

Capítulo 3

LA REPRESENTACIÓN Y RECUPERACIÓN DE LOS CONTENIDOS DIGITALES: DE LOS TESAuros CONCEPTUALES A LAS FOLKSONOMÍAS

José Antonio Moreiro González

Introducción

Dentro de las exigencias planteadas por el nuevo y cada vez más común acceso a la información en documentos digitales, especialmente de los ofrecidos por la Web, se encuentra la revisión de los modos en que se efectuaba la representación de los conceptos principales mediante lenguajes combinatorios. En consecuencia los tesauros, aún manteniendo vigentes su funcionalidad y principios de actuación, han debido evolucionar hacia soluciones más abiertas. Cuando analizamos los documentos intentamos entender las categorías semánticas que el texto contiene, con objeto de identificar sus conceptos principales. Hay que remontarse aquí hasta las propuestas de construcción retórica de los discursos en torno a las ideas más generales a las que alude la expresión del pensamiento, los diez predicamentos de Aristóteles,¹ que luego Leibniz representó en tan solo cinco categorías,² y a los que Kant los limitó a cuatro.³ Estos universales se intensionalizan en las macroestructuras, cuya conformación debe entenderse como una abstracción de las propiedades genéricas, precisamente los focos de atención cuando deseamos hacer recuperaciones o representaciones relevantes y pertinentes.

¹ Substancia, cantidad, cualidad, relación, lugar, tiempo, situación, posesión, acción y pasión, Aristóteles: *Tratados de Lógica (El Organon)*, Madrid: Gredos, 1982, vol. 1, p. 51.

² Substancia, cantidad, cualidad, relación y acción o pasión, Leibniz: *De Synthesi et Analysisi universali seu Arte inveniendi et judicandi*, 1966, incluido en su *Hauptschriften zur Grundlegung der Philosophie*, Hamburgo: Meiner, vol. 1, p. 41.

³ Cantidad, cualidad, relación y modalidad, Kant: *Crítica de la razón pura*, 2 vols., Barcelona, Ediciones Folio, 2000, vol. 1: Crítica analítica.

3.1. Los tesauros. El origen de su red conceptual

La consideración de las propuestas retóricas nos sirve para acercarnos al concepto de *tesauro* en cuanto lista formada con las posibilidades de representación de los conceptos más generales (macroestructurales) en un dominio concreto. En su fase de construcción, los documentos contienen ya muchos tesauros, por las relaciones semánticas que existen entre sus términos componentes. Un mismo concepto puede ser genérico (*genus*), por su relación con las ideas a las que se extiende, y especie (*species*), por su subordinación a otra más general, lo que también sucede en un tesauro:⁴ *Genus*, con un género supremo: *top term* o macrodescriptor, y con géneros y especies intermediarios: *middle term* o submacrodescriptor. Y *species* de diferente nivel en un tesauro, con una especie especialísima: los descriptores genéricos, y con individuos o descriptores específicos.

De manera que, dentro de cada campo semántico, los descriptores mantienen las relaciones interconceptuales propuestas por Aristóteles, en las que está la causa del orden jerárquico (facetas en que se clasifica y desarrollo de estas): género, subtipo y diferencia. Mientras que se deben al movimiento de Port-Royal los conceptos lógicos, asimismo integrados en la estructura de los tesauros, de:⁵

- *Definición*, para explicar la *esencia* de las cosas. En los tesauros, la *scope note* concreta el significado de algunos descriptores.
- *División* que clarificaba una definición por la diferencia de una especie con otras (asociaciones entre términos de la misma categoría).

Finalmente, Raimundo Lulio, en su *Logica nova*, propuso las diez reglas generales de cuestionar o decámetro hermenéutico que subyace a cualquier hecho comunicativo.⁶

⁴ La estructuración del tesauro y las relaciones que se pueden establecer entre sus elementos puede consultarse en G. van Slype: *Los lenguajes de indización. Concepción, construcción y utilización en los sistemas documentales*, Madrid y Salamanca: Fundación Germán Sánchez Ruipérez, 1991.

⁵ Situada por Nidditch en el paso de la lógica clásica a la lógica simbólica, en su obra: *El desarrollo de la lógica matemática*, Madrid: Cátedra, 1978.

⁶ Jordi Gayà: «El arranque filosófico del Ars Iuliana. Constantes y fragmentos del pensamiento Iuliano», en F. Domínguez y J. de Salas (eds.): *Actas del simposio sobre Ramon Llull en Trujillo*, Tübingen: Max Niemeyer, 1994, pp. 1-8. En ellas están ya las razones de asociar términos como estrategia empleada también para relacionar los descriptores en los tesauros: *Utrum; Quid; De quo; Quare; Quomodo; Ubi; Quando; Quantum; Cum quo; Quale*, que además de ser el fundamento lejano de las propuestas lógicas de Port-Royal marcan buena parte de las relaciones asociativas existentes entre los conceptos que un tesauro define.

Los tesauros, que representan el pensamiento, aprovechan precisamente las relaciones semánticas para establecer asociaciones entre los términos que componen el corpus, sobre las que realizan su proceso inductivo que permite a los usuarios navegar coherentemente por los conceptos que representan:

- Integrando términos dentro de una misma categoría.
- Determinando diferencias de una especie con otras.
- Posibilitando la división o análisis del género por las diferencias.
- Concediendo definiciones (*scope note*) para ajustar el significado de un término.

Al construir un tesauro manejamos tres herramientas:

- Una colección de términos extraídos del dominio cuya representación se intenta.
- Una organización macroestructural de la que se siguen los esquemas y la segmentación de aquella colección. Se dispone de una macroestructura global (el dominio de aplicación del tesauro), de unas macroestructuras secundarias (representada por los macrodescriptores que encabezan cada una de las subdivisiones del tesauro) desde las que se clasifican.
- Y unas microestructuras o estructuras de superficie, que se componen por los propios términos y sus relaciones.

Estos principios y aplicaciones sobre las que gira la naturaleza de los lenguajes controlados siguen siendo actuales y, por lo tanto, válidos. Es en el cómo donde se evidencia la adaptación a las nuevas necesidades

Si regresamos al esquema general del acto de comunicación lingüística, vemos que se origina en los *septem loci* retóricos⁷ para la construcción (y representación) de los textos, cuestiones cuyas posibles respuestas llevarían, como muestras, a los predicados universales entendidos como términos constitutivos básicos o no-

⁷ En E. H. Lausberg: *Elementos de retórica literaria*, Madrid: Gredos., 1975, pp. 30-46, se hace evidente cómo la gramática generativa reconsideró el alcance del hexámetro universal hermenéutico, los *septem loci* retóricos, para la construcción (y representación) de los textos. Carnap los denominó «palabras universales» en su obra: *Logical Structure of the World & Pseudoproblems in Philosophy*, Berkeley: University of California Press, 1969, p. 292. Cada pregunta es un tipo semántico y contiene todas las posibles respuestas «ejemplares». *Quis, Quid, Quur (Cur), Quomodo, Ubi, Quando, Quibus*.

ciones puras, mediante los cuales se simbolizan la macroestructura global y las secundarias,⁸ que luego se hacen patentes mediante los descriptores en la indización y mediante las oraciones explicativas en el resumen. Su interpretación más reciente se da con las *wh questions*, inventario temático dispuesto para indagar selectivamente en los mensajes del texto.

