades, conocimientos y puntos de vista: expertos en el área de la biblioteconomía y documentación, en el ámbito de actuación de la organización, conocedores del funcionamiento de la organización, informáticos, diseñadores gráficos, cargos directivos, representantes de los diversos departamentos o unidades de la organización, etcétera. Lo más importante,

2 Información más detallada en:
<http://www.niso.org/committees/MT-info.html>.

sin embargo, es conseguir el compromiso de todos los agentes implicados: todos deben comprender la importancia de lo que se va a efectuar, asumir la responsabilidad y estar dispuestos a participar.

Como en todo proyecto, existen grupos de actividades principales que se deben realizar siguiendo una secuencia ordenada. Estos grandes grupos de actividades o etapas son cinco: planificación estratégica; construcción de la estructura; categorización de contenidos; aplicación y presentación, ciclo de evaluación y mejora continua y, finalmente, difusión.

PLANIFICACIÓN ESTRATÉGICA

El objetivo último de esta primera fase es llegar a disponer de una definición de la taxonomía que se pretende construir. Para ello es necesario realizar un análisis de la situación de partida; se trata, en definitiva, de analizar el trío formado por el contexto, la audiencia y los contenidos.

Análisis del contexto

El objetivo de este paso es obtener una definición del entorno corporativo que contenga:

- Tipología de la organización: pequeña, mediana o gran empresa, pública o privada, de servicios, de fabricación, etcétera.
- Área de actividad a la que se dedica.
- Misión, metas, objetivos y cultura corporativa.
- Dimensiones en cuanto a miembros, clientes/usuarios, producción, facturación, etcétera.
- Organigrama y estructuras de funcionamiento (oficiales e informales).
- Políticas y procedimientos establecidos para el funcionamiento y el desarrollo de las actividades habituales.
- Recursos disponibles: presupuesto, recursos humanos e infraestructura tecnológica.
- Cuestiones legales que pueden afectar la actividad de la organización.

Para obtener estos datos se debe hacer uso de todo el material documental disponible, tanto formal como informal. Son especialmente útiles las memorias o informes de actividad, los organigramas y catálogos de lugares de trabajo con descripción de funciones, los protocolos de actuación y los manuales de procedimientos. Otra fuente de información importante que se

puede utilizar son los propios miembros de la organización, con quienes se realizarán reuniones y entrevistas que permitirán obtener una excelente información complementaria que no suele estar reflejada documentalmente.

El análisis se completará con el examen de otros aspectos relativos a la situación de partida. En primer lugar se trata de conocer la posible existencia de taxonomías previas. Éstas pueden ser visibles y formales (como un vocabulario controlado), u ocultas (como el uso de un lenguaje específico por parte del personal). Es necesario investigar si existen, cuántas hay y de qué tipo son, para qué son usadas y por quién, etcétera.

En numerosas ocasiones la organización ya dispone de un sitio web –que puede ser de carácter público (Internet) o de carácter privado (intranet o extranet)–, con una estructura de contenidos más o menos elaborada a modo de taxonomía. Este sitio debe ser analizado para determinar el tipo de portal, las dimensiones, el criterio de agrupación de contenidos, las relaciones establecidas entre categorías, las formas de presentación, el vocabulario, las aplicaciones implicadas, el grado de difusión y utilización, etcétera.

Análisis de la audiencia³

Esta fase tiene por objeto identificar los diversos grupos de usuarios (internos y externos) a los que se dirige la taxonomía, sus habilidades y comportamiento en recuperación de información, sus deseos y necesidades en cuanto a contenidos y su organización.

En primer lugar hay que identificar la comunidad global de usuarios y segmentarla en los diferentes grupos. Para ello se debe partir de los planes estratégicos de la corporación, que se deben completar con reuniones y entrevistas, y con toda la información que se ha obtenido en el análisis del contexto. En el caso de que ya exista un sitio web, resultará interesante analizar los registros de transacciones y/o estadísticas de uso si se dispone de ellos. Los criterios de segmentación o agrupación de usuarios son diversos: demográficos, geográficos, psicológicos, económicos, tecnológicos, etcétera. *A priori*, deben contemplarse todos los que resultan relevantes para la organización. Posteriormente habrá que escoger el que mejor represente toda la gama de necesidades y características del conjunto.

