La indización asistida para el manejo de conocimiento en organizaciones frente al tratamiento documental en bibliotecas

Por
Mela Bosch
(Argentina)
Docente online de la Cátedra Taller en Tecnología en Comunicación Social, Facultad de Periodismo y Comunicación Social, Universidad Nacional de La Plata.
melabosch@europe.com

Resumen

El trabajo es una revisión del tema de la indización asistida desde la óptica de la Bibliotecología y Documentación. Se parte de conceptos de base sobre la indización y los métodos automáticos de indización. Se realizan precisiones terminológicas y conceptuales sobre los sistemas asistidos de indización automática, metodologías e interactividad, y se presentan las áreas y disciplinas involucradas. Se consideran los aspectos teóricos de las estructuras de representación del conocimiento a través de las taxonomías en estos sistemas. Finalmente se presenta la temática de la información estructurada, en relación con aquella no estructurada así como las tecnologías para su tratamiento considerando su importancia para el desarrollo del Web Semántico. La conclusión es abierta y plantea aspectos que aún debemos afrontar acerca de la determinación de los rasgos distintivos de los documentos digitales en las organizaciones y el papel de los documentalistas en su tratamiento.

Palabras claves:
Documentos digitales, Indización, Información estructurada, Información no estructurada, Taxonomías

Abstract:

This article revises the subject of computer assisted indexing systems by using library and documentation science concepts. We consider terminological aspects and we outline indexing base concepts and methodologies for automatic indexing. Theoretical aspects of the knowledge representation structures through taxonomies in these systems are considered. We also introduce areas and disciplines involved in computer assisted indexing process. Having described these matters, the article reflects upon the role of the structured information and unstructured information emphasizing their importance for the Semantic Web development. We then take a look and the present trends in our field.

Keywords:
Computer-assisted indexing, Digital documents, Structured information, Taxonomies; Unstructured information

1. Objetivos y metodología

El uso de sistemas de análisis de contenido en forma automática se encuentra en este momento en fuerte crecimiento en el contexto del manejo de conocimiento en organizaciones: empresas, oficinas, organizaciones de servicio y logística. El aporte de este trabajo es una revisión del tema de la indización asistida desde la óptica de la Bibliotecología y Documentación. Dada la vastedad de la problemática, nos proponemos un primer acercamiento que ofrezca algunas bases que permitan continuar y complementar el estudio de manera más detallada y aplicada.

El presente estudio surge de los intercambios realizados en el marco de la **comunidad de práctica virtual Taxoar**, (http://www.cnea.gov.ar/cac/ci/taxoAR/index.htm) desarrollada a partir de la iniciativa de colaborar en el aporte en lengua española para la elaboración de la taxonomía para el Computer Assisted Indexing System del INIS (Internacional Nuclear Information System) del OIEA (Organismo Internacional de Energía Nuclear). Este grupo, constituido por profesionales argentinos de diferentes disciplinas, aporta al tema desde diferentes enfoques: bibliométrico, informático, documental.

Para este trabajo la **metodología** es de **tipo empírico**, se han estudiado productos de mercado y asistido a seminarios de presentación de algunos de ellos. Pero no es la intención de la autora hacer una comparación de desarrollos de software, ya que ya se han realizado informes muy rigurosos al respecto, (Nevyjel, 2003), sino considerar los conceptos dominantes en estos desarrollos.
En **síntesis,** este trabajo presenta:

a) Un recorrido de los **conceptos base de las ciencias de la información** y la documentación que deben o deberían animar los *Computer Assisted Indexing systems (CAI)*
b) **Precisiones terminológicas** sobre los conceptos fundamentales involucrados
c) Presentación de las metodologías, disciplinas y áreas involucradas los sistemas de indización automática
d) Los aspectos fundamentales del **paso del tratamiento de información estructurada al paso de información no estructurada**.

2. Conceptos de base: La indización

Según la norma ISO 5127, la **indización** es la **representación del contenido** de un documento mediante **selección de las palabras extraídas** de su contexto o mediante la **atribución de términos** pertenecientes a un lenguaje controlado.