La respuesta a estas preguntas determina el reconocimiento de la materia atendida en el documento y las principales condiciones que cumple. Su esquematización se conoce como *Paradigma de Laswell*⁹ (*Loi de Quintilien*, para los tratadistas franceses, y antes propuestos ya por Aristóteles en la *Retórica*¹⁰ y Cicerón en *De oratore*¹¹), modelización de los movimientos necesarios que deben cumplir las etapas de análisis de la información en torno al qué de las materias y asuntos; al por qué de sus causas y motivos; al cómo o procedimiento para realizar algo; a quién protagoniza el contenido del documento; a cuándo se ejecuta la acción; al dónde se desarrolla la acción, e incluso al con quién.¹²

Estas preguntas actúan como plantilla constructora y destructora del discurso, viniendo las respuestas pertinentes de la terminología que compone un tesoro.

3.1.1. DIFICULTADES SOBREVENIDAS AL EMPLEO DE LOS TESAUROS

Aunque el empleo de los tesauros es fiable y notorio, la ambigüedad, riqueza y capacidad de innovación constante de los lenguajes en los que se encuentran expresados los documentos sobre los que trabaja un tesoro, produce en muchas

⁸ Hatier: *Le compte rendu de lecture. De l'analyse a la créativité*, París: Hatier, 1977, p. 59.

⁹ H. J. Lasswell: «Estructura y función de la comunicación en la Sociedad», en M. de Moragas (ed.): *Sociología de la Comunicación II. Estructura, funciones y efectos*, 4.ª ed., Barcelona: Gustavo Gili, 1994. Podemos ver a las *wh questions* como un molde categorial que nuestro fin analítico utiliza en sentido deconstructivo: *who, when, where, what* y sus subordinadas *how, why* y *what for*, aunque estos interrogantes se hayan ideado primeramente para proporcionar los criterios a seguir en una adecuada redacción informativa.

¹⁰ Aristóteles: *Retórica*, Madrid: Gredos, 1983. Comentado también por T. Albadalejo: *Retórica*, Madrid: Síntesis, 1989.

¹¹ M. Tulli *Ciceronis rhetorica*. 1963. Recognovit brevique adnotatione critica instruxit A. S. Wilkins. Oxonii: Typ. Clarendonianus. Contiene: I. *Tres libros De Oratore*. II. *Brutus orator. De optimo genere oratorum. Partitiones oratoria. Topica*.

¹² La gramática de casos de Fillmore, Ch. J. The case for case, en E. Bach y R. T. Harms (comp.): *Universals in Linguistic Theory*, Nueva York: Holt, Rinehart y Winston, 1968, pp. 1-88, determina los niveles de: *Acción, Agente, Objeto, Instrumento, Modo, Causa, Lugar y Tiempo*, incluyendo los modificadores *asociativo* y de *situación*.

ocasiones el silencio o el ruido, en definitiva, una falta de pertinencia a la hora de la selección de documentos. A lo que se añade el problema del número de documentos que circulan por la Web y la variedad de sus soportes.

Las estructuras que se encuentran definidas en un tesoro hacen que solo puedan recuperar los documentos que posean aquellas relaciones estáticas con las que trabaja y que han sido definidas a priori. Los problemas concretos se traducen, en la práctica, en dos graves inconvenientes:¹³

- 1) Dificultad para describir campos con un nivel de abstracción alto (el de la informática, por ejemplo), o en aquellos en los que, por el contrario, los documentos no estuvieran estructurados en ninguna medida (textos de libre formato).
- 2) Un elevado coste tanto en la creación como en su mantenimiento, que solo se podría eliminar con la automatización o semiautomatización de ambas tareas. Lamentablemente esta automatización difícilmente puede ser llevada a cabo de forma eficaz y eficiente, debido a la riqueza de las estructuras semánticas de los textos.

Históricamente, se han empleado tres soluciones para sortear estos inconvenientes:

- *Lingüística*: consistente en complementar el esquema clásico con nuevos aspectos que ampliaban el tesoro mediante la creación de distintas «vistas» (facetas) del mismo. Esta aproximación facetada complicaba la automatización de los modelos, y los hacía poco intuitivos y manejables para el usuario. Otra de las aproximaciones consiste en la atomización sémica.
- *Estadística*: solución matemática y, por ello, favorable a la automatización. Sin embargo se alejaba de la realidad lingüística la utilización de algoritmos que pretendían una visión simplista del mundo real. Por ello, solo son eficaces en ámbitos muy concretos. Se han empleado algoritmos *k-means*, coocurrencia terminológica o redes neuronales (ART, Kohonen, etc.).
- *Informática*: en los años noventa se empezaron a aplicar los modelos propios de la ingeniería de software a estructuras como el tesoro.

¹³ H. Chen, T. Yim, D. Fye y B. Schatz: «Automatic Thesaurus Generation for an Electronic Community System», *Journal of the American Society for Information Science*, 46, n.º 3, 1995, pp. 348-369.

El proceso de identificación y adquisición de componentes representativos suele determinarse en estos pasos:¹⁴

- Limpieza de los caracteres no procesables.
- Análisis léxico textual:
 - Procesos de filtrado para eliminar previamente los términos vacíos.
 - Tratamiento de los términos flexionados mediante algoritmos de *stemming* (número de *n-grams* grupos de letras).
 - Tratamiento de palabras compuestas.
- Filtrado de términos: discriminación de términos representativos de un texto: algoritmos IDF y *n-grams*.
- Luego, se procede a la obtención e integración de relaciones:
 - Obtención de asociaciones temáticas entre componentes (clusterización en clases de aquellos descriptores con características comunes): clasificadores cienciométricos:¹⁵ Co-wording; estadísticos: Max-min, K-vecinos, K-vecinos incremental, Isodata; neuronales: Kohonen, Art-1, Art-2.
 - Construcción de la representación del dominio: clusterización + extracción de raíces: cálculo del centroide (descriptor más general del cluster).
 - Proceso de generación de relaciones semánticas para evitar relaciones distintas para dos descriptores dados.

La convergencia de estos métodos ha desembocado en una evolución de los estándares sobre tesauros hacia verdaderos mapas de representación del conocimiento.

Esta transformación lleva implícita la creación de relaciones adaptables a cada dominio, así como la inclusión de gran variedad de descriptores cuya categoría gramatical aporta matices en la semántica del mapa conceptual y de nuevas categorías relacionales que ligán los recursos de información y las facetas. Precisamente en los años noventa Internet y su oferta de enlace hipertextual de documen-

¹⁴ Véase el apartado «Indización automática» de José Antonio Moreiro González: *El contenido de los documentos textuales: su análisis y representación mediante el lenguaje natural*, Gijón: TREA, 2004, donde se describe detalladamente el proceso y la intervención de las técnicas de discriminación algorítmicas y de los analizadores lingüísticos.

¹⁵ C. T. Meadow, B. R. Boyce y D. H. Kraft: *Text Information Retrieval Systems*, 2.ª ed., San Diego: Academic Press, 1999.

tos obligó a diferenciar la representación de los contenidos documentales, surgiendo alternativas a los lenguajes combinatorios:

- Tesauros conceptuales.
- Tesauros de verbos.
- Mapas de conceptos.
- *Topic maps*.
- Folksonomías.

Aparece como nota común de estas propuestas su pretensión de activar las correspondencias terminológicas de los tesauros con la idea de mejorar la representación y aumentar la precisión y eficacia mediante nuevas relaciones, mayores en número y vinculadas mediante identificación para favorecer la navegación, visualización o grafos explícitos de conexiones de la información.¹⁶

Desde que comenzó a intentarse la automatización de los tesauros se apreció un incremento pragmático de las relaciones,¹⁷ en especial las de asociación, que son las que más subtipos presentan:¹⁸

1. Ideas combinadas.
2. Términos relacionados conceptualmente.
3. Contigüidad.
4. Relaciones asociativas por definición.
5. Relaciones asociativas con diferente jerarquía.
6. Relaciones asociativas trasladadas por significado.
7. Relaciones asociativas con idéntica jerarquía.
8. Cuestiones de finalidad.
9. Relaciones asociativas sin especificar.

