

POTENCIALIDAD EVOLUTIVA DEL TESAURO: HACIA UNA BASE DE CONOCIMIENTO EXPERTO

María José López-Huertas
Facultad de Biblioteconomía y Documentación
Universidad de Granada

Resumen

Las prestaciones de los tesauros tradicionales en los actuales sistemas de información se reducen a ser uno de sus recursos léxicos y a dirigir vagamente las búsquedas a través de los dispositivos semánticos previstos por tales lenguajes como son los de sugerir búsquedas más genéricas, más específicas o más amplias. La evolución del tesoro tradicional hacia una estructura de gran capacidad para la representación conceptual y para la identificación de numerosas relaciones conceptuales apunta a la posibilidad de poder convertirse en una base de conocimiento experto para la recuperación de la información en un dominio temático determinado.

Palabras clave. Tesauros, recuperación de la información, sistemas de información especializados, estructuras conceptuales, bases de conocimiento, sistemas expertos.

Abstract

Traditional thesauri in new information systems are used as a lexical resource and as a tool for searching by suggesting either a more generic, a more specific or a wider search. The evolution of the traditional thesaurus towards a structure with a high capacity for concept representation and for conceptual relations identification points to the possibility for the thesaurus to become a knowledge base for intelligent information retrieval in specialized domains.

Key words: Thesauri, information retrieval, specialized information systems searches, conceptual structures, knowledge bases, expert systems.

0. INTRODUCCIÓN

Nuevas tendencias en el diseño y construcción de tesauros han puesto de manifiesto que este lenguaje, tal y como se ha venido concibiendo desde décadas pasadas hasta la actualidad, está mínimamente explotado [1]. Podría afirmarse incluso que su potencialidad está poco explorada. Ésta puede ser una de las razones por la que sus prestaciones en los modernos sistemas de información están bastante mermadas, adoleciendo, por regla general, de la eficacia que podría suponersele a una herramienta de estas características. Estas disfunciones han venido siendo denunciadas por algunos especialistas desde los años ochenta [2], [3].

Estudios recientes han demostrado, por un lado, la posible conexión del tesoro con metodologías de la inteligencia artificial para la recuperación de la información [4]. Por otro lado, la utilización de nuevos modelos para el diseño de tesauros basados en modelos cognitivos de los sistemas de información y el uso de métodos poco convencionales hasta ahora para la elaboración de las estructuras conceptuales en las que se apoyan estos lenguajes han permitido la creación de unas estructuras mucho más ricas, más representativas del dominio temático al que se refieren y más dinámicas que las tradicionales partes clasificadas de los tesauros. El resultado de estos trabajos ha sido la obtención de una potente herramienta para la recuperación de la información que tiene más similitudes con las bases de conocimiento que utilizan los sistemas expertos que con la clásica idea de tesoro [5]. El tesoro clásico y el "evolucionado" comparten el hecho de tener una base terminológica, pero presentan profundas diferencias en lo que se refiere a la estructura conceptual que soporta a uno y otro y a las relaciones conceptuales que pueden ser generadas a partir de dicha estructura en uno y otro caso. Si entendemos que la estructura conceptual es la propuesta del sistema de indización, de este lenguaje, para la representación y la organización del conocimiento de un ámbito temático determinado y la que facilita de forma decisiva la recuperación de la información, podría decirse que estamos ante el proceso de creación de un mecanismo o de una herramienta de recuperación de la información que presenta grandes diferencias con el tesoro tradicional y que tiene ciertamente concomitancias con las bases de

conocimiento experto. El objetivo de este trabajo será demostrar cómo es posible desarrollar la capacidad de representación y organización del conocimiento en una estructura conceptual de un tesoro, la incidencia que este hecho tiene en la identificación de numerosas relaciones conceptuales exclusivas de un campo temático determinado y se verá cómo éstas pueden hacer que un sistema de recuperación de la información funcione de forma parecida a como lo hacen los sistemas expertos.