En segundo lugar se deben establecer las necesidades, comportamientos, habilidades y deseos de cada uno de los grupos. Existe una gran variedad de

3 Se utiliza el término audiencia para referirse al conjunto de usuarios, internos y externos, que van a utilizar la estructura. El uso de este término es habitual cuando se sitúa la cuestión en el entorno corporativo, uso originado probablemente por la traducción literal del término inglés *audience* en la literatura sobre el tema.

métodos que se pueden usar: entrevistas, cuestionarios y encuestas, discusiones de grupo, observación y análisis de protocolos en entornos reales o experimentales de recuperación, análisis transaccional, tests de usabilidad, etcétera. Una entrevista nos mostrará deseos y necesidades; un análisis transaccional o un experimento de recuperación nos mostrarán comportamientos y habilidades. Es importante elegir los métodos más adecuados para el tipo de información que se quiera obtener.

Es muy posible que el análisis de la audiencia resulte en una diversidad de grupos de usuarios que difícilmente podrá ser asumida por la taxonomía. En este caso se impone una priorización de grupos que debe ajustarse a los objetivos de la organización.

Análisis de los contenidos

La finalidad de esta etapa es obtener el inventario de contenidos, acompañado de un mapa de contenidos o de relaciones. Este inventario supone una descripción completa de las unidades de contenido mínimo que la taxonomía pretende organizar.

La unidad mínima es la porción más pequeña de contenido que exige un tratamiento individual y debe decidirse para cada caso. Aunque suele equipararse a una URL (una dirección completa de un archivo en la red), la unidad mínima puede consistir en una tipología de contenido, en una instancia o ítem concreto, o en una parte de éste.

En algunos casos, resulta útil aplicar un criterio funcional para identificar los tipos de contenidos. Se trata de identificar las funciones y establecer los contenidos implicados en la realización de cada función. Hay dos tipos de funciones: las comunes a todas las organizaciones y las específicas de cada organización. En cuanto a las generales, los cuadros de clasificación de documentos administrativos pueden resultar de gran ayuda, ya que muestran las categorías tradicionales de documentos corporativos con las que nos vamos a encontrar. En cuanto a las funciones específicas, éstas se obtienen con el análisis detallado de la organización.

Se puede elaborar el inventario de contenidos a partir del análisis de una muestra representativa. Se trata de ir recopilando algunas instancias de objetos de información de diversos tipos a partir de criterios varios: recursos en diferentes formatos y/o tipologías documentales, generados por distintas fuentes, sobre diversas materias, de diferente extensión o volumen, en más de una lengua, para audiencias específicas, etcétera.

Para cada unidad mínima de contenidos hay que determinar los atributos, metadatos o características descriptivas que, conjuntamente, permiten

identificarla y, a la vez, diferenciarla del resto. Una “ficha de inventario” debería contener:

- Nombre de la tipología de recurso.
- Contexto de creación: ¿por qué se crea, recibe o genera y dónde?
- Definición: ¿en qué consiste?
- Contenido: instancias que incluye.
- Condiciones de validez: fecha de caducidad, nivel de normalización.
- Funciones: ¿para qué se utiliza?
- Responsable de la creación y mantenimiento.
- Actualización y conservación: cada cuánto debe actualizarse, y si hay que destruirlo o conservarlo y cómo, previsión de crecimiento, etcétera.
- Audiencia: ¿quién lo ha de usar?
- Accesibilidad: condiciones de acceso para discapacitados y necesidad de aplicaciones concretas.
- Formato/s y dinamismo: ¿en qué formato está y si es de carácter dinámico o estático?
- Aspectos legales que lo afectan.
- Ejemplo de instancia: nombre, localización y reproducción de un ejemplo.