Como cada uno de estos subtipos ofrece subdivisiones, se ha multiplicado de forma notable el número posible de relaciones asociativas presentes en un tesauro.

¹⁶ W. Schmitz-Esser: «New approaches in Thesaurus Application», *International Classification*, 18, 3, 1991, pp. 143-147.

¹⁷ H. Alani, C. Jones y D. Tudhope: «Associative and spatial relationships in thesaurus-based retrieval», *Proceedings of the 4th European Conference on Research and Advanced Technology for Digital Libraries (ECDL2000)*, *Lecture Notes in Computer Science*, Springer, 2000, pp. 45-58.

¹⁸ D. Tudhope, H. Alani y C. Jones: «Augmenting thesaurus relationships: possibilities for retrieval», *Journal of Digital Information*, 1, n.º 8, 2001, pp. 28-38.

3.2. Los tesauros conceptuales

Se entiende por tesoro conceptual una red semántica en la que cada nodo contiene un único concepto semántico que puede tener una serie de descriptores asociados, que también pueden ser identificados en la red de descriptores relacionados según las típicas relaciones de los tesauros: preferenciales, jerárquicas o asociativas. La denominación se fundamenta en la noción de materia (concepto) de la que tratan los textos:

- Un modelo de tesoro de *relaciones asociativas*, que aúna términos y conceptos reales por su similitud de sentido en el contexto específico del usuario.¹⁹
- Una *red semántica conceptual* donde los términos más genéricos de una faceta determinante encaminan hacia los más específicos, y viceversa (navegación vertical), que a su vez permite la transición de una clase hacia otra y de un campo de la ciencia hacia otro mediante las relaciones asociativas (navegación horizontal usando nudos polijerárquicos), e incluso la navegación transversal de la red desde una intersección de árboles.²⁰
- *Espacio conceptual* que concibe el tesoro como un sistema formal en torno a la idea de dominio algebraico. Su modelo *espacial* define las relaciones entre términos con mayor precisión que los tesauros tradicionales.²¹

Respecto a los tesauros convencionales presentaban estas características novedosas:

- a) Listan todas las palabras «no vacías» existentes en las bases de datos;
- b) consideran los términos coloquiales, también las variaciones y truncamientos de los términos reconocidos (los *topic maps* permitirán acceder a un concepto por todos sus sinónimos, incluso en siglas y códigos);
- c) aportan notas definitorias que aclaran las posibles dudas de uso, llegando incluso a dar algunas explicaciones extensas.

¹⁹ Jacques Maniez: «Du bon usage des facettes: des classifications aux thésaurus», *Documentaliste-Sciences de l'information*, 36, n.º 4-5, 1999, pp. 249-262.

²⁰ André Deweze: *Réseaux sémantiques: essai de modélisation-application à l'indexation et à la recherche documentaire*, Lyon: Université Claude-Bernard, 1981.

²¹ P. Schüuble: «Thesaurus based concept spaces», *Annual ACM Conference on Research and Development in Information Retrieval. Proceedings of the 10th annual international ACM SIGIR conference on Research and development in information retrieval*, Nueva Orleans (Luisiana), Nueva York: ACM Press, 1987, pp. 254-262.

- d) razonan las equivalencias existentes entre términos;
- e) contienen numerosas relaciones asociativas entre descriptores, e incluso con los no descriptores;
- f) diferencian los términos de utilización real de aquellos no usados.

En su funcionamiento, ante una demanda, establecen enlaces entre los conocimientos del usuario y las bases terminológicas, mezclando en la recuperación el lenguaje controlado y diccionarios del lenguaje natural, por lo que manejan:

- a) analizadores morfológicos, sintácticos y semánticos, para evitar la indeterminación de las palabras y de sus relaciones;
- b) bancos de datos terminológicos, con los conceptos resultantes del análisis de originales de un dominio;
- c) procedimientos provenientes de la Inteligencia Artificial para bases de conocimientos terminológicas, que permiten búsquedas automatizadas en diferentes bases de conocimientos multilingües, trabajando cooperativamente a través de la red Internet;
- d) además, para mejorar la precisión de las recuperaciones documentales, aprovechan el diseño de ontologías por áreas del conocimiento desde las que se autogeneran tesauros conceptuales que permiten:²²
 - distinguir los sinónimos,
 - suprimir los homónimos e
 - inducir relaciones asociativas entre los descriptores.

Una ontología para una base de conocimientos de IA debe abarcar:

- los diferentes tipos de documentos,
- las descripciones conceptuales,
- las relaciones entre dichos documentos (citas), y
- las de estos con los diferentes problemas científicos; además de índices, descripciones bibliográficas, tesauros, códigos clasificatorios, formalizaciones de validez, información terminológica, etcétera.

²² M. Uschold y M. Gruninger: «Ontologies: Principles, Methods and Applications», *The Knowledge Engineering Review*, 11, n.º 2, 1996, pp. 93-136.

Su aplicación debe proporcionar una visión general de la estructura y de la terminología del dominio que facilite recuperaciones muy relevantes. Pese a todo, los tesauros conceptuales presentan evidentes limitaciones:

- a) *Desorientación de los usuarios* durante su navegación a través de los textos o de los hipertextos para encontrar documentos relevantes o conceptos apropiados. Se causa por la sobresaturación de enlaces en los nodos que conjuntan información de numerosas fuentes.²³
- b) La normalización de la terminología de los documentos. Solución: contar con bases de conocimientos terminológicos, con gran parte del corpus lingüístico de cada una de las lenguas más usadas.
- c) Las *interfaces hombre-máquina deben ser capaces de relacionar los conceptos a partir de ontologías preestablecidas*. Se facilitará el acceso mediante representaciones gráficas de la Red (tabla de contenidos gráficos con las conexiones visibles mediante enlaces hipertextuales: mapas).

Los tesauros conceptuales y su elaboración no están muy lejos de los resultados que se obtienen en los programas de indización automatizada. Resultan similares la extracción y la normalización de vocabulario. Si después se representan los términos mediante algún tipo de relación, como las infográficas, estamos cerca de formalizar un lenguaje combinatorio:

- Los términos se extraen de documentos a 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.
- Se obtiene un corpus 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.

²³ W. Schmitz-Esser: «New Approaches in Thesaurus Application», *International Classification*, 18, n.º 3, 1991, pp. 143-147.

- Los conceptos solicitados se confrontan con la red terminológica, cuyos elementos están diseñados como mapas representativos de los textos, y que actúan así como lenguajes controlados que organizan la información de cualquier objeto contenido en la Red.

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 simples,
 - luego los sintagmas.
- Con los términos identificados se establecen bancos de datos terminológicos.
- Finalmente, mediante redes neuronales, se crean bases de conocimientos terminológicas con las que se establecen búsquedas automatizadas en Internet.

Sugieren siempre a los usuarios una ampliación de las búsquedas y suponen una mejora de los resultados de la recuperación, pues permiten a los usuarios buscar mediante una gran cantidad de términos:

- entre los suyos propios,
- los procedentes de la base de datos
- e incluso de indizaciones previas.

Además, los usuarios controlan interactivamente el proceso de navegación por el tesauro, frente a la rigidez que suponen los procedimientos automáticos.

3.3. Tesauros de verbos

Otra de las propuestas de mejora de los tesauros consiste en la inclusión de verbos que complementen los tesauros estáticos tradicionales de sustantivos.²⁴

²⁴ B. Levin: *English verb classes and Alternations: a preliminary investigation*, Chicago: University of Chicago Press, 1993.