El concepto básico es la representación del contenido. La norma alude luego a dos tipos de procedimientos complementarios para hacerlo: la selección de palabras pertenecientes al contenido del documento o la asignación de términos de un lenguaje documental externo al documento. Es decir que se realiza una extracción o bien, o complementariamente, se realiza una asignación, siempre de unidades lingüísticas.

Es útil tener en cuenta estos procedimientos porque son muy diferentes en cuanto al tipo de tarea y a las unidades lingüísticas que involucra: para la **extracción**, se seleccionan las **palabras de un contexto** y en el segundo se **asignan términos de una herramienta externa al mismo.** Las unidades lingüísticas que se extraen de un conjunto discursivo son **palabras, unidades léxicas**. En tanto que, aquello que es asignado, es una unidad lingüística que proviene de una herramienta terminográfica, es decir que es un término. Un **término**, según la norma ISO 1087, es la designación por medio de una unidad lingüística de un **concepto definido en una lengua de especialidad** o en un **lenguaje controlado**.

Muchas veces la extracción de palabras tiene como objetivo transformarlas en términos que representen el contenido documental, ya sea considerándolas como tales luego de alguna normalización, o bien pasando del momento extracción al de asignación.

En el paso del momento de extracción al de asignación hay un proceso de traducción que convierte las unidades discursivas, las palabras, en términos. Sabemos que hablar de traducción es controvertido, nos limitamos a entenderlo en el sentido amplio de la representación de un sistema semiótico en otro. (Eco, 2002).

Este proceso de traducción implica un análisis de la estructura conceptual de las palabras extraídas de un documento y la adecuación de estas palabras a los términos del lenguaje controlado que se usa para representar el conocimiento del dominio del sistema documental. De manera que **la indización de un documento** es siempre una **función relativa** al dominio de **conocimiento** y al **lenguaje controlado** que se usa para representarlo.

Esta es una tarea que puede realizarse en forma **manual**, valiéndose de **soporte papel** en todo el proceso, en forma **semi automática**, utilizando **lenguajes controlados en soporte digital**, y finalmente de manera **automática** en dos modos: **totalmente automático**, y en forma **asistida**.

3. Los métodos automáticos de indización

La generación de **índices** en una base de datos con las palabras extraídas del **título**, **resumen** y **palabras claves** del autor, fue el primer método utilizado para indizar documentos en bases de datos bibliográficas. Los **algoritmos de indización** usan una combinación de **frecuencias absolutas y relativas**. Se calcula la frecuencia en el documento y se los compara con la ocurrencia en el resto de la base de datos. Pueden usar **criterios lingüísticos**, e**stadísticos**, o **mezcla de ambos**, (Gil Leiva, 1996).

En la **extracción de ocurrencias** los sistemas informáticos han mostrado una enorme eficacia. Pero no olvidemos esta es sólo la **primera parte del proceso de indización**, aunque realmente la mayor parte de los llamados **motores de búsqueda**, se quedan en este primer paso y generan índices a partir de la **simple extracción de palabras**.

El paso siguiente es la **asignación**, que implica, o debería implicar, la **traducción** de las **palabras extraídas** en **términos** que describen el contenido. Los **algoritmos** que se usan para este proceso, llamados en inglés de *assignment indexing* son varios: en algunos casos el sistema posee un **tesauro**, que funciona para el **control terminológico** y que dispone de conjuntos de términos que permiten vincular los términos extraídos con los del lenguaje controlado. En este caso la intervención humana antes o después del proceso de asignación consiste en establecer o validar los criterios de inclusión de los términos y el perfil de las palabras con las que se relacionan.

Como indicamos más arriba, históricamente la indización de palabras extraídas de registros en bases de datos fue el **primer estadio en la indización automática**, en forma **complementaria**

se fueron desarrollando los **sistemas de asignación**, en principio en forma independiente de la extracción como apoyo a la indización manual y luego en forma complementaria a la misma.

