

# La recuperación de la información patrimonial en el marco de la lingüística cognitiva: su aplicación a la construcción de un Macrotesauro del Patrimonio Cultural-Artístico de Extremadura

**Miguel Ángel López Alonso**

Universidad de Extremadura (España)

## Resumen

Se analiza la influencia reciente de la teoría cognitiva y las anteriores aportaciones de la lingüística del texto a la construcción de los lenguajes documentales. Se continúa con los requerimientos mínimos de estos lenguajes para la recuperación de la información en los sistemas de información del patrimonio. Se finaliza con las propuestas que hagan viable el *Macrotesauro del Patrimonio Cultural-Artístico de Extremadura*.

**Palabras clave:** Lingüística textual. Diseño de tesauros ontológicos. Recuperación semántica de la información en la Web. Sistemas de información sobre patrimonio cultural. Extremadura.

## Abstract

Analysis of the theoretical background and practical implications of a project for the development of a macrothesaurus of the Cultural and Artistic Heritage of Extremadura. The recent influence of the cognitive model —and the previous contributions of the Textual Linguistics— in the construction of documentary languages are extensively explored.

**Keywords:** Textual Linguistic. Ontological Thesauri Design. Web Semantic Information Retrieval. Cultural Heritage Information Systems. Extremadura.

## 1. Introducción

La edición digital de documentos en los nuevos soportes informativos permite variar la forma y contenido de estos dinámicamente, de manera que al variar el contenido varíen igualmente los elementos que los describen y relacionan. Por otra parte, la organización y representación de sus contenidos conceptuales avanza

de la mano de “los modelos de organización y representación de la información semántica para la gestión del conocimiento”, mediante la definición formal y estructural de los términos y las relaciones más relevantes; el desarrollo de las referencias cruzadas entre los términos preferentes; y el desarrollo de interfaces capaces de redirigir en tiempo real las oraciones de búsqueda de los usuarios a sus descriptores más adecuados, a partir de las herramientas conceptuales de control terminológico que actúan de intermediarios en el proceso comunicativo entre el lenguaje natural y el lenguaje informático. Por tanto, el actual paradigma de representación de la estructura conceptual y la dinámica de la visualización de estos modelos deberá variar cualitativamente para lograr la modificación o actualización dinámica de sus descripciones conceptuales (por ejemplo, mediante sistemas de conocimientos, redes conceptuales, ontologías, etc.), y la visualización gráfica de sus conceptos en los sistemas integrados de gestión de la información (por ejemplo, mediante mapas conceptuales, *topic maps*, etc.). Entre las técnicas que se utilizan desde los años sesenta con éxito creciente, destaca la utilización de los mismos vocabularios controlados primero en la descripción de los objetos y posteriormente en la formulación de las ecuaciones de búsqueda.

Las aproximaciones tecnológicas de los servicios de recuperación de la información de la Web tienen fuertes limitaciones en cuanto a la precisión de sus resultados, pues la mayoría de los motores de búsqueda tratan de localizar objetos documentales (por ejemplo, páginas web) que contengan las palabras clave de las preguntas del usuario. Pero, por el contrario, las palabras clave del lenguaje natural de los usuarios son un medio muy pobre de recuperación conceptual, debido a su ambigüedad en la captura de la semántica conceptual de las preguntas y de los objetos buscados. En los servidores de la Web la cosa va aún más lejos; los objetos deberían ser previamente “definidos codificadamente” mediante los metadatos: datos sobre los conceptos abarcados, legibles de forma automática, previa normalización semántica de dichos conceptos. El reto ineludible es abarcar el espacio conceptual de la futura Web Semántica, Web inteligente del futuro, denominada también WWG (World Wide Grid), como un “Espacio autonavegable y autocomprensible en el que los conceptos se definan para que sean comprendidos y utilizados automáticamente por las máquinas (o agentes expertos), sin necesidad de intervención humana”. En esta tarea colabora el W3C (World Wide Web Consortium) (2006), que crea los estándares de dicha Web Semántica, como universo de información accesible en red que permita expresarnos en términos que los ordenadores puedan interpretar e intercambiar.