1. MÉTODOS PARA LA ELABORACIÓN DE LA ESTRUCTURA

Una estructura que tenga las características necesarias para permitir una búsqueda y recuperación inteligente de información ha de basarse en un diseño y en unos métodos de elaboración que hagan posible esta funcionalidad y que nada tienen que ver con los tradicionales supuestos o métodos de elaboración de tesoros. Los fundamentos conceptuales y la metodología de construcción son aspectos cruciales del proceso. Las bases teóricas en las que se inspira el nuevo diseño hay que buscarlas en los modelos de sistemas de información propuestos por los teóricos del entorno cognitivo en los que textos y usuarios son los pilares del sistema, entendidos los primeros como depósitos de conocimiento público que está formado por conocimiento objetivo básicamente, pero también por una serie de otros elementos entre los que se pueden citar la intencionalidad del autor, sus objetivos, etc., y los segundos, los usuarios, representados por sus necesidades de información que también son la suma del tema requerido y de otros aspectos, como pueden ser la tarea concreta a realizar, objetivos, etc.

En trabajos recientes [5], se ha podido demostrar cómo la estructura puede crearse directamente a partir de estos parámetros, de forma que esté en sintonía con el sistema, que sea parte del mismo, y no un elemento impuesto. El referido trabajo se centra sólo en textos especializados para demostrar que es posible reproducir o trasladar la estructura de éstos a la estructura conceptual del tesoro. Para ello, se parte de un grupo de descriptores ya seleccionados previamente, representantes del campo temático elegido que en el ejemplo son los instrumentos musicales de cuerda, y se analizan las definiciones que sobre esas expresiones dan los mejores diccionarios de la materia. El primer resultado de este análisis es la identificación de características relevantes para el concepto definido, expresadas en el propio texto, bajo las cuales es representado posteriormente en forma de categoría. Para organizar las categorías aisladas en la estructura se sigue un criterio de relevancia, entendido éste como la frecuencia con que son citadas las categorías en textos analizados, lo que, al mismo tiempo, incorpora la importancia que el autor de esos textos les asigna. Un ejemplo resumido puede verse en la fig.1.

1.1. Representación conceptual

Una estructura como la descrita permite que todas las características relevantes de cada uno de los conceptos que van a ser representados tengan entrada y representación en la misma, lo que significa que se podrá llegar al concepto desde cada uno de los aspectos que los autores de los textos hayan considerado relevantes. Se produce así un fenómeno de polirrepresentación conceptual que significa poder representar un concepto en una estructura tantas veces como sea relevante para él según los textos, proporcionando en consecuencia una representación exhaustiva para cada concepto. Este importante aspecto de la estructura concentra características que otros autores han considerado necesarias en un sistema de recuperación de información como son el principio de variedad [1] y el principio de redundancia [1], al tiempo que sustenta un principio enunciado por Bates, *hit side of the barn*, [6]. Como resultado de las cualidades descritas, es de esperar que una estructura así eleve el nivel de exhaustividad en la recuperación de la información, que es uno de los problemas más evidentes de los sistemas de información tanto tradicionales como no convencionales sin perder por ello precisión.

Para ilustrar un poco el tipo de representación obtenida, se puede decir, volviendo al ejemplo de los instrumentos musicales de cuerda, que cada uno de ellos estará representado al menos bajo las categorías *fuerza del sonido*, *características físicas*, *funciones* y *materia* si mencionamos sólo las más representativas. Según descendemos en cada una de ellas, la representación conceptual aumenta. Un determinado instrumento, por ejemplo el **violoncelo**, está representado directa o indirectamente, al menos, en 26 categorías según se muestra en la fig. 2.