El mapa de contenidos que acompaña al inventario consiste en una representación gráfica que refleja las relaciones detectadas entre los contenidos descritos. Debe representar la estructura existente; en ningún caso se trata de adelantar la estructura de relaciones “ideal”.

Se finaliza esta etapa de análisis con un informe de síntesis que incluye comentarios basados en el análisis comparativo de contexto, audiencia y contenidos.

Definición de la taxonomía

En este momento ya sabemos lo suficiente para poder establecer las características básicas que deberá tener la taxonomía que se elaborará. Concretamente, hay que definir:

- Alcance y dimensiones: temática, usuarios, contenidos y lenguas.
- Usos y funciones: para qué debe servir y qué tipo de aplicaciones soportará.
- Tipo de taxonomía: visible al usuario o únicamente soporte para un sistema informático.

- Grado de complejidad de categorización: número cerrado de categorías o crecimiento continuado.
- Recursos necesarios: aproximación a los costos económicos y materiales, humanos y de tiempo.

Además, deberían incorporarse en el documento de definición:

- La misión, metas y objetivos de la taxonomía.
- Una enumeración de las herramientas ya existentes que van a usarse: otras taxonomías, tesauros, sistemas de clasificación, vocabularios, cuadros de clasificación de documentos administrativos, sitios web de organizaciones similares, etcétera.

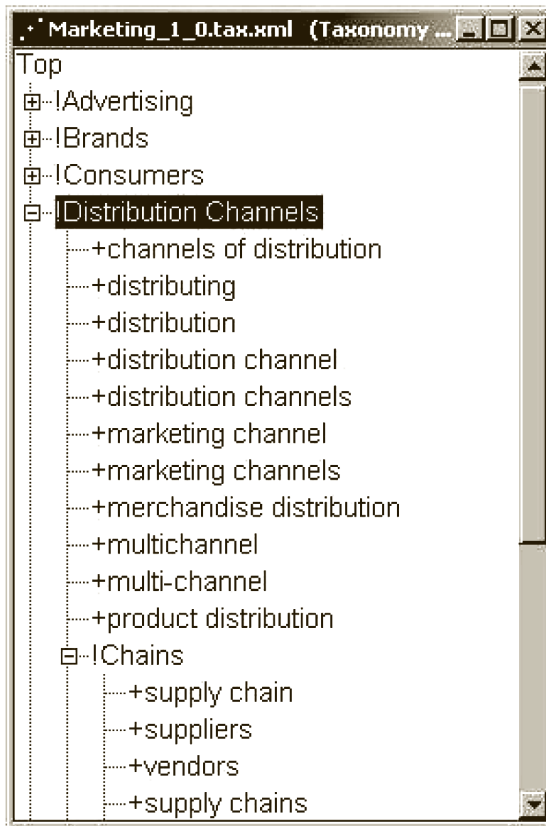
SOLUCIONES INFORMÁTICAS PARA EL DESARROLLO DE TAXONOMÍAS CORPORATIVAS

Con la definición de la taxonomía concluyen las fases fundamentalmente estratégicas sobre la elaboración de taxonomías corporativas. A continuación se desarrollan los procesos de construcción, categorización, aplicación y presentación, y el ciclo de evaluación y mejora continua. Estas fases se fundamentan en los análisis del contexto, la audiencia y los contenidos, e incorporan actividades de carácter técnico, donde la eficiencia de los sistemas informáticos pueden complementar la comprensión e inteligencia humanas. Actualmente existen numerosas soluciones informáticas para desarrollar taxonomías corporativas. Éste es un sector muy dinámico –son constantes las apariciones estelares, las alianzas y los fracasos estrepitosos–, y en franco crecimiento que según la consultora Merrill Lynch doblará su valor económico en el 2005.⁴