Está causado por la necesidad de mejorar la precisión cuando se trata de reutilización de software o de recuperación de imágenes en movimiento. Las ventajas que ofrece el uso de descriptores verbales:

- Posibilidad de indizar imágenes de vídeo mediante gerundios.
- Identificación mediante verbos de asociaciones funcionales mucho más adaptables a dominios concretos.
- Posibilidad de mostrar la relación existente entre dos conceptos usando las posibilidades innumerables del lenguaje natural (categorías verbales a modo de relaciones facetables).
- Desambiguación conceptual.

La inclusión de verbos en labores de recuperación plantea posibilidades semánticas diferentes a las de los tesauros tradicionales:²⁵

- Se pueden mostrar los agentes que interactuarán con el sistema y cómo lo hacen (casos de uso).
- También permite mostrar los diagramas de secuencia y actividad.

La tipología de relaciones de asociación del lenguaje UML se integra así en la tendencia de ampliar el número de relaciones de asociación de los tesauros y de que no planteen ninguna ambigüedad:

- Relación de agregación, en la que la desaparición del todo no implica la desaparición de las partes.
- De composición, en la que la desaparición del todo implica la desaparición de sus partes.
- Información de multiplicidad (es decir, cuántos objetos pueden interactuar en una misma relación).
- Dirección de la relación.
- Tipificación de relaciones.

²⁵ J. Rumbaugh, M. Blaha, W. Premerlani, F. Hedf y W. Lorenzen: *Modelado y diseño orientados a objetos Metodología OMT*, New Cork: Editorial Prentice Hall, 1996. El oMT —Object Modeling Technique— puede considerarse un precursor de UML —Unified Modeling Language—.

Esta aproximación mediante la integración verbal procede del área pedagógica, donde esta forma de relacionar los conceptos mediante verbos recibió la denominación de mapas conceptuales (*concept maps*).

En especial, el uso de los *topic maps* posibilita la recuperación por verbos, ya sean como descriptores o como identificadores de las relaciones existentes entre descriptores. Su antecedente se encuentra en la base de datos de referencia léxica WordNet, con relaciones de equivalencia y de jerarquía entre diferentes categorías gramaticales.²⁶ La utilización de esta red semántica multidisciplinar en inglés ha estado vinculada a la desambiguación conceptual, en especial mediante el uso de verbos:²⁷

1. Se seleccionan de la frase todas las parejas sustantivo-verbo.
2. Se escoge el significado más probable del término (desambiguación terminológica):
 - a. Las palabras de la frase se agrupan en parejas.
 - b. Se buscan en WordNet los distintos significados de cada término
 - c. Se forman todos los posibles pares de conceptos.
 - d. Se busca cada par en Internet. Luego se ordenan los resultados según los conceptos más frecuentes.
3. Teniendo en cuenta los conceptos más frecuentes, se seleccionan todos los sustantivos de los «glosarios» de cada verbo y sus descendientes jerárquicos.
4. Se calcula mediante una fórmula los conceptos comunes entre los sustantivos del punto anterior.
5. Se ordenan todas las parejas de conceptos de sustantivo-verbo según el resultado de la fórmula.

Este proceso, aunque fructífero para la identificación y desambiguación, no permite la identificación de las relaciones que establecen los distintos verbos con los que trabajan los sustantivos.

²⁶ Rebecca Green, J. Dorr Bonnie y Philip Resnik: «Inducing Frame Semantic Verb Classes from WordNet and LDOCE», en *Proceedings of the conference / 42nd Annual Meeting of the Association for Computational Linguistics*, 21-26 de julio del 2004, Barcelona, España, East Stroudsburg (EE. UU.): ACL. (Barcelona: Copistería Miracle), 2004, pp. 213-224. Rebecca Green, Lisa Pearl, Bonnie J. Dorr y Philip Resnik: Mapping lexical entries in a verbs database to wornet senses, en *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, Toulouse, 2001, Morgan Kaufmann Publishers, 2001, pp. 244-251.

²⁷ Dan I. Moldovan y Rada Mihalcea: «Using WordNet and Lexical Operators to improve Internet Searches», *IEEE Internet Computing*, 4, n.º 1, 2000, pp. 34-43.

Este es el proceso que sigue la automatización de la construcción de tesauros por la aparición de sustantivos próximos a estructuras verbales:

1. Análisis de documentos relevantes para la extracción de su vocabulario (glosarios, diccionarios, estándares sobre vocabulario, etc.).
2. Depuración manual del vocabulario extraído, obteniendo los descriptores de los tesauros.
3. Indización con estos descriptores de documentos textuales maestros (manuales, estándares, artículos, etc.).
4. En esta etapa se almacenan principalmente aquellas frases del documento en las que aparece uno o varios descriptores del tesoro en el sintagma nominal sujeto, y uno o varios descriptores en el sintagma verbal.
5. Posteriormente, los conceptos dinámicos se agrupan, clasifican y se asimilan a las relaciones del tesoro que se deseen identificar.
6. Manualmente, se revisan en el tesoro las relaciones obtenidas.

Desde luego, la distribución de las relaciones de los tesauros de manera lógica mantiene los tres tipos de jerarquía, equivalencia y asociación, ampliándose a otras relaciones.

3.4. Mapas conceptuales de navegación por redes semánticas

El estudio de los mapas temáticos se originó en la necesidad de crear índices en torno a algún concepto o materia. Las redes semánticas son un método común de representar el conocimiento en el campo de la Inteligencia Artificial, que busca establecer comunicación entre las personas y las máquinas.²⁸

- Grafos o redes conceptuales constituidos por conceptos y relaciones de conceptos.²⁹
- Son colecciones ordenadas de nodos conectados por arcos que se usan para representar documentos.

²⁸ D. H. Jonassen: *Handbook of research for educational communications and technology*, Nueva Jersey: Prentice Hall, 1996.

²⁹ D. E. Rumelhart: «Schemata and the Cognitive System», en R. S. Wyer y T. Srull (eds.): *Handbook of Social Cognition*, Hillsdale, NJ, Erlbaum, 1984.

- Un tipo de grafo es la red semántica, que representa las relaciones semánticas que se establecen en el texto.

Su antecedente inmediato fueron los *mapas lexicográficos*,³⁰ representación gráfica de los índices que permitía mostrar los términos relacionados en mapas lexicográficos, con la ventaja de que cada palabra aparece representada en su contexto, con lo que se desambigua su significado.

- Tras asignarse algorítmicamente un determinado peso a las palabras significativas, se establece una representación asociativa entre ellas, mediante redes basadas en la coocurrencia,
- Se consigue una representación gráfica en forma de red o mapa lexicométrico, en cuyo interior se sitúa cada palabra en relación con las otras palabras, lo que permite desambiguar su significado sin tener que recurrir al análisis semántico.
- Esas palabras asociadas conforman el espacio de indización del documento, pues reflejan las relaciones que realmente existen entre las palabras que lo representan.

Los *mapas conceptuales* son una técnica para representar el conocimiento en gráficas cognitivas → redes de conceptos que se componen:

- de nodos (puntos/vértices) que representan conceptos y
- de enlaces (*arcs*: arcos/*edges*: extremos, satélites) que representan las relaciones entre los conceptos.

Esta técnica fue desarrollada por Joseph Novak en los años sesenta, como herramienta para ayudar a enseñar y aprender en las clases.³¹ Se basó en la idea de que el conocimiento está representado por frases simples con estructura gramatical: sujeto-verbo-predicado; como una relación: concepto → relación → concepto.

³⁰ Así actúa el sistema LEXIMAPPE: J. Chaumier, M. Dejean: «L'indexation documentaire: de l'analyse conceptuelle humaine à l'analyse automatique morphosyntaxique», *Documentaliste-Sciences de l'information*, 27, n.º 6, 1990, pp. 265-279.