## 2. La teoría cognitiva en la recuperación de la información

El problema clave de la recuperación de la información para Blair (1990) y Soergel (2001) pasa por la búsqueda de los procedimientos teóricos más aproxi-

ados para su representación. Se trata de un problema de uso del lenguaje que, siguiendo la lingüística del texto, se acomete desde el análisis de contenido de los textos. A los problemas teóricos de la sintaxis o la semántica se añade la dificultad de representar el conocimiento pragmático obtenido durante el procesamiento contextual del lenguaje natural de los documentos. La cuestión se complica con la masificación de la información digitalizada que, como sustituto del documento original en formato papel, utiliza simultáneamente textos digitales o soportes multimedia: audio, vídeo, programas informáticos, etc. Su procesamiento implica el escaneado de los textos e imágenes, la comprensión del sonido, la compilación de los programas, etc., para obtener el conocimiento derivado representable en la Web.

Los problemas de la representación codificada de este conocimiento preocupan igualmente a los especialistas en psicología, lingüística, documentación e inteligencia artificial, dentro del ámbito interdisciplinario de la teoría cognitiva. La clave parece ser: ¿cómo incorporar las entidades abstractas de la teoría sintáctica y semántica, y sus relaciones contextuales, al complejo mundo de los sistemas informáticos, de tal forma que participen como intermediarios en el proceso comunicativo de codificación de la información analizada (indización o resumen) y su posterior decodificación para la recuperación del conocimiento contenido, mediante el lenguaje natural utilizado por los usuarios de los sistemas de información? En resumen, lo que nos interesa es el establecimiento de las técnicas y los modelos de representación del conocimiento que, teóricamente (a priori) soporten las necesidades de información, variables y mal definidas por los usuarios.

Como la información se obtiene de las complejas interacciones entre las fuentes de conocimiento disponibles y las representaciones de las necesidades del usuario (Cortez, et al., 1995), este puede ayudarse, para la enunciación de sus preguntas, con la información adicional de una herramienta conceptual del espacio semántico: tesauro, ontología, *topic map*, etc. De esta manera, la recuperación es considerada como un proceso de razonamiento en el que la base de conocimientos de los contenidos de la pregunta y la de los documentos se comparan (espacio documental), para estimar la probabilidad con la que un documento se ajusta a una pregunta. Pero, mientras que en la comunicación lingüística las palabras se clarifican entre sí según el contexto, en las búsquedas documentales la ambigüedad del lenguaje natural de los usuarios es un obstáculo que les conduce a encontrar documentos distintos de los buscados. Esto obliga al establecimiento de los términos de indización de la forma más unívoca posible, para evitar las respuestas parásitas o ruido documental en las recuperaciones.

El proceso de entendimiento del lenguaje natural por los ordenadores es un fenómeno de comunicación recíproca que obliga a profundizar en el estudio previo de los niveles lingüísticos del texto (lógico, semántico, pragmático, etc.), y

en las técnicas de análisis conceptual que conlleven consideraciones pragmáticas (heurísticas o de inferencia), para mejorar la cooperación sistema-usuario.

Esta comunicación presenta dos problemas fundamentales que deberán superarse, para que sea plenamente efectiva: la eliminación de la ambigüedad de dicho lenguaje y el desarrollo de interfaces inteligentes, que trasladen las preguntas desde dicho lenguaje natural (mediante la semiótica del texto) a los lenguajes informáticos (mediante la lingüística computacional) entendidos por el sistema de información.

### **2.1. Integración teórica de la descripción documental**

A mediados del siglo XX se había llegado al establecimiento del “modo correcto” de representar los conceptos, categorías y clasificaciones, como revisión de la posición clásica de Aristóteles. Sin embargo, en las últimas décadas, con el auge de las ciencias cognitivas, las ideas sobre la manera de clasificar el conocimiento han sufrido numerosos ataques, y esa concepción clásica ha sido reemplazada por una concepción natural de los conceptos. Se plantea el problema conceptual de determinar cómo la capacidad humana permite agrupar, de manera unívoca, elementos como miembros de una categoría y luego distinguirla de otras categorías pertenecientes al mismo dominio general de la experiencia. Como resultado ha ido imponiéndose que esa categorización es relativamente arbitraria, y que la cultura determina las entidades o clases que el clasificador humano discrimina.