Además es un sector con un alto grado de diversificación producida ésta por la incidencia de diferentes factores. Sin duda el factor que ha conducido la evolución del sector ha sido el grado de intervención humana en los procesos de construcción de la taxonomía y/o la categorización. En un extremo se hallan las soluciones centradas en un algoritmo o un conjunto de algoritmos que permiten prescindir, casi totalmente, de la intervención humana. Sin embargo, sus resultados no han sido muy alentadores: Ian Hersey, vicepresidente de desarrollo y estrategia corporativa de la empresa Inxight indica que las mejores herramientas automáticas de categorización pueden obtener de un

50% a un 80% de fiabilidad, lo cual no puede considerarse aceptable.⁵ El mercado ha introducido dos alternativas:

- Taxonomías preelaboradas, especializadas en diferentes sectores industriales, que las corporaciones pueden adquirir e implementar en sus sistemas de gestión de contenidos, adaptándolas a las condiciones específicas de su contexto, audiencia y contenidos. Por ejemplo, SemioTaxonomy (Entrieva).⁶



Presentación de la taxonomía de Marketing de SemioTaxonomy
(Fuente: SemioTaxonomy: jumpstart your way to value
<http://www.entrieva.com/entrieva/downloads/SCTSds051402.pdf>)

⁵ Citado en Letson (2001).

⁶ Más información en:
<http://www.entrieva.com/entrieva/products/scts.asp?Hdr=scts>.

Subject matter area	
Automotive	High tech
Chemicals	Human resources
CRM	Human rights
DOE	Information technology
Drug/Disease	Legal
Energy	Marketing
Engineering	Medical/Health (CRISP)
European Union vocabulary	MEDRA
Federal legislation	Military
Financial	Military (DTIC)
Genes	Petroleum
Geography	Pharmaceutical
Global law	Physics
Healthcare	

Relación de taxonomías que ofrece SemioTaxonomy
 (Fuente: SemioTaxonomy: jumpstart your way to value
<http://www.entrieva.com/entrieva/downloads/SCTSDs051402.pdf>)

- Soluciones semiautomáticas en que se armonizan las capacidades de los algoritmos y la inteligencia humana. En este segmento pueden incluirse sistemas de gestión de tesauros adaptados al entorno corporativo. Por ejemplo, Wordmap Taxonomy Management System (Wordmap).⁷



Muestra del módulo de creación y gestión de taxonomías de Wordmap Taxonomy Management System (Fuente: Wordmap applications: understanding how Wordmap is applied in information systems
<http://www.wordmap.com/downloads/whitepapers/wordmap-applications.pdf>).

7 Más información en:
<http://www.wordmap.com/product/asp.html>.

<p>you are in: science > astronomy > galaxies</p> <p>words mapped under: galaxies</p> <p>active galaxies black holes elliptical galaxies irregular galaxies lenticular galaxies spiral galaxies</p> <p>other words mapped under: astronomy</p> <p>Solar System amateur astronomy cosmology eclipses extrasolar planets extraterrestrial life observatories planetariums space</p>	<p>wordmap taxonomies</p> <p>Wordmap is the leading enterprise strength taxonomy organisations can build accurate taxonomies quick variety of content applications. For more information</p> <p>search terms</p> <p>Wordmap is searching for :</p> <input type="text" value="galaxies astronomy"/> Add your own terms and press 'search' to narrow <p>results</p> <p>Astronomy Plus CCD Images of Galaxies A list of images and text from the University of Oregon More Low Surface Brightness Galaxies Clusters of Example Digital Spectra ... Different Kinds of Stars A: URL: http://zebu.uoregon.edu/galaxy.html found by: All the Web</p> <p>Astronomy 162: Stars, Galaxies, and Cosmology Entire Astrophysics course online. Very extensive in of astronomy we were concerned primarily with ...</p>
--	---

Muestra del módulo de edición de taxonomías para la navegación y búsqueda de Wordmap Taxonomy Management System (Fuente: Wordmap applications: understanding how Wordmap is applied in information systems <http://www.wordmap.com/downloads/whitepapers/wordmap-applications.pdf>).