Luego la presencia de grandes **masas textuales en la Web** y la difusión de los **lenguajes de marcado** produjo una **modificación en las perspectivas de trabajo**. (Bosch, 2001) A partir de ese momento con la **complementariedad** de formas de **extracción** y **asignación** en ambiente Web, podemos decir que nos encontramos ya frente a lo que se conocen como sistemas asistidos de indización automática, en inglés, **Computer Assisted Indexing systems (CAI).**

4. Sistemas asistidos de indización automática, metodologías e interactividad

La terminología en este punto resulta poco precisa. Para aclararla proponemos observar algunos rasgos: Los **Competer assisted Indexing Systems**, **CAI**, en el uso más difundido del término, se caracterizan por su capacidad de manipular documentos **extrayendo palabras significativas** del título, resumen y palabras claves asignadas por los autores, o del texto del documento, transformando las **palabras extraídas** mediante algún **procedimiento de análisis**, en una **asignación de términos** que pretende **representar el contenido**.

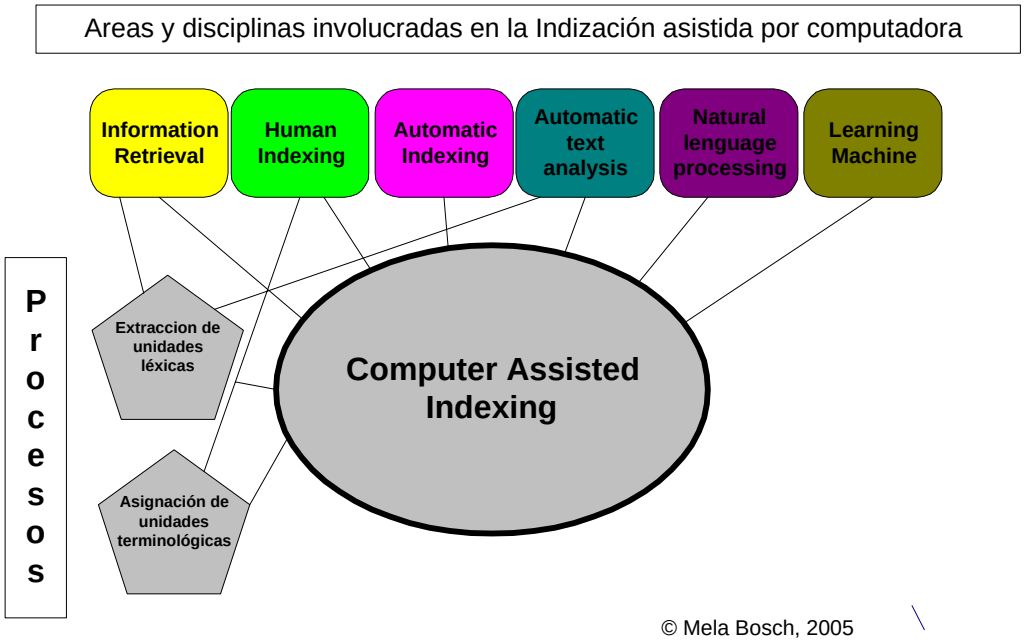
Por tanto, en los CAI hay que considerar cuáles son los procedimientos de **extracción** y de **análisis.** Metodológicamente utilizan diferentes enfoques: Análisis semántico, análisis conceptual con base estadística y análisis en base a reglas de reconocimiento, entre otros. (Nevyjel, 2003).