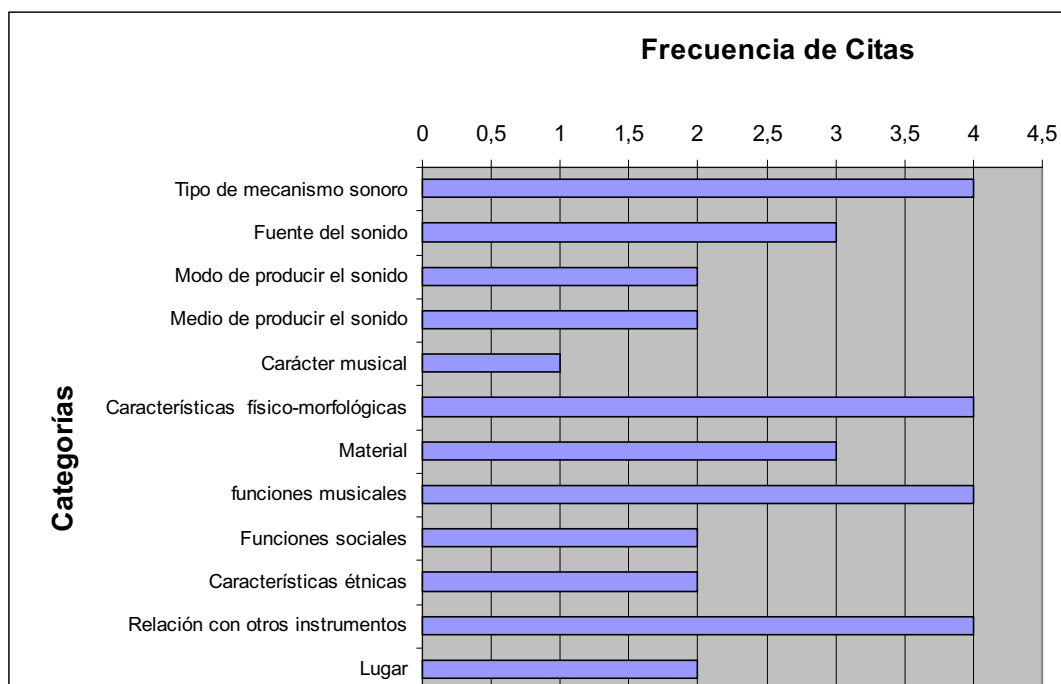
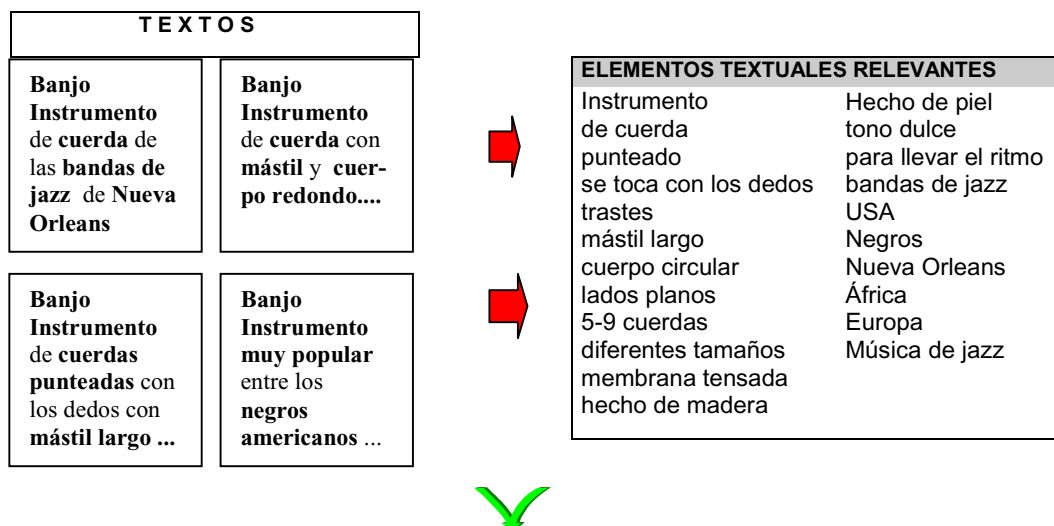
La polirrepresentación conceptual tiene dos importantes consecuencias:

- 1) permite la identificación de temas secundarios fundamentales para la articulación conceptual del tema que se quiere representar, con lo que se asegura que el campo semántico total para nuestro tema; es decir, núcleo temático y temas auxiliares, está

- completo o al menos dotado de los más significativos, según los especialistas en la materia.
- 2) favorece el establecimiento de relaciones conceptuales mucho más numerosas que las que aparecen en el tesoro tradicional y factibles de ser identificadas individualmente.