Otro factor relevante de diversificación del mercado es la configuración del producto en relación a los tres procesos básicos de creación de taxonomías: la construcción de la estructura, la categorización de documentos y la presentación de la taxonomía. Es habitual la especialización en uno solo de estos procesos, la categorización, aislándolo de los otros dos. Ejemplos de este tipo de soluciones son el módulo Collect del programa Quantum (Entopia, Inc.),⁸ Moho Classifier v2.3 (KOFAX),⁹ y Textology Categorizer (Textology, Inc.).¹⁰ No obstante, conforme se ha ido imponiendo el modelo semiautomático y se han incorporado los sistemas de gestión de tesauros, la integración de procesos ha ido ganando terreno. Por ejemplo, el producto IDOL Classification Server (Autonomy, Inc.)¹¹ dispone de tres módulos:

8 Más información en: http://www.entopia.com/products_pg3.1.1.htm.

9 Más información en: <http://www.kofax.com/products/mohomine>.

10 Más información en: <http://www.textology.com/categorizer.html>.

11 Más información en: <http://www.autonomy.com/Content/Products/IDOL/f/Classification>.

Autonomy Taxonomy Generation para la construcción de la estructura; Automatic Categorization para la construcción de la estructura y la categorización; y Automatic Clustering para la agrupación de documentos (*clustering*) y la visualización de las agrupaciones resultantes. Tampoco son extrañas las integraciones en plataformas más amplias situadas en el ámbito de la gestión de contenidos; es el caso de Collaborative Taxonomy Engine y Custom Directory (Wherewithal), que constituyen un componente de la plataforma Enterprise Knowledge Portal 3.0 para el desarrollo de portales corporativos.

Otros factores de diversificación de este mercado son: el carácter monousuario o multiusuario de la aplicación; la orientación al desarrollo centralizado o descentralizado (colaborativo) de la taxonomía; la tipología de información que permiten procesar (estructurada o desestructurada, textual o no textual, monolingüe o plurilingüe...); etcétera.

Si se plantea la adquisición e implementación de una aplicación para el desarrollo de la taxonomía corporativa, deben tenerse en cuenta todos estos factores de diversificación y otros criterios generales de evaluación, como son la solvencia y experiencia del productor, la fiabilidad y escalabilidad de los procesos, las posibilidades de integración con otras aplicaciones de desarrollo de portales y de gestión de contenidos y, por supuesto, el precio. Una alternativa, inicialmente más económica, es el desarrollo de una aplicación *ad hoc*, aunque ya son conocidos los costos a medio y largo plazo de este tipo de soluciones.

CONSTRUCCIÓN DE LA TAXONOMÍA CORPORATIVA

La fase de planificación estratégica proporciona los límites de la realidad que será objeto de representación mediante la taxonomía corporativa; o dicho en términos técnicos, un dominio. En esta fase, deben identificarse los conceptos que integran el dominio, la designación o designaciones posibles de cada concepto,¹² y las relaciones que se establecen entre ellos.

Los aspectos fundamentales de esta etapa del proceso están contemplados en las normas para el establecimiento y desarrollo de tesauros monolingües. A partir de los textos normativos indicados y de los trabajos de revisión ya comentados, es posible diferenciar cuatro procesos necesarios para la construcción de la taxonomía: (1) la identificación de criterios de facetación de los contenidos; (2) la extracción del léxico; (3) el control del léxico; y (4) el desarrollo de la estructura de la taxonomía.

12 El conjunto de designaciones se denomina, técnicamente, el léxico del dominio. Cada designación se denomina término o, más comúnmente en el contexto de los sitios web, categoría.

Identificación de criterios de facitación de los contenidos del sitio web

Entendemos por facetas los aspectos, propiedades o características de una realidad específica claramente definidos, mutuamente exclusivos y, en conjunto, exhaustivos. Cada una de las facetas se descompone en categorías de diferentes niveles de especificidad; por ejemplo, en el caso del vino, una faceta podría ser el color, y ésta podría descomponerse en las categorías: blanco, tinto, y rosado. Para identificar correctamente las facetas debe tenerse en cuenta el resultado del análisis del contexto, de la audiencia y de los contenidos. Posteriormente, en la fase de presentación de la taxonomía, debe decidirse si se prioriza un número limitado de facetas y el orden en que éstas se presentan.¹³

Extracción del léxico

El objetivo de esta fase es identificar todos los términos o categorías que designan los conceptos propios del dominio. Cada categoría debe vincularse, como mínimo, a una de las facetas establecidas en la fase anterior. Para la extracción deben utilizarse diferentes tipos de fuentes : personales, documentales, etcétera, e incluso taxonomías preexistentes. Para cada tipo de fuentes, es posible localizar recursos propios (internos) de la organización, y recursos externos.