Otro aspecto importante aparte del enfoque **metodológico para la extracción y el análisis**, es el nivel de **interactividad,** entendido como el grado de mayor o menor intervención humana en los diferentes procesos. En algunos casos el CAI da mucho margen a la **intervención humana**, se limita a sugerir descriptores del lenguaje controlado ordenándolos por relevancia, en estos casos se habla de **ayuda automática a la indización, Computer-aided indexing**, (Browne, 1996). En otros, tenemos una interface de usuario para el aprendizaje interactivo por parte del sistema y la asistencia humana se realiza para apoyar la capacidad de proceso del sistema, estamos en lo que usualmente se denomina **Computer assisted Indexing Systems**, **CAI.**

Con el avance de los sistemas los **limites** entre **Computer-aided indexing** y **Computer assisted Indexing Systems** se han hecho **difusos**. En el primer caso el énfasis pareciera estar puesto en el trabajo humano, y en el segundo en el de los sistemas, encontramos que en los últimos años el **uso** parece haber consagrado la expresión **Computer assisted Indexing Systems, CAI**, para todo sistema más o menos automático que involucre procesos de extracción y análisis de palabras significativas de documentos. Por ello hemos preferido utilizar ese término. Pero creemos que más allá de la estrategia de presentación de los productos la **presencia humana** es en todos los casos **indispensable.**

Y esto es así porque los CAI resultan un **punto de encuentro** de muchos **estudios y desarrollos** en el ámbito de las **Ciencias de la Información** y de la **Informática.**

El gráfico que presentamos realiza una síntesis de estos aportes:



5. Estructuras de representación del lenguaje controlado: El lenguaje en los sistemas informáticos, las taxonomías

Cualquiera sea el nivel de interactividad que se prevea todo **CAI** requiere de un cierto **nivel de implementación y adaptación**. Habitualmente el proveedor del sistema ofrece el soporte de instalación, una categorización temática básica, y un nivel inicial de formación para los usuarios. Luego es necesario una **tarea de refinamiento** que se realiza con los especialistas del dominio temático del sistema.

Tanto wn el momento de la **asignación,** como en el momento del **refinamiento** de la misma, aparece un dispositivo de **lenguaje controlado**, es la **taxonomía.**

El uso de **taxonomías** es un recurso que tiene un **valor de procedimiento**, ya que la posibilidad de **computar las estructuras jerárquicas** ha sido ampliamente estudiada y desarrollada en el mundo de las **Ciencias de la Computación**.

Lo que no está totalmente claro es la **vinculación entre las estructuras jerárquicas los sistemas conceptuales.** La consideración del **conocimiento como el proceso de referencia y comparación de datos sensibles a un modelo abstracto** tiene su origen en **Kant** y fue retomado, aunque con diferencias, por **Peirce**, (Peirce, 1948). La propuesta de este último de aprehender el mundo a través de un **proceso de construcción de símbolos que se basan a su vez en símbolos que ya poseemos** fue aplicada en el ámbito informático ya desde en las primeras teorías y desarrollos de inteligencia artificial.

Allan Newell y Herbert Simon fueron quienes formularon explícitamente la concepción del s**istema informático como un sistema de manipulación de símbolos físicos**. Utilizan la palabra físico para indicar que estos símbolos suponen alguna materialidad en sus componentes, (Newell, A.; Simon, H. 1994).

Según estos autores pioneros de la inteligencia artificial, los **símbolos** deben tener dos características fundamentales para poder funcionar **dentro de un sistema informático**: son la **integridad** y la **conclusión**. Deben ser íntegros porque al ser arbitrarios requieren que se mantengan como tales en su designación mientras no cambie su interpretación y viceversa. La conclusión indica que tienen un inicio y final constantes para el sistema, es decir que tienen siempre el mismo valor.

De manera que cuando operamos símbolos en un sistema informático lo hacemos de una manera particular, se requiere llevar los **símbolos del lenguaje natural**, que son a veces **polisémicos** o fluctuantes, a símbolos físicos con integridad y conclusión, por lo cual el proceso de designación e interpretación debe regirse por los principios constantes y consistentes con un dominio de conocimiento.

Este proceso difiere del que se realiza usando una lengua en uso cotidiano, donde es posible desplegar estrategias de comunicación que van desde lo estilístico a lo gestual. Rápidamente se comprendió que o bien se usaban símbolos matemáticos, o bien se usaba un **lenguaje con un nivel de control y coherencia muy preciso.**

Este tipo de lenguajes se desarrolló en el ámbito de las **ciencias naturales** y en el de la **bibliotecología**.