³¹ J. D. Novak, D. B. Gowin y D. Johansen: «The Use of the Concept Mapping and Knowledge Vee Mapping with Junior High School Science Students», *Science Education*, 67, n.º 5, 1983, pp. 625-645. J. D. Novak y D. B. Gowin: *Aprendiendo a aprender*. Barcelona: Martínez Roca, 1988.

- Estas relaciones pueden representarse mediante grafos en los que las ideas o los conceptos se sitúan en los nodos y las relaciones en los enlaces.
- De forma que un documento o dominio del conocimiento puede representarse por un conjunto de conceptos interrelacionados y dispuestos como un grafo.

3.4.1. RASGOS DE LOS MAPAS CONCEPTUALES

- Los mapas conceptuales ofrecen una red de relaciones más rica que los tesauros.
- El concepto que origina su estructura (nodo-enlace-nodo) es próximo a su equivalente en las redes hipertextuales, por lo que soportan la navegación de un modo muy natural.
- El uso de mapas conceptuales permite desarrollar mejores mecanismos de representación y recuperación, ya que las relaciones entre los conceptos se eligen teniendo en cuenta la modelización de las necesidades y expectativas de cada usuario.

La construcción de mapas de conceptos puede perseguir diferentes objetivos:

- Para generar ideas (tormenta de ideas, etc.).
- Para diseñar estructuras complejas (textos largos, sitios web, etc.).
- Para comunicar ideas complejas.
- Para ayudar al aprendizaje, al hacer patente la integración del nuevo conocimiento con el viejo.

*Proceso de construcción de un mapa conceptual:*³²

- 1) Selección de los conceptos que se representarán en el mapa.
- 2) Listado de esos conceptos.
- 3) Agrupación de los conceptos relacionados.
- 4) Ordenación de los mismos en forma bidimensional o tridimensional.
- 5) Enlace de cada par de conceptos mediante líneas etiquetadas.

³² A. Ontoria: *Mapas Conceptuales*, Madrid, Narcea, 1992.

La pretensión de visualización que persiguen los mapas conceptuales, como procedimiento para la representación esquemática de las nociones esenciales, estaba planteada ya en el *Árbol de Porfirio*, primera red semántica o, si queremos, primer mapa conceptual.³³

3.5. Topic maps

Un *topic map* es un documento, o un conjunto de documentos SGML o XML interrelacionados en un espacio multidimensional en el que las localizaciones son *topic*. La distancia entre dos conceptos es medida por el número de conceptos intermedios que son necesarios para viajar de uno a otro.³⁴ Aparecen como ideas fundamentales la representación de los conceptos o elementos de recuperación de la información, los diferentes casos que ofrecen (conjuntos de objetos informativos en tono a los conceptos), y las asociaciones entre conceptos y objetos. El *topic* y las asociaciones de *topics* construyen una red semántica estructurada que se acopla con recursos de información. Los elementos clave que conforman un *topic map* son:³⁵

- **Topic** (p. ej., <ISKO>).
- **Topic type** (p. ej., <jornadas científicas>).
- **Association** <tiene lugar en>.
- **Association type** <tener lugar en> (localización).
- **Scope** (ámbito en el que una relación tiene sentido) *theme* <Representación documental>.
- **Topic occurrence** (p. ej., <<http://www.isko.org>>).
- **Occurrence type** (p. ej., Página web) [filtro por facetas].

³³ J. F. Sowa: *Knowledge representation: Logical, Philosophical and Computational Foundations*, Pacific Grove: Brooks/Cole Thompson Learning, 2000.

³⁴ ISO/IEC JTC 1/SC34. 2004. *Information Technology-Document Description and Processing Languages. SGML-Topic Maps*, <<http://www.topicmaps.com/content/resources/iso13250/iso13250-2000-fcd.htm>>. ISO/IEC JTC 1/SC34. 2004. *Information Technology-Document Description and Processing Languages. Topic Maps-XML Synta3*, <<http://www.jtc1sc34.org/repository/0495.htm>>.

³⁵ J. A. Moreira, S. Sánchez Cuadrado y J. Morato: *Panorámica y tendencias en topic maps* [en línea], 2002, en C. Rovira y L. Codina (dirs.): *Documentación digital*, Barcelona: Sección Científica de Ciencias de la Documentación. Departamento de Ciencias Políticas y Sociales. Universidad Pompeu Fabra, <<http://www.documentaciondigital.org>>, [consulta 23-07-2005].

Se destacan como características de los *topic maps*:

- Control de vocabulario ad hoc.
- Dominios con validez y expansión potencialmente ilimitados.
- La tipología de *topics* es potencialmente infinita.
- Relaciones: se establecen usualmente mediante verbos.
- Tipología de relaciones: número potencialmente infinito.
- Una relación puede ser un *topic*.
- Se aceptan todas las variaciones ortográficas, son normalizadas a posteriori.
- Los *topics* se definen mediante sus *occurrences* y *associations*.

De entre ellas, destacamos tres como principales:

- La denominación (*topic names*), que acepta las diferentes formas de denominación que puede tener un *topic*, incluidas todas las variaciones formales, las de carácter simbólico, y hasta los apodos.
- Las apariciones o casos (*occurrences*), guía dirigida a los recursos de información internos (una definición, por ejemplo) y externos que contienen el *topic*.
- La función de las asociaciones (*role associations*), ya que la multiplicidad de asociaciones entre los conceptos, de acuerdo con las posibilidades que ofrece la lógica y la vida, engloba cuantas asociaciones puedan darse: sinonimia, ubicación, generalización, vivir en... Mostrando como rasgo más original la *association type*, que está definida por la forma verbal que une los *topics*.

Las propiedades que caracterizan a las *association type* implicadas en un *topic map* determinan las inferencias que se pueden realizar entre los componentes de un conjunto de asociaciones:

1. Simetría.
2. Transitividad.
3. Mejora de la recuperación.

Una de las principales propuestas de los *topic maps* es la de visualizar la Web semántica. El hecho de que un *topic map* puede presentar miles de asociaciones de diferente tipología (*association types*, *roles*, *occurrences*, etc.) supone un problema. A la hora de analizar estas representaciones se debe tener en cuenta que

muestren tanto información local de los *topics* que interesan al usuario como información sobre la localización de esos *topics* en el conjunto del *topic map*. Usar para ello árboles con hipervínculos es uno de los principales inconvenientes de este planteamiento, pues si bien la percepción del *topic map* es más sencilla para el usuario, no pasa lo mismo con los distintos elementos del *topic map*.³⁶ La tipología de los *topic maps* (*topics*, *occurrences*, *roles*, *types*) hace que el usuario pueda desorientarse ante un árbol de estas características.

3.6. La indización colaborativa de contenidos: la folksonomía

Este neologismo surgió para denominar una práctica de clasificación colaborativa en la que se emplean de forma libre palabras clave. La propuesta del término *folksonomía* se atribuye a Thomas van der Wal como resultado de combinar las palabras *folk* (gente, popular) y *taxonomía* (gestión —*taxis*— de la clasificación —*nomos*—), de forma que *folksonomía* viene a ser etimológicamente «clasificación gestionada popularmente». En su elaboración interviene, pues, gente que coopera de forma espontánea para organizar la información en categorías,³⁷ de donde se deriva la ventaja de ser una arquitectura de la información que se construye de manera social y cooperativa, en la que cada usuario escoge lo que más se ajusta a su entendimiento. Se permite, así, que los internautas asignen libremente palabras clave personales mediante etiquetas en las que se categoriza todo tipo de contenidos, desde enlaces de noticias a fotografías, pasando por artículos especializados, canciones, en fin, todo tipo de información y de objetos con tal de que dispongan de una URL, siempre con el propósito de prepararlos para su recuperación.