Es necesario registrar todas las designaciones posibles de un mismo concepto y todas las formas que puede adoptar una designación. Además, es conveniente consignar algunos detalles de cada designación: fuentes en las que aparecen, frecuencia de aparición, coocurrencia con otras designaciones, etcétera.

Control del léxico

Es posible que en la fase anterior se hayan identificado diferentes designaciones posibles para un mismo concepto (sinónimos o cuasisinónimos) y/o que una misma designación presente diferentes formas posibles (gramaticales, ortográficas, etcétera). La eficacia de una taxonomía corporativa recomienda el establecimiento de un término preferente para representar cada uno de los conceptos que integran su dominio. El control del léxico tiene por objetivo determinar cuáles son esas categorías preferentes, teniendo en cuenta que, debido al carácter personalizado de la taxonomía, deben preferirse las alternativas registradas en fuentes internas, a las registradas en fuentes externas. En este punto se procede a establecer relaciones de equivalencia entre las designaciones preferentes

13 Así, por ejemplo, en el portal Wine.com (<http://wine.com>), las facetas seleccionadas son: tipo (*type*), región (*region*) y bodega (*winery*).

y las no preferentes, así como a restarle ambigüedad a los términos polisémicos aceptados como categorías preferentes.

Desarrollo de la estructura de la taxonomía

El objetivo de esta fase es identificar y establecer dos tipos de relaciones entre las categorías de la taxonomía: la relación jerárquica y la relación asociativa. El resultado será un vocabulario controlado en forma de estructura conceptual.

La relación jerárquica se basa en grados o niveles de orden superior y subordinación, en que un término general representa un todo o clase y los términos subordinados corresponden a sus miembros, partes o instancias. Todas las categorías de una faceta deben estar conectadas por relaciones jerárquicas.

La relación asociativa conecta categorías (de diferentes facetas) a partir de asociaciones de ideas que se pueden dar entre: una acción y su resultado o producto; un concepto y una propiedad del mismo; un producto y el material con que está hecho; etcétera. La conexión a través de relaciones asociativas se establecerá únicamente cuando pueda revelar términos alternativos potencialmente útiles para el usuario al momento de recuperar información.

Se recomienda la participación de la audiencia al establecer las relaciones jerárquica y asociativa mediante la aplicación de técnicas como la ordenación de fichas (*card sorting*).¹⁴

Es importante que todos los criterios establecidos durante el proceso sean recogidos en un manual de directrices y concreciones, que permitirá el mantenimiento y la evolución de la taxonomía de forma consistente.

CATEGORIZACIÓN

La categorización es el proceso de asignación de una o más categorías a un documento. Los diferentes métodos existentes pueden contemplarse como un continuo: en un extremo se encuentran los métodos intelectuales (o manuales), mediante los que un agente humano analiza un documento y le asigna las categorías que lo representan; en el otro, se encuentran los métodos automáticos en los que una aplicación informática agrupa documentos (*clustering*) a partir de criterios preestablecidos de identificación de similitudes, etiqueta las agrupaciones (*clusters*) mediante categorías extraídas de los propios documentos, y organiza estas categorías de forma jerárquica, alfabética,

14 Información completa sobre esta técnica en: Maurer, Donna; Warfel, Todd. Card sorting: a definitive guide. Consultado en: 26-4-2004.
http://www.boxesandarrows.com/archives/card_sorting_a_definitive_guide.php.

etcétera. Los métodos automáticos se basan en algoritmos estadísticos y/o procesamiento del lenguaje natural, y su aplicación implica la integración del desarrollo de la taxonomía y la categorización en un único proceso.