En ambos casos, aunque para diferentes fines, existe un esfuerzo de **sistematización** que se propone **describir biunívoca mente seres y objetos** a partir de un **conjunto definido de propiedades.**

6. Sistemas asistidos de indización automática y las taxonomías

Como indicamos, los **CAI** hacen uso de **estructuras taxonómicas**. Desde el punto de vista de su implementación una **taxonomía** es una estructura predeterminada que se usa para **dividir un área temática en otras áreas progresivamente más pequeñas**.

Cuando se piensa en el mundo informático la referencia a la estructura en árboles de directorios y archivos salta rápidamente a la mente. Eco nos hace notar que la **diferencia** fundamental entre una estructura jerárquica de un **árbol de directorios** y una **taxonomía** es que la **categoría de mayor nivel imprime sus propiedades a las de los niveles que le suceden**. (Eco, 1997).

La agrupación a partir de propiedades de los conceptos por parte de las taxonomías requiere un trabajo particular de análisis conceptual que es el de determinar los elementos diferenciadores entre los conceptos y a la vez la agrupación de los conceptos que poseen similares características. Este ha sido su gran poder de organizar el conocimiento científico desde su desarrollo en el siglo XVII llegando a las grandes reestructuraciones de las categorías en el siglo XIX. (Perelló, 2002)

Pero debemos tener presente que el uso de las **taxonomías en los CAI <u>no</u>** reúne los requisitos de **organización del conocimiento científico**: *"Un aporte intelectual excepcional de la ciencia*

consiste, justamente, en ordenar de una forma racional, coherente y verificable, la caótica multiplicidad e inagotable diversificación de los procesos y estructuras del mundo real". (García, 2002)

Se suelen generar confusiones al utilizar el término taxonomía, pero debemos hacer una delimitación: Las **clasificaciones bibliográficas**, los **tesauros** y las **taxonomías**, en todos los **sistemas de indización**, desde los de uso manual hasta su aplicación en los CAI tienen en común que se ocupan de representar **no** al mundo real, ***sino a objetos que hablan de ese mundo***. Tiene un valor de **meta representación.**

Es por ello que la información que agregan a los documentos es **metadata**, tanto en el sentido lógico, como en el sentido de la implementación, ya que muchos CAI trabajan utilizando etiquetas de lenguaje de marcado.

También aquí la **terminología** es confusa, en líneas generales se utiliza el término **taxonomía** para indicar un **lenguaje controlado de estructura jerárquica** que **no** tiene una **estructura clasificatoria numérica**, tal como las grandes clasificaciones como la CDU. También puede ocurrir que la taxonomía se ordene numéricamente, pero lo que **no puede cambiar es su valor lógico** de que cada categoría reúne **elementos con un conjunto de propiedades comunes**. Aún en las taxonomías facetadas que están surgiendo en este momento esta estructura lógica se mantiene.

Sin embargo, los usos no son claros y los debates y confusiones son muchas, (Warner, 2004. Edols, 2004) Sin embargo hay otro **rasgo** que parece caracterizar a las **taxonomías para CAI** y es que no responden a una estructura abstracta del mundo, sino a un **específico y determinado dominio**.

Pero aquí también hay diferencias: *"Library catalogers and indexers tend to focus on content when developing taxonomies because that's all they have to work with. Database designers tend to focus on making a single business process more efficient. Journalists tend to focus on user needs and interests."* Montague Institute, 2002.

Para dar una visión más completa de los peculiaridades de este problema es útil mirar otro aspecto fundamental de la información, no sólo su contenido, si no su forma.

7. Información estructurada, información no estructurada: gestión de la información no estructurada

Una distinción tradicional en el mundo de la informática alude a que hay dos tipos de información: la **estructurada** que encontramos en las **bases de datos** y la **no-estructurada** en formatos muy diversos como **texto**, **videos**, **audio** o **imágenes**.