1.2. Sistema relacional

Los tesauros tradicionales presentan unas relaciones conceptuales entre los distintos descriptores de su vocabulario que son a todas luces insuficientes y poco precisas. Esta situación, aunque denunciada desde finales de los ochenta [2], [3], ha tenido algunas respuestas recientes por parte de especialistas enfocadas a profundizar en las clases de relaciones conceptuales que se esconden bajo las comúnmente llamadas jerárquicas [7], qué otras relaciones han de proponerse [8] o qué tipo de relaciones son necesarias para la búsqueda de información [9]. El método descrito para la elaboración de esta nueva estructura lleva a la identificación de abundantes y pertinentes relaciones conceptuales para el campo temático que se está trabajando. Por ejemplo, para el campo relacionado con los instrumentos de cuerda, para el que sólo hay dos tipos de relaciones estructurales en todos los tesoro al uso: las jerárquicas y las asociativas, se han aislado hasta 21 clases de posibles relaciones para cada concepto a partir del estudio de 10 instrumentos de cuerda: *procedencias del sonido, formas de producir el sonido, medios de producir el sonido, carácter musical, técnicas de interpretación, géneros que interpreta, formas que interpreta, orquestas o conjuntos relacionados, descripciones físico-morfológicas, relaciones con otros instrumentos, métodos de aprendizaje, escuelas, compositores relacionados, intérpretes relacionados, artesanos relacionados, material del que está hecho, grupos étnicos relacionados, grupos religiosos/militares relacionados, composiciones, lugar y tiempo.*





ESTRUCTURA
Tipo de mecanismo sonoro/solo
Fuente del sonido
Forma de producir el sonido
Medio de producir el sonido
Carácter musical
Técnicas de colocación/interpretación
Funciones
Funciones musicales
Funciones musicales/géneros
Funciones musicales/orquestas
Funciones sociales
Instrumentos populares
Material
Tipo de mecanismo sonoro/orquestas
Géneros musicales
Formas musicales
Características étnicas
Lugar
Tiempo

Figura 1. *Modelo para la representación y organización conceptual a partir de textos especializados*

Esta exhaustiva identificación de relaciones junto con la polirrepresentación conceptual con la que está muy relacionada porque es su consecuencia inmediata va a hacer posible que el usuario tenga una recuperación de información asistida de forma inteligente.

El ejemplo de la figura 2 muestra cómo estaría representado en la estructura cada uno de los conceptos que, a su vez, representan un ámbito temático especializado. Se produce un fenómeno de polirrepresentación altamente pertinente para cada uno de los conceptos que va a suponer un paso hacia adelante en la configuración de estructuras inteligentes, a partir de las identificadas hasta ahora como las partes sistemáticas de los tesauros.

2. RESULTADOS: HACIA UNA BASE DE CONOCIMIENTO EXPERTO.

Una de las características más importantes de estructuras como la aquí descrita es que se alcanza una gran exhaustividad y pertinencia tanto en la representación individualizada de conceptos como en las dinámicas relacionales entre ellos. La consecuencia más destacada de esta potencialidad es que el sistema realmente "conoce" toda la información relevante referida a cada concepto y a sus relaciones con otros de su entorno. Estaríamos, por tanto, hablando de un sistema que presenta bastantes concomitancias funcionales con las bases de conocimiento experto e incluso con lo que llamamos sistemas expertos.

Las características más importantes de un sistema experto podría decirse que son su especialización temática y la posibilidad de facilitar la toma de decisiones o resolver los problemas concretos planteados. Para ello, el sistema debe anticiparse, prever esas posibles cuestiones para presentar alternativas o formas de resolver el problema o la pregunta dirigida al sistema. Es decir, debe conocer exhaustiva y pertinentemente, lo mismo que un experto, el ámbito temático para el que se crea y su potencialidad relacional, características ambas presentes en la estructura que se acaba de describir.

2.1 RECUPERACIÓN INTELIGENTE DE LA INFORMACIÓN

¿Qué clase de sistema experto podría construirse a partir de estos supuestos? Esta sería la siguiente cuestión que tendríamos que aclarar. Principalmente, serviría para la recuperación inteligente de la información. Sería un sistema experto en la recuperación de información especializada, aunque subsidiariamente también podría contestar preguntas sobre la especialidad elegida que facilitarían la búsqueda.