Entre los dos extremos se encuentran los métodos semiautomáticos, que combinan la inteligencia humana y la eficiencia de las máquinas. Las dos variantes más utilizadas son la categorización basada en reglas de búsqueda y la categorización basada en conjuntos de documentos ejemplares o de entrenamiento.

En el primer modelo el administrador del sistema elabora una ecuación de búsqueda “magistral”, que se vincula a una o más categorías de la taxonomía y que especifica las condiciones que debe cumplir un documento para que se le asigne aquella categoría o categorías determinadas.

En el modelo de entrenamiento, el administrador selecciona los documentos que, de forma ejemplar, corresponderían a una categoría determinada. Estos documentos son analizados mediante algoritmos y darán como resultado un perfil de comportamiento. Se vincularán automáticamente a una categoría determinada todos aquellos documentos que, analizados mediante el mismo algoritmo, presenten el mismo perfil de comportamiento que los documentos ejemplares.

Los sistemas semiautomáticos proporcionan resultados más relevantes que los automáticos, y se muestran más eficientes que los intelectuales.

APLICACIÓN Y PRESENTACIÓN DE LA TAXONOMÍA

En la definición de la taxonomía, con lo que concluye la etapa de planificación, se han determinado las aplicaciones que la estructura deberá soportar (sistemas de organización y exploración o *browsing*; sistemas de recuperación; sistemas de filtraje; ...) Cuanto más numerosas sean las aplicaciones mayor será la rentabilidad de la inversión realizada.

El desarrollo de la taxonomía ha dado como resultado un vocabulario controlado en el que las categorías se relacionan mediante tres tipos de relaciones: de equivalencia, de jerarquía y de asociación. Esta construcción conceptual no tiene que ser necesariamente trasladada tal cual al sitio web para su presentación al usuario. En el caso de un sistema de exploración o navegación, por ejemplo, puede resultar eficaz una presentación doble de forma paralela, en las formas de estructura jerárquica de categorías y de relación alfabética. En el caso de una aplicación de recuperación directa (buscador), el uso de la taxonomía puede limitarse a la relación de equivalencia a modo de anillo de sinónimos para ampliar el alcance de las consultas planteadas por los usuarios.

También se puede optar por presentaciones textuales, gráficas o metafóricas, por organizaciones cronológicas o geográficas, unidimensionales o arbóreas, etcétera. Lo cual dependerá de las preferencias y habilidades detectadas en el análisis de la audiencia. Esta vocación de adaptación al usuario implica que el ciclo evaluativo de la taxonomía debe iniciarse ya en esta fase, mediante la realización, por ejemplo, de tests de usabilidad sobre las diversas presentaciones posibles.

CICLO DE EVALUACIÓN Y MEJORA CONTINUA

Debido al dinamismo del trinomio básico –contexto, audiencia y contenidos–, una taxonomía debe ser una herramienta en continuo cambio.

Es necesario establecer un sistema de toma de decisiones sobre los cambios (quién los propone y quién los acepta) y un sistema de comunicación de las decisiones sobre éstos (quién, dice cómo y a quién le o les dice). El mantenimiento o gestión de cambios no es el único mecanismo para la mejora continua de la herramienta; es necesario establecer un sistema de evaluación. El ciclo evaluativo debe iniciarse antes del lanzamiento real de la taxonomía corporativa, durante la fase de aplicación y presentación. Además, debe mantenerse una vez que la taxonomía esté ya en operación mediante mecanismos indirectos –test de relevancia (*relevancy testing*), reutilización de ítems (*item reuse*), etcétera,– y directos –tests de usabilidad, encuestas de satisfacción de usuarios, etcétera. La selección de los mecanismos dependerá, entre otras consideraciones, de las aplicaciones que exploten la taxonomía.