Tal como indica Kugel, (Kugel, 2003) la información no estructurada, *unstructured information* hasta hace poco aparecía en las Bases de Datos como *binary large objects* (BLOBs). Ahora ya se encuentra referida en las bases de datos pero a la vez con su propia entidad. En el ámbito de la **información no estructurada** este autor identifica seis **áreas** diferentes a tratar: **manejo de documentos**, **manejo de contenidos Web**, **manejo de registros**, manejo de **derechos en el ámbito digital**, **control y seguimiento de la colaboración**.

En este momento con el **cambio de escenario** se impone no sólo **indizar y describir** la referencia a los documentos por **medio de registros de información estructurada** sino llegar a **administrar información no-estructurada.** La respuesta es la sinergia de dos tipos de

tecnología: XML y SQL, es decir lenguaje de marcado y la gestión bases de datos, esto permite la representación y administración de información no-estructurada con los recursos de manejo de la información estructurada.

Los documentos se estructuran con el lenguaje de marcado, así el límite entre información estructurada y no estructurada también va siendo progresivamente borrado, ya que no sólo se indican aspectos formales sino que se incluyen marcas con valor semántico, es decir que entramos ya en el campo de la indización temática.

8. Conclusión: las puntas del iceberg

Mientras en el mundo de la **Ciencias de la Información y Documentación** se discutía y analizaba la **calidad y pertinencia** de la **indización automática** respecto de la realizada por **documentalistas o por expertos**, y a la vez se avanzaba en la **calidad de los servicios** de **resumen e indización en bases de datos bibliográficas** (Hawkins, 2001, 2003), la masificación de Internet y su entrada en el mundo de las administraciones empresariales trajo a la luz nuevas necesidades en la indización de documentos.

Cobra ahora fuerza el interés no sólo por las bibliotecas institucionales de publicaciones técnicas, científicas y especializadas, aparece en escena la marea de **la producción de las organizaciones**. Unido a esto encontramos el desarrollo de los recursos informativos en Web: sitios, portales, intranets.

En este contexto Tim Berners Lee, pionero del mundo Internet, plantea lo que considera el futuro: la **Web Semántica**. (Berners Lee, 2001), la posibilidad de que los **documento**s expresen no solamente sus propiedades como objetos digitales: fechas, tipo de archivo, tamaño, dirección, sino también **datos inherentes al contenido**. La forma de realizarlo es a través de la **metadata** y las **prestaciones de base de datos**.

¡No es ninguna novedad! Se escucha decir y con razón, en el ámbito de la documentación y la bibliotecología. Los registros documentales son metadata, se trata de una referencia externa al documento fuente, y deben tener además una normalización precisa, construida a lo largo de años de experiencia en el desarrollo de catálogos.

Pero, ¿cuál es el camino? ¿Normalizar bibliotecológicamente toda la producción de una organización? Y por otra parte, ¿el fondo documental de una organización, no es acaso una tarea de archivistas?
No, -dicen unos- pues se trata de archivos activos, pero, ¿cuál es, en el vertiginoso mundo digital, el lapso que separa un archivo activo de uno intermedio, a otro histórico? Ni antes ni ahora, en la vinculación entre bibliotecarios y archivistas las relaciones han sido fáciles. (Murtomaa, 2001)

Las preguntas son entonces ¿**qué son los documentos digitales de las organizaciones**, de las empresas e instituciones? y ¿cómo representarlos? En este sentido también se hicieron muchos esfuerzos. (Bosch, 2003)

Debemos aún **orientar la formación y la experiencia** en este sentido ya que el vendaval de la *new economy* avanzó sin detenerse a realizar estas disquisiciones y ahora nos encontramos con una realidad:
Considerando que lo que **importa no es sólo la información**, sino la vinculación de de ésta con la historia organizacional y las habilidades acumuladas, es decir, el **conocimiento corporativo**, es en el ámbito de la administración de las organizaciones y desde las empresas de desarrollo de

software para ese mercado, donde encontramos ya operando muchos y eficaces sistemas de tratamiento documental de masas textuales para las organizaciones.