2.2. Funcionalidad del sistema

El mecanismo de funcionamiento se basa en un sistema de búsqueda a dos niveles que en ningún momento son percibidos por el usuario ni se le exige una opción al respecto ya

que el sistema decidirá por él de acuerdo con el tipo de pregunta que se formule. Los niveles básicos son los siguientes:

- 1) Nivel de búsqueda. Cuando el usuario plantea una búsqueda concreta de información. Ej. " Géneros musicales que se tocan con el violonchelo". Si en el transcurso de esta búsqueda se produce una pregunta, se pasa automáticamente al nivel de pregunta.

Categorías Generales	Sonido	Funciones	Características físico-morfológicas	Relación con otros instrumentos	Personas	Enseñanza	Tiempo
Violonchelo	ABCD	AB	ABCD	ABC	A	AB	S.XVI,XVII,XVIII
S U B C A T E G O R Í A S							
SONIDO	Tipo de mecanismo sonoro	Fuente del sonido	Forma de producir el sonido	Medio de producir el sonido	Carácter musical	Modo de colocación	Técnicas de interpretación
Violonchelo	instrumento	cuerda	frotación	arco	instrumento cantante, registros bajos, sonoridad profunda, rango de 3 octavas 1/2, afinado do, sol, re, la	sobre el suelo. Entre las piernas	Pizzicato
Relaciones con otros instrumentos	Relación/origenes	Relación/familia	Relación/semajanzas		Relación/competencia		
Violonchelo		relación familia de la viola de brazo, y viola	Familia del violín		viola da gamba	desplazó a la viola da gamba	
CARACT. FÍSICO-MORFOLÓGICAS		Cuerdas/número	Cuerdas/dimensión	Tamaño/instrumento	Con Cuello	Cuello sin trastes	
Violonchelo		4 cuerdas	704mm.	grande	con cuello	sin trastes	
PERSONAS	Personas/compositores			Personas/intérpretes		Personas/artesanos	Personas/teóricos
Violonchelo	Haydin, Beethoven, Handel, Bach, Mozart, Jacchini, Lanzatti, Cirri, Cervetto, Vivaldi, Schumann, Dvorak, Prokofiev, Saëns, Shostacovich, Debussy, R.Strauss, Casella, Britten, Rachmaninoff, Fauré			Bastitini, Corelli, Tartini,Franciscello, Bertaut,Janson, Duport,Cervetto, Casals, Maréchal y Cassadó		Stradivarius	Mich, de Cupis y Duport
FUNCIONES MUSICALES	Func. musicales/géneros	Func. musicales/formas	Func.musicales/orquestas	Func.musicales/partituras	Func. musicales/importancia		
Violonchelo	toca música de cámara, instrumental y dramática	Toca sonatas, conciertos y rícercares	se toca en la orquesta	instrumento solista. Clave de fa en 4ª	papel muy destacado en la música instrumental		
Enseñanza/métodos							
Violonchelo	métodos de Mich, Corrette, de Cupis y Duport						

Figura 2. Representación del concepto violonchelo en la estructura

- 2) Nivel de pregunta. El usuario tienen que hacer un trabajo sobre las formas musicales que se interpretan con instrumentos de cuerda sin trastes y no sabe cuáles son esos instrumentos. Ej. Si empieza con una pregunta: "¿cuáles son los instrumentos de cuerda sin trastes?" El sistema contestaría dándole la lista de ellos, el usuario elegiría uno y seguiría automáticamente el nivel 1.

Nivel de búsqueda. En este modo, el sistema proporciona una visualización del campo solicitado "géneros musicales" a partir del instrumento de su interés que es el violonchelo porque existe una relación establecida entre éste y los géneros musicales en la estructura (véase fig. 2). Al abrir los géneros, el usuario descubre cuáles son, en este caso: Música de cámara, música instrumental y música dramática. A partir de aquí, el usuario podrá optar por el o los que satisfagan su necesidad de información. Si el usuario escoge música dramática porque es el género que le interesa de los tres posibles, seguirá profundizando en su búsqueda y abrirá la música dramática y en ella se encontrará: ópera, zarzuela, etc. Se puede seguir profundizando, si es solicitado, hasta llegar a saber qué piezas concretas son o cuáles fueron sus compositores porque existen categorías en la estructura que permiten establecer este tipo de relaciones en una sola búsqueda. Todo el proceso se desarrolla presentando al usuario posibilidades de búsqueda ante las que sólo tendrá que elegir de acuerdo con sus

necesidades. Se establece así un diálogo con el usuario que se ve atendido por un especialista en el tema y en recuperación de la información, Al mismo tiempo el proceso se desarrolla con una mínima carga cognoscitiva para el usuario, aunque con gran potencialidad informativa, lo que dota al sistema con la interesante cualidad de ser docente ya que el usuario aprende del sistema con el mínimo esfuerzo.