De la síntesis de los postulados de algunas disciplinas integradas en la teoría cognitiva, la psicología, la antropología y la filosofía, ha ido surgiendo la solución. La categorización no tiene nada de artificial o inmutable, sino que se basa en la información procedente del mundo natural. Las categorías o clases facetadas poseen una estructura interna centrada alrededor de estereotipos o marcas de clase, y los demás elementos están más o menos próximos a estas marcas según el grado en que comparten determinados rasgos comunes con el elemento que da lugar a la clase. Como los esquemas clasificatorios no reflejan todas las relaciones individuales existentes entre los términos, se les compara con los tesauros facetados post-coordinados desde un punto de vista cognitivo. En estos, los campos temáticos se subdividen en clases de materias por agrupación de numerosas subclases de facetas, elegidas por poseer una característica común divisoria. Estos esquemas proporcionan un orden jerárquico a muchos pequeños subconjuntos de conocimientos que se toman como punto de partida para la disposición de los descriptores en un orden determinado dentro de dichos tesauros. Se trata de obtener el mismo grado de detalle con el planteamiento de “abajo-arriba” del análisis terminológico (segmentación conceptual) que con la clasificación de “arriba-abajo” (taxonomía documental).

A partir del análisis estructural del documento primario podremos describir dos niveles del sistema de clasificación: una macroestructura global constituida

por un sistema de conocimiento determinado y otra microestructura variable según el modelo del sistema de clasificación (por ejemplo, enumerativo, facetado, etc.) y de sus correspondientes códigos de notación (Pinto y Gálvez, 1987). Por tanto, durante el análisis documental la clasificación opera con objeto de organizar el conocimiento de acuerdo con un esquema clasificatorio conceptual y unificar las partes del documento, situándose al nivel más sintético posible, mientras que la indización y el resumen documental adoptan una perspectiva analítica o de descripción frente al documento a nivel macroestructural. Los sistemas de información documental han recurrido a la utilización de los descriptores de los tesauros de un área del conocimiento para indizar los documentos por sus títulos o resúmenes; y, tras la ponderación de los conceptos más relevantes abarcados por estos, los han vuelto a utilizar como descriptores para la recuperación documental, eliminando con ello el ruido documental o la falta de precisión derivados de la tradicional utilización de las categorías clasificatorias como elemento de acceso en las recuperaciones documentales. A partir de dichas características comunes entre ambos lenguajes documentales —los sistemas de clasificación y los tesauros—, diferentes profesionales de la información como Chan (1995), Vizine-Goetz (1996) o McKiernan (1997) señalaron las ventajas de utilizar las clasificaciones jerárquicas universales para organizar el universo de Internet. Encadenándolas con tesauros integrados por un amplio número de relaciones conceptuales en un espacio terminológico de redes semánticas, investigadores como Kohonen o Xia Lin (Zavrel, 1996) empezaron a desvelar sus analogías con el ámbito cognitivo del mundo desarrollado por el ser humano.

Los primeros resultados de la utilización de los tesauros conceptuales en la expansión de las ecuaciones de búsqueda de los usuarios en grandes bases de datos documentales periódicas indicaron un aumento de la exhaustividad de las recuperaciones, pero no así de la precisión (Kristensen, 1993). Resultados más recientes, que utilizan la inferencia de las redes bayesianas junto con dichos tesauros en un modelo de red extendido, e incluyen la combinación de los descriptores con otras cinco fuentes de evidencia —notas de alcance, descriptores más generales y más específicos, conceptos sinónimos y descriptores relacionados, en las búsquedas, en una base de datos específica del ámbito jurídico—, han aportado la evidencia de ganancias mantenidas de precisión de hasta el 30% (Silveira y Ribeiro-Neto, 2004).