El ciclo evaluativo debe incorporar el análisis de costos y beneficios que comporta para la organización el diseño e implementación de la taxonomía. Sin duda una taxonomía bien construida puede mejorar la productividad de sus usuarios, sean estos miembros de una organización o usuarios externos. Aunque la cuantificación exhaustiva de los beneficios en términos de ROI (retorno sobre la inversión) es complicada, se están trabajando propuestas para desarrollar valoraciones parciales igualmente efectivas (Berlind 2004).

Finalmente, no debe olvidarse la necesidad de difundir la taxonomía corporativa entre todas las instancias que intervienen en su desarrollo y mantenimiento, y entre los usuarios que deben beneficiarse de su implementación.

CONCLUSIONES

La taxonomía corporativa es una herramienta estratégica para las organizaciones. En su diseño pueden aplicarse las técnicas tradicionales de construcción

de vocabularios controlados en general y, más concretamente, de tesauros. No obstante, deben tenerse en cuenta una serie de cuestiones específicas fundamentales.

- El diseño de una taxonomía corporativa debe plantearse como un proyecto riguroso, con etapas definidas, recursos asignados que hay que administrar, un calendario y plazos que hay que respetar, etcétera. La improvisación sólo se traducirá en una pérdida de tiempo y dinero, tanto para el equipo de trabajo como para la organización en su conjunto.
- La planificación de una taxonomía debe basarse en un análisis previo de tres elementos clave de la organización en la que se aplicará: el contexto, la audiencia y los contenidos. Este trinomio básico constituye, además, un marco de referencia para desarrollar el resto de las etapas del proyecto de taxonomía.
- En el desarrollo del proyecto será necesario lograr la participación de todos los agentes de la organización, en especial de los órganos decisivos, de la audiencia y de los productores de contenidos. Sin esta implicación resultará difícil, si no imposible, lograr el resultado deseado.

BIBLIOGRAFÍA

- Berlind, David. Taxonomy today, ROI tomorrow. Consultado en: 23-4-2004.
http://techupdate.zdnet.com/techupdate/stories/main/Taxonomy_today,_ROI_tomorrow.html.
- Bryar, J. V. Taxonomies: the value of organized business knowledge. Consultado en: 16-4-2004.
http://www.thomson.com/cms/assets/pdfs/biz_info_and_news/whitepaper_newsedge_taxonomies.pdf.
- Conway, Susan; Sligar, Char. "Building taxonomies", en *Unlocking knowledge assets*. [S.l.]: Microsoft Press, 2002, pp. 105-124. Capítulo también disponible a través de Internet:
<http://www.microsoft.com/mspress/books/sampchap/5516.asp#SampleChapter>.
- Edols, Liz. Taxonomies are what? Consultado en: 13-4-2004.
<http://www.freepint.com/issues/041001.htm#feature>.
- Gilchrist, Alan. "Thesauri, taxonomies and ontologies: an etymological note", en *Journal of documentation*, 2003, v. 59, n. 1, pp. 7-18.
- Gilchrist, Alan; Kibby, Peter; Mahon, Barry. *Taxonomies for business: access and connectivity in a wired world*. London: TFPL, 2000. ISBN 1-870-889-83-5.

- Letson, Russell. Taxonomies put content in context. Consultado en: 20-1-2004.
http://www.transformmag.com/db_area/archs/2001/12/tfm0112f1.shtml.
- National Information Standards Organization. Developing the next generation of standards for controlled vocabularies and thesauri. Consultado en: 23-3-2004.
<http://www.niso.org/committees/MT-info.html>.
- Pack, Thomas. Taxonomy's role in content management. Consultado en: 23-4-2004.
<http://www.econtentmag.com/Articles/ArticleReader.aspx?ArticleID=867>.
- Rosenfeld, Louis; Morville, Peter. *Information architecture for the world wide web*. 2nd ed. Beijing [etc.]: O'Reilly & Associates, 2002. ISBN 0-596-00035-9.
- Taxonomy and content classification: market milestone report. A Delphi Group white paper. Consultado en: 13-4-2004.
<http://www.dmreview.com/whitepaper/WID371.pdf>.