Se viene utilizando el marcado con etiquetas de metadatos (*tagging*) como un modo de indizar la información porque hace más fácil la navegación. Esta nueva forma de búsqueda la emplea Technorati para determinar las categorías de los blogs <www.hechnorati.com>, sin duda debido a que las etiquetas pueden ser públicas y vistas por todos. De manera que en la era de la globalización que supone Internet, el conocimiento se organiza localmente, lo que significa que ninguna categoría puede ser considerada universal, y se hace mediante la libre participación

³⁶ Jack Park y Sam Hunting (eds.): *XML Topic Maps: Creating and Using Topic Maps for the Web*, Addison-Wesley Professional, 2002.

³⁷ Thomas van der Wal fue el creador del término: *folk + taxonomy* = gente + taxonomía, clasificación cooperativa.

a la hora de asignar etiquetas para identificar y clasificar información digitalizada. El marcado con etiquetas se efectúa con carácter social y cooperativo, pues se lleva a cabo por la misma persona que usa la información, pero a su vez es compartido y está abierto a otros, produciéndose así una rentabilidad social, pues todos contribuyen al etiquetado y se benefician del que hacen los demás.

Así pues, el empleo de las folksonomías debe entenderse dentro de las propuestas que buscan llegar a la tan procurada Web semántica, donde cada página web dispone de unos metadatos en los que se describe su contenido. La dificultad real de que los autores asignen metadatos a las páginas que han construido, ha llevado a proponer la folksonomía como sistema de indización distribuido cuyos costes de entrada son bajos, pues elude el coste económico que supone aprender a usar los metadatos o cualquier vocabulario controlado y luego asignarlos.

La principal novedad que aportan las folksonomías está en que se retroalimentan, a diferencia de lo que sucede con cualquier otro vocabulario.³⁸ Este uso colectivo de etiquetas genera un sistema de categorización no jerárquico, ya que, como defiende Thomas van der Wal, construir una jerarquía no es algo esencial, mientras que sí lo es un instrumento que ofrezca la posibilidad de emplear etiquetas semánticas para efectuar la indización. Al estudiar los metadatos generados por los usuarios y ver cómo son empleados en los servicios web para compartir y organizar la información, Adam Mathes observó que las folksonomías no presentan relaciones jerárquicas ni de otro tipo,³⁹ si bien encontró que, pese a que no existe una relación formal entre los términos, se establecen relaciones de forma natural mediante las URL que aquellos describen.

De forma que nos movemos dentro de unas categorías que evolucionan según vayan participando los usuarios, en vez de contar con rígidas categorías predefinidas. No solo son importantes los términos más populares, sino también aquellas etiquetas no dominantes (*meta-noise*) pero que añaden comprensión semántica.⁴⁰ El elemento clave de las folksonomías está en que se fundamentan sobre unos metadatos de asignación barata, que deben cumplir, al menos, estas condiciones:

³⁸ Permite que alguien marque una página web que encuentra interesante, que organice estas páginas por categorías usando etiquetas libremente y que comparta sus descubrimientos con otras personas. Jon Udell publicó una serie de artículos acerca de Del.icio.us en su blog: <weblog.infoworld.com/udell/> [consulta 29-07-2005]. *How Do You Use del.icio.us?* <http://www.primidi.com/2004/11/08.html#a1019> [consulta 29-07-2005].

³⁹ <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html> [consulta 30-07-2005].

⁴⁰ *Folksonomie. 2004. Many 2 many. A group weblog on social software*, <www.corante.com/many/archives/2004/08/25/folksonomy.php> [consulta 28-07-2005].

1. Están hechos por cualquier persona.⁴¹
2. Para crearlos se requiere aprender muy pocas reglas.
3. Se producen sin interés propio.
4. Las ventajas de los usuarios crecen con la agregación.
5. No se rompen cuando hay datos incompletos.
6. Se presentan en forma de conjuntos, y no de jerarquías.
7. No están diseñados a priori, lo que les hace más flexibles.
8. No pertenecen a nadie, pues nadie los centraliza ni controla.
9. Son relevantes para los propósitos e intereses de un sitio web.

Dos sitios web donde el término cobra sentido son Del.icio.us social bookmarks y Flickr. Los creadores de Del.icio.us se definen como gestores sociales de sitios Web favoritos (*bookmarks*, almacén de enlaces). Su producto permite añadir a los favoritos las páginas deseadas, clasificar mediante palabras-clave los sitios para poder recuperarlos después, e intercambiar esta información con otros usuarios. A través de una palabra clave propuesta por un usuario se puede navegar por los contenidos etiquetados. Se capta, pues, el conjunto de términos que unos usuarios emplean para clasificar un contenido. Un problema de Del.icio.us es que, cuando un usuario va a asignar una etiqueta o palabra clave, este no permite introducir espacios, con lo que la folksonomía se conforma tan solo con unitérminos.⁴²

Por su parte, Flickr (<http://www.flickr.com/>) es un sitio web sobre fotografía en el que usuarios de cualquier lugar pueden compartir fotos poniéndolas en disposición de consulta mediante un sistema organizativo basado en asignar palabras clave en un entorno colaborativo. Es el propio usuario el que clasifica cada fotografía de acuerdo con lo que su contenido le sugiere.

En esos dos ejemplos se puede observar que existe una distinción entre:⁴³

- **Broad folksonomy** o folksonomía genérica, en que diferentes usuarios etiquetan el mismo objeto y cada uno puede hacerlo con sus propias etiquetas o vocabulario, tal como sucede con Del.icio.us, por eso se las conoce también como *re-lacionadas*. En este caso son sitios web los que se etiquetan y comparten.

⁴¹ A. Weiss: «The power of collective intelligence», *netWorker*, 2005, 9, n.º 3: 16-23.

⁴² Otra iniciativa similar a Del.icio.us pero quizás menos popular, es CiteULike.com. A free online service to organise your academic papers, <www.citeulike.org/> [consulta 31-07-2005].

⁴³ Thomas Vander Wal: *Explaining and Showing Broad and Narrow Folksonomies*, <http://www.vanderwal.net/random/entrysel.php?blog=1635> [consulta 23-07-2005].

- **Narrow folksonomy** o **folksonomía específica**, en que una o pocas personas etiquetan la información. Así sucede con Flickr, cuyo uso está determinado por el hecho de poder alojar fotografías propias y etiquetarlas para que luego otros puedan buscar mediante esas etiquetas. De manera que en esta folksonomía se clasifica según la forma en que los usuarios etiquetan sus contenidos.

3.6.1. FOLKSONOMÍAS FRENTE A TESAUROS

Aunque, por su utilización novedosa, puede dar la sensación de estar frente a una moda tanto en lo que respecta al significado como al propio término, las folksonomías permiten estudiar qué términos son los más empleados y cómo evolucionan de acuerdo con las tendencias y según las comunidades que los empleen. Los vocabularios controlados están principalmente destinados a encontrar la información. Al aumentar tanto la cantidad de información existente, puede volverse imposible o ser económicamente inviable localizar cualquier asunto. Lo que corrobora que mantener una taxonomía organizada de arriba hacia abajo se ha hecho muy difícil.

Por otra parte, una folksonomía, aunque naturalmente sea menos exacta, invita a los usuarios a investigar, a descubrir el contenido de los sitios web. Implica la tarea de recuperar la información en un contexto más amplio, en donde el proceso de búsqueda se desgrana en otros más pequeños dentro de una experiencia orgánica enriquecedora.

Los vocabularios controlados no son aplicables ni económicamente extensibles a la mayoría de los casos donde podrían utilizarse metadatos. Su construcción, mantenimiento y aplicación resultan a menudo demasiado costosos tanto en tiempo de desarrollo como en el tiempo que necesita el usuario del sistema para aprender el esquema de la clasificación.