### **3. La lingüística del texto aplicada a la construcción de lenguajes documentales conceptuales**

En un sistema del lenguaje, las bases teóricas de las distinciones dicotómicas entre lengua-habla y eje paradigmático-eje sintagmático fueron desarrolladas por Sausure (1985). De la articulación de ambos ejes se entiende el proceso de cómo

un sistema conceptual o lenguaje genera sistemas físicos, los textos en lenguaje natural, de cómo a partir de estos textos se genera un sistema conceptual, el sistema de clasificación, y de cómo este sistema genera un sistema físico o notación clasificatoria. El antecedente inmediato del análisis textual desde la perspectiva lingüística se encuentra en las investigaciones sobre el discurso de Zellig Harris a comienzos de los años sesenta (1959-1961). En España, Gili Gaya (1969) mencionaba en su *Curso superior de sintaxis española* que

La unidad total del discurso, a la cual sirven las oraciones que lo componen, obedece a leyes psicológicas, y según ellas percibe el oyente o el lector la coherencia o incoherencia del discurso que se le dirige.

De lo anterior cabe afirmar que la lingüística del texto es un instrumento enriquecedor para el análisis documental como análisis de contenido de los textos. El análisis documental es un modelo de análisis textual de características propias, impuestas por las reglas de la ciencia de la documentación; esto es, el objetivo final de esta técnica analítica es la difusión de la información previamente analizada. En muchos aspectos, el análisis documental ofrece un modelo de análisis y una metodología más completos que otros tipos de análisis, debido a la propia finalidad del análisis (el contenido textual); el tipo de textos que analiza (los documentos informativos o texto documental); y la oferta de unos elementos sumamente claros de análisis, con una metodología precisa (Bernárdez, 1982).

En cuanto a esta metodología, se pueden establecer unos puntos comunes a las múltiples definiciones que las distintas corrientes de la lingüística del texto han generado, a saber: la existencia de dos niveles lingüísticos en el texto: estructura superficial o gramatical (morfológica o sintáctica) y estructura profunda o textual (semántica); la integración de los elementos pragmáticos, extralingüísticos, o sociales del texto; y la base textual como planificación previa del texto, que el autor o emisor realiza antes de comenzar a elaborar el texto.

A partir de la consideración del texto documental como un nuevo tipo de texto, la lingüística del texto tiene que incluirlo en su tipología, si bien es cierto que añadiendo un tercer nivel en el análisis. Al propio nivel lingüístico y al pragmático habrá de unir el nivel de análisis documental, semántico o conceptual, si se pretende difundir el texto y que se pueda recuperar por el receptor o usuario. La visión del estudio de los textos como estructura unitaria dentro del marco de la lingüística nos los muestra como conjuntos de unidades lingüísticas que, dentro de unos límites contextuales, contienen objetos semióticos con unas especificaciones culturales, y vienen caracterizados como entes dinámicos.

El análisis del discurso se incorpora así a la descripción sustancial de los documentos, en la búsqueda de unos criterios objetivos que nos permitan identificar las estructuras que soportan las informaciones más relevantes dentro de

los textos. No podemos olvidar que actuamos sobre discursos de manera global, pues nos interesa la explicación fundamental del contenido de un documento y no sus enunciados aislados. En consecuencia, nuestra labor se centra en identificar la organización metodológica del discurso, y encontrar las informaciones relevantes desde una perspectiva documental. Debe constarnos de forma explícita la articulación del texto en torno a una gramática que implica factores estructurales, contextuales y textuales.