Nivel de pregunta. Ha de entenderse aquí preguntas necesarias para la búsqueda y recuperación de la información, fundamentalmente aquellas referidas a aclaraciones o puntualizaciones conceptuales. Un caso típico puede ser cuando el usuario necesita recuperar información sobre un tema que no conoce del todo, como es el caso del ejemplo anterior: "quiere saber las formas musicales interpretadas por instrumentos de cuerda sin trastes, pero no sabe exactamente cuáles son esos instrumentos". En este caso, puede preguntar al sistema por los instrumentos de cuerda sin trastes y obtendrá la lista de ellos y, a partir de ahí, podrá seguir normalmente la búsqueda como en el nivel de búsqueda. Si el usuario está aún más perdido y su pregunta es más difusa, el sistema igualmente le ayuda a tomar decisiones porque puede mostrar la parte de su estructura relacionada con la pregunta, ofreciendo distintas rutas posibles de búsqueda pertinente que el usuario irá seleccionando en función de sus necesidades. Por ejemplo, si tiene que hacer un trabajo sobre algún aspecto relacionado con el papel o las funciones de los instrumentos de cuerda en la sociedad, pero no sabe nada más, será suficiente con que teclee instrumentos de cuerda y se abrirá una pantalla que le mostrará que estas funciones son: características étnicas o grupos étnicos relacionados con ellos, grupos militares o religiosos, festividades, estaciones del año, costumbres sociales, raigambre popular del instrumento, etc. A partir de aquí, podrá saber qué función o funciones le interesan y que instrumento o instrumentos están relacionados con ellas, porque ni todos los instrumentos tienen reconocidas funciones sociales, ni todos tienen las mismas y en la estructura están consignadas esas funciones y qué instrumentos musicales las desempeñan. Cuando está completamente informado, podrá hacer su búsqueda sobre los instrumentos de cuerda que estén relacionados con la función social que la haya interesado.

Por tanto, el sistema, por medio de su estructura principalmente, asiste al usuario desde el momento que contacta con él, en cualquiera de sus niveles o modos, hasta que éste consigue aclarar, refinar y estar seguro de haber representado correctamente su necesidad de información. Se trata de una ayuda que lo conduce de forma inteligente a navegar por la estructura del sistema hasta que tiene claro lo que busca y puede expresarlo adecuadamente en la frase de búsqueda.

2.3. Mecanismos de recuperación

El propio sistema es bastante autosuficiente, razón por la cual no son necesarias muchas técnicas que son habituales en los sistemas de recuperación de información. Concretamente, merece especial mención el mecanismo tan extendido de los operadores booleanos. Para plantear búsquedas como las anteriormente mencionadas, por ejemplo "Géneros musicales interpretados con el violonchelo" o "el papel de los instrumentos de cuerda en la sociedad", se usarían los operadores booleanos para combinar los conceptos deseados: instrumentos de cuerda y sociedad; géneros musicales y violonchelo. Aquí no es necesario este planteamiento porque, en el primer caso, se llega a la información sobre cuál o cuáles instrumentos son los que tienen una relación con su entorno social y de qué tipo de relación se trata en el diálogo con el sistema. Sabido esto, es más que probable que los documentos que traten sobre esos instrumentos satisfagan la necesidad de información del usuario, habida cuenta de que la vertiente social es un rasgo pertinente en ellos y de que esta pertinencia se ha extraído precisamente del análisis de textos especializados sobre ese concepto. La búsqueda de este tema usando operadores booleanos ocasionaría bastante ruido por lo impreciso del término sociedad en este contexto al tiempo que escondería mucha información en teoría disponible porque no se expresa qué instrumentos son los que tienen esta vertiente social y probablemente el resultado esté próximo al silencio documental. En el segundo caso, "Géneros musicales interpretados por el violonchelo" todavía es más evidente lo que se acaba de decir. La combinación booleana sobra aquí desde el momento que el propio sistema responde a esta necesidad (véase fig.2) y lo único que habría que pedir al sistema sería información sobre los géneros musicales concretos que está relacionados con el violonchelo: música de cámara, instrumental y dramática, o uno de ellos si así lo decide el usuario. Los documentos recuperados tratarían sobre esos géneros musicales concretos en los que también aparecería el violonchelo como su instrumento musical ejecutor que es. El planteamiento con los operadores conseguiría, al igual que en el caso anterior, una respuesta llena de imprecisiones y carente de exhaustividad.