De manera que las folksonomías son una solución cuando la indización o clasificación tradicionales no son viables.⁴⁴

En contraste con los autoritarios sistemas taxonómicos organizados de arriba hacia abajo, las categorías de las folksonomías surgen del compromiso de un in-

⁴⁴ Clay Shirky: *Folksonomies + controlled vocabularies*, 2005, <http://www.corante.com/many/archives/2005/01/07/folksonomies_controlled_vocabularies.php> [consulta 7-08-2005].

dividuo con el contenido que etiqueta, por lo que la categoría creada es al mismo tiempo personal, social y (en cierto grado) sistemática, sin olvidar nunca que son al mismo tiempo imperfectas y provisionales. Se trata de sistemas de indización abiertos que emergen de forma espontánea, desordenada y sin coordinar, frente a los sistemas tradicionales de indización, que son jerárquicos y tienen unas categorías prefijadas de forma rigurosa, en ellas un concepto se clasifica en especificidad descendente según el nivel que ocupe en su categoría conceptual.⁴⁵ Mientras que los nuevos servicios, como el sitio para albergar fotos Flickr o el servicio para guardar enlaces del.icio.us, dan libertad al usuario para que cree su propia etiqueta, después emerge la folksonomía del uso colectivo. Así, en Flickr, lo que comienza siendo una etiqueta personal se convierte en una categoría en la que se agrupan cientos de fotos en torno al mismo asunto, que además el propio servicio de Flickr explota. A partir de una de las fotos puede saltarse a otras de la misma etiqueta pertenecientes a otras decenas de internautas. Navegando a través de una etiqueta se puede llegar a decenas de artículos sobre el mismo tema que han sido guardados con etiquetas coincidentes asignadas por otros. De este modo resulta una alternativa fácil y espontánea a la categorización jerárquica tradicional, pues los metadatos se asignan con rapidez y facilidad, sin necesidad de recorrer ninguna estructura jerárquica, por lo general de difícil comprensión para quienes no sean expertos en el dominio reflejado en ella.

Es evidente el valor de muchos de los términos propuestos como candidatos a integrarse en vocabularios controlados, ya que se trata de términos o conceptos comúnmente empleados por los usuarios de Internet, aunque esos usuarios no sean profesionales de la información, a diferencia de lo que sucede con los tesauros y sus creadores.

Por otra parte, las etiquetas de las folksonomías pueden facilitar conexiones transculturales e interdisciplinarias. Además, compensan los términos que faltan en las clasificaciones particulares o en la taxonomía de un sitio web.

De momento, su aplicación presenta muchas debilidades. La folksonomía no ofrece un lenguaje controlado, pues no tiene jerarquías ni otro tipo de asociaciones más allá de los posibles clusters de recursos que se puedan formar según las clasificaciones o palabras clave de los usuarios.

Así pues, al carecer de un lenguaje controlado, la ambigüedad es muy alta. Para un usuario, un término tiene un significado o puede expresar una realidad com-

⁴⁵ B. Berlin: *Ethnobiological Classification*, Princeton: Princeton University Press, 1992.

pletamente distinta a la de otro usuario. No hay guías de uso ni limitaciones o restricciones a cómo etiquetar los contenidos.

Entre sus debilidades está también la falta de consistencia subsiguiente a la ausencia de control de los sinónimos. Puede considerarse como ejemplo el que, a la hora de denominar un ordenador Macintosh, puedan emplearse las etiquetas *mac*, *apple*, o *macintosh*.⁴⁶ También, cierta imprecisión derivada de que se utilizan etiquetas que solo permiten manejar unitérminos. Igual sucede con el empleo indistinto de singulares y plurales, así en Flickr se admite la búsqueda y el etiquetado según plurales o singulares, con recuperaciones dispares, pues se trata de distintas fotografías que hacen referencia a un mismo concepto, pero que se buscan por términos diferentes. Y, por supuesto, cuanto se deriva de la ausencia de jerarquía y de los propios tipos de contenidos analizados (*bookmarks*, fotos), que son bastante simples.

3.6.2. ¿PARA QUÉ SIRVE UNA FOLKSONOMÍA?

Es evidente que las folksonomías están funcionando debido a que resultan efectivas para los usuarios a la hora de ordenar su información (en principio, para ser recuperada por ellos mismos, aunque luego la puedan compartir). Además, como se trata de usuarios inexpertos en la organización del conocimiento, el tiempo que tardarían con complejos sistemas de clasificación jerárquicos queda minimizado con este método.

Dado que quienes organizan la información en los blogs y los defensores del uso social del software suelen ser los principales usuarios, las folksonomías producen resultados que reflejan con exactitud el modelo de conceptos que tiene la gente. Se trata de unos sistemas simples y emergentes, que se alejan de la manera tradicional de clasificar desde arriba hacia abajo, por lo que su ventaja proviene de la capacidad de emparejar las necesidades verdaderas de los usuarios y la lengua, no de buscar su precisión.⁴⁷

Abandonar las taxonomías en favor de las listas de palabras clave no es algo nuevo, si bien la diferencia fundamental entre ambas está en que las folksonomías ofrecen la opción de retroalimentarse. Se benefician del vocabulario personal y de

⁴⁶ Gene Smith enumera algunos inconvenientes de las folksonomías en su *Folksonomy: social classification* <http://atomiq.org/archives/2004/08/folksonomy_social_classification.html> [consulta 30-07-2005].

⁴⁷ Emanuele Quintarelli: *Folksonomies: power to the people*, 2005, <<http://www.iskoi.org/doc/folksonomies.htm>> [consulta 8-08-2005].

la contribución social para ayudar a encontrar en la Web los objetos que son interesantes y controlarlos. Es importante entender quién está marcando con etiquetas y cómo se aprovechan después.

Otra propiedad de las folksonomías es que incluyen cada una de las palabras propuestas, sin dejar ninguna fuera. No hay una autoridad central que imponga su visión, de forma que cada voz tiene su lugar. Este aspecto implica que tenemos también la opción de derivar hacia ideas no entroncadas con la corriente principal, que pueden emerger del interés de una parte pequeña de la población.⁴⁸

Marcar con etiquetas de carácter social puede considerarse «un puente de baja inversión entre la clasificación personal y la clasificación compartida».⁴⁹ De hecho, la clasificación facetada ofrece más flexibilidad y una aproximación semántica. Sin embargo, las facetadas no podían ser la respuesta correcta cuando la gente está obligada a clasificar el contenido que ella misma produce, como sucede hoy. Si el número de términos y de sus combinaciones se amplía, las clasificaciones facetadas multiplican el número de las decisiones requeridas para clasificar un documento, lo que agrega un significativo coste cognitivo a la clasificación. A menudo, esta inversión cognitiva es demasiado elevada para los *bloggers*, que prefieren describir sus envíos con palabras claves libres. De modo que, por encima de su posible eficacia en la recuperación, el uso de las folksonomías tiene como valor principal su carácter social y cooperativo, pues por su medio los servicios de almacenamiento y acceso a la información crean nexos entre los usuarios.⁵⁰

Entre las perspectivas de uso es conveniente hablar del lugar llamado a ocupar por las folksonomías en las empresas, puesto que las folksonomías son generadas por los usuarios y es, por lo tanto, barato ponerlas en marcha, ya que suponen una alternativa barata y más útil que las tradicionales taxonomías institucionales o los vocabularios controlados. Las folksonomías pueden además resultar útiles para facilitar la democracia dentro del lugar de trabajo y la distribución de las tareas de gestión entre la gente que trabaja allí.

Entre las dudas que nos suscita este modo de indizar hay que notar que todavía muchos especialistas o conocedores de temas específicos echan en falta la presen-

⁴⁸ Joshua Porter: *Controlled vocabularies cut off the long tail*, 2004, <http://bokardo.com/archives/controlled_vocabularies_long_tail/> [consulta 8-08-2005].