En este contexto creemos que es **importante que los documentalistas seamos protagonistas** de este proceso aportando nuestras capacidades en **organización semántica de los contenidos documentales** para dar la visibilidad a las grandes cantidades de información no estructurada en la Web y en el mundo de las intranets, cuya pobreza actual contrasta vivamente con la **disponibilidad y calidad de la recuperación de la información bibliográfica,** fruto de nuestro trabajo y experiencia**.**

9. Bibliografía

Adams, Katherine C. (2003) Word Wranglers. Automatic classification tools transform enterprise documents from "bags of words" into knowledge resources. [Disponible en http://www.intelligentkm.com/feature/010101/feat1.jhtml?_requestid=80050] Consultado el Marzo de 2005.

Anderson, James D; Pérez-Carballo, J. (2001) The nature of indexing: how humans and machines analyze messages and texts for retrieval. Part II: Machine indexing, and the allocation of human versus machine effort. Pergamon: Information Processing and Management 37 255-277.

Anderson, James D; Pérez-Carballo, J. (2001) The nature of indexing: how humans and machines analyze messages and texts for retrieval. Part I: Research, and the nature of human indexing, Pergamon: Information Processing and Management 37 231-254.

Berners-Lee, T.; Hendler, J.; Lassila, O. (2001) The Semantic Web, A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities Scientific American, May. [Disponible en http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html] Consultado Marzo de 2005.

Bosch, M. (2001) Documentos y lenguaje de marcado: conceptos, problemas y tendencias: El profesional de la información, Barcelona, 10, (11): 4-9.

Bosch, M. (2002)La gestión del conocimiento en el medio digital. Viejos problemas de tratamiento de información y aspectos nuevos: Ciencias de la Información, La Habana, 33, (1):35-43.

Bosch, M. (2003). Modelo conceptual de objetos para la representación y rastreabilidad de documentos digitales. In: Frías, José Antonio; Travieso, Críspulo (eds.): Tendencias de investigación en organización del conocimiento = Trends in knowledge organization research. Salamanca: Universidad de Salamanca, p. 313-320.

Browne, Glenda. (1996) Automatic indexing. Online Currents, the AusSI Newsletter 20(6):4-9, July and LASIE 27(3):58-65

Eco, U. (1997). Kant e l´ornitorinco. Milano: Bompiani.

Eco, U. (2002) Dire quasi la stessa cosa. Milano: Bompiani.

Edols, Liz, Taxonomies are what? (2004) [Disponible en http://www.freepint.com/issues/041001.htm#feature] Consultado Marzo de 2005.

García, Hugo. (2002). La documentación científica. En: Dai, Daniel (coord) Pensar y hacer ciencia. Buenos Aires: Hernandarias.

Gil Leiva, Isidoro; Rodríguez Muñoz, José Vicente. (1996) Tendencias en los sistemas de indización automática. Estudio Evolutivo. En: Revista Española de Documentación Científica, 19 (3).

Gilchrist, Alan (2003) Thesauri, taxonomies and ontologies, an etymological note. Journal of Documentation: Vol. 59 (1): 7-18.

Hawkins, Donald T. (2001) .Information Science Abstracts: Tracking the Literature of Information Science. Part 1: Definition and Map. Journal of the American Society for Information Science and Technology, 52(1): 44–53.

Hawkins, Donald T.; Larson, Signe E.; Caton, Bari Q. (2003) .Information Science Abstracts: Tracking the Literature of Information Science. Part 2: A New Taxonomy for Information Science. Journal of the American Society for Information Science and Technology, 54(8): 771–781.

ISO 1087-1:2000.Terminology work.Vocabulary,Part 1: Theory and application. Technical committee / subcommittee: TC 37/SC 1.