El discurso científico se establece como marco desde el que se comprende a la ciencia de la documentación. Se compone aquel de todos los contenidos científicos conectados al macrotexto del conocimiento, las grandes teorías generales, además de una división en ámbitos disciplinares diferenciados por una conceptualidad y metodología propias. Encuadrado en él se da el producto científico concreto, el texto-discurso, resultado de una investigación que fija el contenido en un documento en función de su aportación al macrodiscurso de la ciencia. Desde este universo documental valoramos su aportación de noticias respecto a una situación dada (Moreiro, 1992). De esta manera el discurso se fragmenta en unidades documentales individuales, la relación de cuyos contenidos pertenece a un campo especializado desde el que se comprende tanto la forma compositiva como el acceso a los datos. Este proceso no se completa hasta que no sea consultado el mensaje documental. Recordemos los ruidos e interferencias que acontecen en la transmisión de los mensajes.

Precisamente los textos, como recursos sobre los que se produce la información científica, vienen estudiados por el factor más genuino de la lingüística del texto: su atención a lo pragmático o entorno extensivo, por la que atiende al intercambio comunicativo entre emisor y receptor a través del discurso, cómo se relacionan ambos entre sí y el contexto. Para que el proceso textual sea completo, el mensaje tiene que ponerse en disposición de consulta. Tiene que llegar al receptor y ser captado por este. Para que esto acontezca hay que sortear previamente unas dificultades que, a modo de ruidos e interferencias en el acto comunicativo, impiden el acceso directo del usuario al mensaje documental.

La fuente y el destino informativos, por lo general separados en tiempo y espacio, no logran que el canal transmita los significados deseados. Los obstáculos forman conceptos ya clásicos en la literatura documental: el número, la dispersión, la obsolescencia, la creciente especialización se han establecido como hechos problemáticos en la transmisión del saber. Los textos informativos ponen en comunicación a los individuos que integran la comunidad científica. Los textos son el objeto de la comunicación científica. La actuación lingüística sucede dentro de un contexto informativo y comunicativo. El hecho documental es por ello un acto de comunicación lingüística.

De la necesidad de “consulta personal” del mensaje documental para que se complete el proceso comunicativo surge el concepto de la información como algo subjetivo e individualizado (Cole, 1994), que forma parte del proceso pragmático extensivo que sigue cualquier usuario en la relación con su entorno. Resalta, pues, lo limitado del enfoque clásico centrado en el documentalista como un intermediario y en la información como un objeto físico, y la necesidad de enfocar el estudio del fenómeno informativo desde enfoques interactivos centrados en el usuario final (Fox et al., 1993). (Teorías del *common sense* de Olaisen, 1996, o del *sense making* de Soergel, 1997, entre otras).

#### **4. Requerimientos mínimos de un lenguaje documental para su utilización en los sistemas de información del patrimonio cultural**

Históricamente, a partir del reconocimiento internacional de la necesidad de una normalización de los diferentes lenguajes documentales con vistas a la indización y la posterior recuperación de la información documental contenida en las bases de datos. Las diferentes organizaciones internacionales (UNESCO, OCDE, OIT...) fueron desarrollando unas terminologías básicas que les permitieron establecer un “eje conceptual integrador” con los tesauros especializados, desarrollados en diferentes entornos documentales a partir de los años cincuenta. Recientemente, con el crecimiento acelerado de la Web y la explosión de los recursos informativos almacenados en ella, se vuelve a pensar en los tesauros (ahora conceptuales) como herramienta de precisión para la recuperación del conocimiento contenido en la información de dichos recursos.

Como innovación cabe señalar que el tesauro conceptual propuesto difiere plenamente de sus antecesores los índices jerárquicos en que su compilación se hace a partir del sublenguaje científico de bases de conocimientos de dominios específicos concretos (por ejemplo, cultura e historia de Extremadura); y sus relaciones entre términos (jerárquicas, de equivalencia y asociativas) se diseñan para ser utilizadas preferentemente en la recuperación de documentos a texto completo, en bases de datos documentales que no han sido previamente indizadas con ningún otro lenguaje controlado (por ejemplo, las contenidas en los servidores automatizados de Internet o en las inmensas bases de datos periodísticas).