3. CONCLUSIONES

Desde un punto de vista teórico, puede concluirse que estructuras conceptuales como las aquí descritas constituyen en sí mismas bases de conocimiento para entornos temáticos especializados que funcionalmente están próximas a los llamados sistemas expertos. Este tendría como característica más destacada la de ser un sistema inteligente para la recuperación de la información especializada y para responder a preguntas que normalmente el usuario haría a un profesional de la documentación especializado. El proceso de búsqueda está asistido en todo momento por ese "profesional" que no es otro que el propio sistema y su funcionalidad hace que la carga cognoscitiva sobre el usuario sea mínima, al no tener que aprender cómo es el sistema, y que consiste en ir eligiendo opciones conforme el sistema se las vaya presentando. Como además se ha previsto que todos los aspectos relevantes de los conceptos que forman el campo temático elegido estén representados en la estructura, todas las rutas conceptuales pertinentes entre conceptos estarán igualmente identificadas, lo que significa, en teoría, que las recuperación de la información sea elevada en sus parámetros de exhaustividad y pertinencia.

BIBLIOGRAFÍA

- ¹ LÓPEZ-HUERTAS, María J. El diseño de tesauros: Consideraciones teóricas y metodológicas. En: II *Encuentros sobre Organización del Conocimiento en Sistemas de Información y Documentación*. Zaragoza, Isko-Universidad, 1998,.
- ² BATES, M. Subject access in online catalogs: a design model. *Journal of the American Society for Information Science*, 37(6), 1986, 357-376
- BATES, M. Rethinking subject cataloging in the online environment. *Library Resources and Technical Services*, 34(4), 1989, 400-412
- ³ BATES, M. Subject access in online catalogs: a design model. *Journal of the American Society for Information Science*, 37(6), 1986, 357-376
- ⁴ FISCHER, Dietrich. From Thsauri to ontologies? En: *Structures and relations in knowledge organization. Proceedings of the Fifth International ISKO Conference, 25-29 August, 1998*. Ed. by Widad Mustafa el Hadi. Würzburg, Ergon Verlag, 1998, 18-30.
- ⁵ LÓPEZ-HUERTAS, María J. Thesaurus structure design: a conceptual approach to improved interaction. *Journal of Documentation*, vol. 53(2), 1997, 139-177.
- ⁶ BATES, Marcia. The design of browsing and berrypicking techniques for the online search interface. *Online Review*, 13(1), 1989, 407- 424.
- ⁷ BEAN, Carol. The nature of explicit parent-child relationships in Mesh tree structures. En: *Structures and relations in knowledge organization. Proceedings of the Fifth International ISKO Conference, 25-29 August, 1998*. Ed. by Widad Mustafa el Hadi. Würzburg, Ergon Verlag, 1998, 132-137.
- ⁸ HUDON, Michèle. A preliminary investigation of the usefulness of semantic relations and of standardized definition for the purpose of specifying meaning in a thesaurus. En: *Structures and relations in knowledge organization. Proceedings of the Fifth International ISKO Conference, 25-29 August, 1998*. Ed. by Widad Mustafa el Hadi. Würzburg, Ergon Verlag, 1998, 138-144.
- ⁹ LIKKE NIELSEN, Marianne. Future thesauri: What kind of structures and relations do searchers need? En: *Structures and relations in knowledge organization. Proceedings of the Fifth International ISKO Conference, 25-29 August, 1998*. Ed. by Widad Mustafa el Hadi. Würzburg, Ergon Verlag, 1998, 152-159.