⁴⁹ Jess, McMullin: *The cognitive cost of classification*, 2005, <<http://www.interactionary.com/inde3.php?cat=7>> [consulta 18-08-2005].

⁵⁰ Thomas Vander Wal: *Off the Top: Folksonomy Entries*, <<http://www.vanderwal.net/random/category.php?cat=153>>, 2 de noviembre del 2005 [consulta 11-11-2005].

cia de términos que describan sus intereses. Asimismo es notable el esfuerzo por alcanzar estándares de etiquetar abiertos para todos con formatos flexibles.⁵¹

Consideración final

Ante los inconvenientes que para su tratamiento plantea el número de documentos que circulan por la Web y la variedad de sus soportes, se han buscado soluciones de afán convergente a partir de la lingüística, la estadística y la informática. Por una parte, se entró hace más de una década en un proceso evolutivo de los estándares sobre tesauros hacia mapas de representación del conocimiento y, por otra, se potenció el uso de lenguajes libres, en especial desde la perspectiva del usuario.

Estas transformaciones han llevado a la creación de relaciones adaptables a cada dominio, así como la ampliación del concepto tesaorial de descriptor por la admisión de nuevas categorías gramaticales que han enriquecido con nuevos matices la semántica del mapa conceptual y, desde luego, aumentando con nuevas categorías las relaciones interconceptuales, que han alcanzado incluso a los recursos de información, y que han extendido las posibilidades de asociación conceptual aproximándolas a la riqueza casuística del lenguaje natural.

Con ello, se ha roto la rigidez que gobernaba la construcción y el uso de los lenguajes combinatorios para pasarse a aceptar nuevas propuestas que o buscan enriquecer la identificación de los conceptos o plantean una visualización cuya más funcional mediante mapas, una mayor exactitud del significado de acuerdo con el contexto de origen y de uso, o aceptan la indización hecha individual o socialmente por los usuarios propensos a consultar un tipo de documentos digitales.

Estas alternativas a los lenguajes combinatorios buscan activar la elaboración de tesauros representativos de los contenidos de un documento o de un grupo de ellos pertenecientes al mismo ámbito cognitivo, identificar adecuadamente los conceptos que contienen los documentos, a la vez que especificar las correspondencias terminológicas para mejorar la representación de las materias y aumentar la precisión y eficacia en un contexto que admite nuevas y más ricas relaciones, expresadas con explicitud gráfica para favorecer la navegación, la visualización y las conexiones que en la información siempre están presentes.

⁵¹ Mary Hodder y otros: «Engines of meaning: How will we scale our understanding?», *Symposium on Social Architecture*, 2005, noviembre, 14-15, Harvard University, Cambridge, MA, <<http://www.corante.com/events/ssa/program.php#session3>>. La iniciativa cuenta con el apoyo de Technorati, Feedster y Pubsub.

HERRAMIENTAS PARA CONSTRUIR TOPIC MAPS, ONTOLOGÍAS Y MAPAS CONCEPTUALES

En la dirección <<http://www.ontopia.net/download/freedownload.html>> puede bajarse una versión libre de Ontopia Omnigator para uso no comercial. Este programa se ha construido usando la Ontopia Topic Map Navigator Framework. Se facilita una Demo de Omnigator en: <<http://www.ontopia.net/omnigator/models/index.jsp>>.

Demo de *topic maps*: Hyperbolic Tree <<http://k42.empolis.co.uk/demo/demo.html>>.

Una página muy completa sobre cualquier tema relacionado con *topic maps*: <<http://www.topicmap.com/>>.

Home of SC 34/WG3 Information Association <<http://www.isotopicmaps.org/>>.

Algermissen. J. Semantic web technologies. Knowledge representations (Thesauri, Ontologies) <<http://www.topicmapping.com/v.html>>.

Metamap: <<http://mapageweb.umontreal.ca/turner/meta/francais/metrometa.html>>, [consulta 30-04-2004].

Visualthesaurus: <<http://www.visualthesaurus.com>>, [consulta 30-04-2005]. <<http://www.techquila.com/topicmaps/tmworld/11584.html>>, [consulta 13-11-2005].

<<http://erpus.com>>, [consulta 15/11/2005].

HERRAMIENTAS PARA HACER ONTOLOGÍAS

<<http://protege.stanford.edu/>>.

<<http://www.mindswap.org/~aditka/editor.shtml>>.

<<http://jena.sourceforge.net/ontology/>>.

<<http://www.ksl.stanford.edu/software/chimaera/>>.

EJEMPLOS DE MAPAS CONCEPTUALES COMO RECURSO DIDÁCTICO

<<http://www.conceptmaps.it/KM-MapExamples-esp.htm>>.

<<http://www.ihmc.us/users/acanas/Publications/RevistaInformaticaEducativa/HerramientasConsConRIE.htm>>.

<<http://www.kartoo.com>>.

Pueden bajarse programas para hacer mapas conceptuales en:
<<http://www.conceptmaps.it/KM-Download-esp.htm>>.

SITIOS WEB RELACIONADOS CON LAS FOLKSONOMÍAS

CITEULIKE: <www.citeulike.org/>.
DEL.ICIO.US: <<http://del.icio.us/>>.
FEEDSTER: <<http://feedster.com/>>.
FLICKR: <www.flickr.com/>.
IFAVORITOS: <www.ifavoritos.com/>.
TAGCLOUD: <www.tagcloud.com/>.⁵²
GENIELAB: <<http://genielab.com/>>.
PUBSUB: <<http://pubsub.com/>>.
TECHNORATI: <www.technorati.com/>.

⁵² TagCloud es un clasificador automatizado de fuentes RSS de noticias, busca cualquier número de las alimentaciones de RSS (formato de archivo derivado del XML y usado para la distribución de contenidos) que se le especifique, extrae las palabras clave del contenido y las lista de acuerdo con su predominio dentro de las alimentaciones de RSS. Pinchando sobre el enlace de la etiqueta aparecerá una lista de todos los resúmenes de artículos asociados con esa palabra clave.

Capítulo 4

GESTIÓN DEL CONOCIMIENTO, GESTIÓN DOCUMENTAL Y GESTIÓN DE CONTENIDOS

Mario Pérez-Montoro

4.1. Introducción

En los últimos años, y cada día de una forma más clara, se está consolidando a un ritmo constante un nuevo escenario económico internacional. Este nuevo escenario se caracteriza, principalmente, por una globalización de los mercados y por una nueva etapa en las relaciones económicas entre las empresas. Ya no existen mercados cerrados. Las empresas ya no solo tienen que competir con otras corporaciones situadas en la misma comarca o región, sino que, en muchas ocasiones, las amenazas vienen de firmas situadas a miles de kilómetros de distancia.

En medio de este nuevo escenario de feroz competencia y cambio continuo, las empresas, para garantizar su supervivencia, se ven obligadas a buscar nuevas estrategias adaptativas. Dentro de este contexto de cambio, se empieza a percibir de una manera clara que los activos materiales o de capital ya no son exclusivamente el único criterio para conseguir la supremacía del mercado. Se comienza a entender que son los activos de naturaleza más intangible que posee una organización —entre los que destaca de forma especial el conocimiento— los que pueden ofrecer ese valor añadido que permita a una empresa diferenciarse, de una manera clara, del resto de la competencia frente sus clientes o potenciales consumidores y garantizar, también, el correcto funcionamiento y la supervivencia de esa organización.

Es en ese nuevo panorama organizacional, y relacionada de forma estrecha con la búsqueda de soluciones que puedan cubrir esas nuevas necesidades de adaptación y supervivencia, donde emerge con fuerza una nueva disciplina: la gestión del conocimiento. Esta nueva disciplina persigue el claro objetivo de implementar programas