Entre los tesauros similares realizados por grupos españoles, destaca el Tesauro del Centro de Documentación del Patrimonio Histórico Andaluz, que ha integrado las distintas disciplinas relacionadas con el patrimonio histórico en un macrotesauro. Este se ha adoptado como lenguaje documental del Sistema de Información del Patrimonio Histórico de Andalucía, al ser terminado en 1998 bajo la dirección del Dr. Antonio García Gutiérrez (2006), catedrático de la Facultad de Ciencias de la Información de la Universidad de Sevilla.



Igualmente, a un nivel internacional, cabe destacar el Getty Vocabulary Program del Getty Research Institute, que elabora, mantiene y difunde diversos tesauros para las artes visuales, arquitectura, archivos de arte y materiales culturales. Entre ellos destaca el Art & Architecture Thesaurus, diseñado para ser utilizado tanto de herramienta conceptual para la catalogación e indización documental como de ayuda a la recuperación de la información en sistemas locales, a partir de su consulta directa online.

Un lenguaje documental, para cumplir con las funciones propias de un sistema de información sobre patrimonio cultural, debe reunir ciertas características que potencien el cumplimiento de su misión como instrumento terminológico organizador del patrimonio. Entre ellas, se da más importancia a las que refuerzan la relación o interconexión documental en este ámbito, pues las principales tareas de estas herramientas son facilitar la integración de las disciplinas relacionadas con el patrimonio entre sí y, facilitar el control terminológico de un amplio conjunto formado por todos los objetos, documentos, imágenes y tradiciones que conforman nuestro patrimonio.

Por ello, como comenta Teresa Abejón Peña (2006), un lenguaje documental adecuado a sistemas de información sobre patrimonio debe proporcionar términos con un alto nivel de precisión y especificidad. Debe tenerse en cuenta especialmente lo relativo a objetos, procesos y técnicas, ya que estos aspectos implican una variedad terminológica considerable que hay que controlar, además de matices locales o jergales. Por ello, se debe atender particularmente la denominación de los objetos porque estos constituyen una gran parte del patrimonio (patrimonio material), y porque este campo está ampliándose continuamente, y por ello exige mucho trabajo de mantenimiento. En segundo lugar, como ya decíamos, debe ser adecuado para una gran variedad de aplicaciones, y poder ser utilizado para ejercer control terminológico sobre una gran cantidad de objetos documentales como material bibliográfico, archivos, recursos visuales, y también para indizar y resumir información sobre muchas áreas y disciplinas (antropología, arqueología, arte, cultura, etc.).

Es importante establecer diferencias entre los vocabularios desarrollados para catalogar exclusivamente objetos documentales en sus distintos soportes y los diseñados para el control de todo tipo de materiales (objetos de museo, imágenes, bienes culturales...). Y es que, si ese lenguaje documental debe ser capaz de interconectar todos los sistemas de información existentes sobre patrimonio, se deberá crear teniendo en cuenta que muchos de ellos están altamente especializados. Derivado de lo anterior, por ello, es importante que sea flexible y adaptable; pues un lenguaje documental demasiado restringido en sus normas o en su jerarquía plantearía dificultades de implantación en algunos de los sistemas de información en los que se pretendiera introducir, la mayoría de ellos con

características documentales y planteamientos informáticos muy variados. En cambio, si es flexible y abierto se facilita la relación entre estos sistemas.

Otra de las condiciones que se deberán considerar en la elaboración de lenguaje documental relativo a patrimonio es, por ejemplo, la utilidad de combinar conceptos para describir los objetos del patrimonio artístico en cuanto a sus aspectos técnicos y estilísticos.

En resumen, un lenguaje documental, para que pueda ser capaz de relacionar sistemas de información sobre el patrimonio, debe ser integrador y servir para toda la gran variedad de aspectos del ámbito del patrimonio; no limitado a la catalogación de solo objetos de museo o solo documentos, debe poder catalogar cualquier tipo de material patrimonial material o inmaterial; ser suficientemente amplio; y derivar su utilidad como instrumento terminológico idóneo del cumplimiento de los estándares de los modernos tesauros conceptuales.