ISO 5127-2:1983 Documentation and information.Vocabulary. Part 2: Traditional documents . Technical committee**:** ISO/TC 46/SC 3**.**

Kugel, Robert D. (2003) Unstructured Information Management. IntelligentBPM. White paper, from Ventana Research. [Disponible en http://www.intelligentbpm.com/feature/2003/12/0312feat2_1.shtml] Consultado Marzo de 2005.

Montague Institute.Ten taxonomy myths, (November 27, 2002). [Disponible en http://www.montague.com/review/myths.shtml] Consultado Marzo de 2005.

Murtomaa, Eeva. (2000,13-18 August)"Podría tratarse del inicio de una buena amistad": comparación de la descripción y acceso a objetos de interés entre bibliotecas y archivos" Jerusalem: IFLA Council and General Conference, 66th.

Nevyjel, Alexander. (2003 September) Computer Assisted Indexing Functionality concept, software evaluation and implementation plan. Salzburg: ODOK'03, International Atomic Energy Agency.

Newell, A.; Simon, H. (1994). 9. La ciencia de la Computación como investigación empírica: Símbolos y Búsqueda. In: Boden, M. (Comp.) La filosofía de la inteligencia artificial. México: FCE.

Peirce, Charles S. (1948) Collected papers. Cambridge: Harvard University Press.
Sebastiani, Frabrizio. (2002?) . Research in automatic text classification, trends and perspectives. Pisa: ISTI-CNR.

Perelló, Javier Gimeno (2002) De las clasificaciones ilustradas al paradigma de la transdisciplinariedad,. [Disponible en: El Catoblepas: http://www.nodulo.org/ec/2002/n010p13.htm] Consultado Marzo de 2005.

Warner Amy J. (2004) A Taxonomy Primer. [Disponible en www.lexonomy.com,.] Consultado Marzo de 2005.

Datos de la autora

Mela Bosch

- Nacida en Santa Fe, Argentina en 1951. Graduada en Letras, realizó los cursos de la Maestría en Ingeniería de Software en la Facultad de Informática de la Universidad Nacional de La Plata. Realizó una especialización en e-learning y virtual community en el Istituto Orga de Milán y una especialización en manejo informático de documentos en empresas en la Universidad Jaume I de Barcelona.
- Actualmente es consultora en tecnologías de la información para Knowledge Management en la empresa Cogent de Milán donde reside. Es docente on line en la Cátedra Tecnologías en Comunicación Social de la Facultad de Periodismo y Comunicación Social de la Universidad Nacional de La Plata, Argentina.
- Como docente universitaria de grado fue Profesora Titular Ordinaria de la Cátedra de Metodología del trabajo intelectual, en el Dpto. de Bibliotecología de la Universidad Nacional de La Plata e Investigadora científica en la Facultad de Periodismo y Comunicación Social de la misma universidad. Enseñó también en la carrera de informática de la Universidad Católica Argentina y dictó de seminarios en el Departamento de Documentación de la Universidad Nacional de Mar del Plata, Argentina, en la modalidad a distancia.
- A nivel posgrado fue docente invitada para cursos en tratamiento automatizado de la documentación, terminología y traducción automática en las universidades: Nacional de Córdoba, La Plata, Mar del Plata y del Nordeste en Argentina, y de Asunción en el Paraguay.
- En el ámbito del Estado argentino fue Coordinadora Alterna de la Subcomisión de Terminología, perteneciente a la Comisión de Sistemas de Información de la Reunión Especializada en Ciencia y Tecnología del MERCOSUR. Coordinó el Proyecto del Catálogo de Bases de Datos en Ciencia Tecnología e Innovación de la República Argentina de la Secretaría de Ciencia y Tecnología de la Nación. Participó en el desarrollo del Módulo de Análisis Documental para el tratamiento automatizado de normativa legal para el MERCOSUR en un proyecto de Banco Mundial y el Sistema Argentino de Informática Jurídica.
- En el ámbito municipal realizó el análisis documental y diseño de la Bases de Datos para la Digitalización del Archivo Histórico de la Ciudad de Buenos Aires.
- Publica en revistas técnicas, foros y congresos de nivel nacional e internacional.