## 5. Conclusiones

Las conclusiones de este estudio son tanto de carácter teórico como empírico. En el plano teórico de aplicación de los principios de la lingüística cognitiva al desarrollo de la lingüística computacional destacan cuatro conclusiones. En primer lugar, se precisa el desarrollo de una nueva generación de sistemas de gestión de la información que identifique las relaciones contextuales entre los conceptos y permita obtener la mayor independencia posible del área del conocimiento abarcada. Estos objetivos se alcanzan más fácilmente con un grado mayor de integración del tesoro en el sistema de gestión de la información, ya que la facilidad de actualización del tesoro depende, en gran parte, de las funciones abarcadas por el gestor de tesauros integrado en dicho sistema.

En segundo lugar, se deben potenciar los sistemas de gestión de la información originados en el campo de la inteligencia artificial, pues tienen capacidad de conocimiento para analizar la información recibida de las ecuaciones de búsqueda de los usuarios (PLN, procesamiento lenguaje natural) y reutilizarla, cruzándola con la información terminológica de los distintos corpus lingüísticos disponibles en la futura Web Semántica, como realimentación de las nuevas búsquedas.

En tercer lugar, se necesita un modelo conceptual integrado de la información que tome como punto de partida la “interacción total” entre los documentos y las preguntas, los considere como entidades del mismo tipo y normalice sus conceptos. Para obtener un espacio conceptual que refleje las relaciones asociativas más estrechas e ignore las menos relevantes, se debe cuantificar el grado de solapamiento entre conceptos (*clustering matching process*) y tomar aquellos que incrementen la relevancia de cada recuperación.

En cuarto lugar, se propone que, cuando se detecte una “relación global” entre la pregunta del usuario y un documento de la base de datos, se analicen sus oraciones principales buscando entidades de argumentación —bloques básicos de conocimientos, marcados previamente por el indizador, o descriptores seleccionados por el usuario en su argumentación— para ser reutilizados en la enunciación de ecuaciones de búsqueda más ajustadas que lleven a una recuperación potente.

Finalmente, para relacionar los esquemas mentales del conocimiento humano con el razonamiento formal silogístico de los programas informáticos, se debe tomar como modelo teórico la analogía entre el razonamiento argumental y el proceso de recuperación del conocimiento documental.

En el plano empírico de la construcción del Macrotesauro del Patrimonio Cultural-Artístico de Extremadura, es preciso estudiar su interrelación con otros vocabularios utilizados en los incipientes sistemas de información patrimonial de las diferentes Comunidades Autónomas y de los demás países de habla hispana, y su interrelación con el Tesoro de la Herencia Cultural Europea (European Heritage (<http://www.european-heritage.net/en/Thesaurus/Contenu.html>)); delimitar su conexión con los vocabularios internos de la mayoría de los museos estatales españoles, ciñéndolos a áreas muy específicas en una primera fase: mobiliario, utensilios, etc., y su conexión con los vocabularios propios de cada sistema de información en bibliotecas, centros de documentación y archivos visuales especializados en patrimonio; y, finalmente, abordar su conversión en un tesoro multilingüe que abarque las distintas lenguas nacionales y la portuguesa, etc. Terminamos haciendo nuestra la afirmación de Teresa Abejón Peña:

El intercambio de información entre una colectividad, en la época de Internet, es algo común y casi imprescindible. Por ello, no debemos trabajar con códigos diferentes según se trate de materiales y centros diferentes (bibliotecas, fondos de museo, archivos)...; porque si eso, en un solo centro, ya produce confusión y pérdida de efectividad, no digamos si se trata del intercambio de información entre varios centros, de varias tipologías además: institutos, servicios de patrimonio, museos, bibliotecas especializadas u otras instituciones culturales. Se debe adoptar un instrumento versátil y lo más universal (por tanto, flexible) posible.

## Referencias

- Abejón Peña, T. (1999). Un instrumento versátil e idóneo para la interconexión de sistemas de información: Tesoro de Patrimonio Histórico Andaluz. URL: <[http://www.juntadeandalucia.es/cultura/iaph/portal/Productos/BasesDatos/Tesoro\\_PHA/dossier7art1.html](http://www.juntadeandalucia.es/cultura/iaph/portal/Productos/BasesDatos/Tesoro_PHA/dossier7art1.html)>. Consultado: 2006-09-30.
- Bernárdez, E. (1982). Introducción a la lingüística del texto. Madrid: Espasa, 1982.
- Blair, D. C. (1990). Language and Representation in Information Retrieval. Ámsterdam: Elsevier, 1990.

- Chan, L. M. (1995). Classification, present and future. // *Cataloging & Classification Quarterly*. 21:2 (1995) 5-17.
- Cole, Ch. (1994). Operationalizing the notion of information as a subjective construct. // *JASIS*. 45:7 (1994) 465-476.
- Cortez, E. M.; et al. (1995). The hybrid application of an inductive learning method and a neural network for intelligent information retrieval. // *Information Processing and Management*. 31:6 (1995) 789-813.
- Fox, E.; et al. (1993). Users, user interfaces, and objects: Envision, a digital library. // *JASIS*. 44:5 (1993) 480-491.
- García Gutiérrez, A. (1999). Itinerarios de representación del conocimiento en la sociedad de la información: bases para la construcción epistemológica del Patrimonio Histórico Andaluz. URL: <[http://www.juntadeandalucia.es/cultura/iaph/portal/Productos/BasesDatos/Tesouro\\_PHA/dossier7art1.html](http://www.juntadeandalucia.es/cultura/iaph/portal/Productos/BasesDatos/Tesouro_PHA/dossier7art1.html)>. Consultado: 2006-09-30.
- Gili Gaya, S. (1969). *Curso superior de sintaxis española*. (3ª ed.). Barcelona: Bibliograf, 1969.
- Harris, Z. S.; et al. (1959-61). *Transformations and discourse analysis projects*. Dept. Linguistics, Phi.: Univ. Pennsylvania, 1961.
- Kristensen, J. (1993). Expanded End-user's Query statements for free text searching with a search-aid thesaurus. // *Information Processing & Management*. 29:6 (1993) 733-4.
- McKiernan, G. (1997). The new/old World Wide Web order: the application of "neo-conventional" functionality to facilitate access and use of a WWW database of science and technology Internet resources. // *Journal of Internet Cataloging*. 1:1 (1997) 47-55.
- Moreiro, J.A. (1992). Consideraciones acerca de la intermediación discursiva de la documentación en la ciencia. // *BILDUMA*. 6 (1992) 147-153.
- Olaisen, J. (1996). *Personal Communication*. Copenhagen, 1996.
- Pinto, M.; Gálvez, C. (1987). *Hacia una teoría integradora de la clasificación documental*. // *Manual de Clasificación Documental*. Madrid: Síntesis, 1987. 33-48.
- Sausure, F. de. (1985). *Curso de lingüística general*. París: Fayot, 1916. Barcelona: Planeta-Agostini, 1985.
- Silveira, M. L.; Ribeiro-Neto, B. (2004). Concept-based ranking: a case study in the juridical domain. // *Information Processing and Management*. 40:6 (2004) 791-805.
- Soergel, D. (1997). *An Information Science Manifesto*. // *ASIS Annual Meeting Coverage* (Dec. 1997).
- Soergel, D. (2001). *ASIST SIG/CR Classification Workshop 2000: Classification for User Support and Learning*. // *Bulletin of ASIST*. 27:4 (May 2001).
- Vizine-Goetz, D. (1996). *Using library classification schemes for Internet resources*. // *Proceedings of the OCLC Internet Cataloging Colloquium*; Dublín, Ohio: OCLC.
- World Wide Web Consortium (2006). URL: <<http://www.w3.org/Consortium>>. Consultado: 2006-09-30.
- Zavrel, J. (1996). Neural Navigation Interfaces for Information Retrieval: Are They More Than an Appealing Idea? // *Artificial Intelligence Review*. 10:5-6 (1996) 477-504.