

Evaluación de softwares de gestión de tesauros

Gregorio Moya Martínez

Isidoro Gil Leiva

RESUMEN

Se presenta la evaluación de cinco paquetes de softwares de gestión de tesauros (BEAT, MultiTes, Stride, TCS-y2k y SGAT). Para llevar a cabo dicha evaluación, se utiliza una serie de criterios específicos de este tipo de software: por un lado, los criterios relativos a la construcción y mantenimiento de tesauros (definición de términos y sus atributos, relaciones entre los términos y edición y mantenimiento del tesauro); por otro lado, criterios sobre las salidas que ofrecen estos paquetes de softwares (salida a pantalla, y salida a impresora y archivo, o a ambos). Los resultados de dicha evaluación se presentan en un conjunto de tablas acompañadas de comentarios junto a cada uno de los criterios analizados.

ABSTRACT

This paper presents the evaluation of five softwares packages of thesaurus management (BEAT, MultiTes, Stride, TCS-y2k and SGAT). For it, we have used a series of specific criteria of this type of software: on the one hand, the criteria relative to the construction and maintenance of thesauri (definition of terms and its attributes, relations between the terms and editing and maintenance of the thesaurus); on the other hand, criteria on the outputs that offer these softwares packages (output to screen and output to printer and/or file). The results of this evaluation appear in a group of tables accompanied by comments next to each one of the analyzed criteria.

Introducción

Los trabajos publicados sobre software de gestión de tesauros giran alrededor de tres temas: unos sobre especificaciones para el desarrollo de estos programas Milstead [1], Interián [2] y Wu [3]; en otros artículos se describen aplicaciones para la gestión automática de tesauros como en Iljon [4] que presenta Astute, desarrollado en Euronet; Burkart [5] quien reseña tres versiones de Proterm; Miranda, Medeiros y Sujii [6] que hacen lo mismo con Tecer, una herramienta desarrollada en el Instituto Brasileño para la Información Técnica y Científica; y López Alonso [7] quien revisó programas de gestión de tesauros independientes como Stride, TCS o Topic y gestores integrados en sistemas de recuperación como Basisplus o API-Cais.

Por último, en otro grupo de artículos se proporcionan criterios para la evaluación como en Ganzmann [8],

Milstead [9] y Milstead [1], o bien se evalúan directamente como en el caso de Rohou [10] que analiza los tipos y números de relaciones, niveles de jerarquía, control de duplicados, visualizaciones, funciones de indización, modos de búsqueda y coste, entre otros aspectos, de diez programas de gestión de tesauros utilizados en Francia (Astute, Alexis, Basis, Mistral, Minisis, Jlb/Doc, Micro-Cairs, Sacado, Liber y Campus/Doc); Ritzler [11] evalúa Index, Proterm y TMS desde el punto de vista de la interfaz de usuario, funcionalidad y fiabilidad; o como Schonfeldt [12] que realiza una evaluación de seis programas manejados en diferentes instituciones alemanas (Adis, Index, Lidos, Pflsaurus, Proterm y Stride) para comprobar los tipos de relaciones permitidas, la consistencia, las interfaces de usuarios y sus capacidades hipertextuales.

En el presente trabajo, nos centramos exclusivamente en paquetes de software de gestión de tesauros que pueden utilizarse de forma independiente. Los programas evaluados se han obtenido de manera gratuita en Internet (excepto SGAT). Aunque algunos de ellos son versiones de demostración, son totalmente operativos (únicamente estaban limitados en cuanto al tiempo o al número de veces que se pueden usar).

Material y método

El software de gestión de tesauros seleccionado es el siguiente:

- BEAT Thesaurus Software 3.2
- MultiTes 6.0 para Windows, de Multisystems.
- Stride 6.2.1 de Questand Ltd.
- TCS-y2k (Thesaurus Construction Systems)
- SGAT 2.0

Tras la revisión bibliográfica sobre la evaluación de programas informáticos, finalmente se decidió utilizar el trabajo de Ganzmann [8] para la obtención de los criterios de evaluación.

Con respecto al método hay que señalar que se han seguido los siguientes pasos:

1) Recopilación de las aplicaciones de gestión de tesauros

Todos los tesauros evaluados se han obtenido de manera gratuita en Internet, si bien en algunos casos se trata solamente de una demostración. A continuación se ofrecen las URL:

- <<http://www.willpower.demon.co.uk/thessoft.htm>>, donde se presenta una lista de software para la creación y gestión de tesauros, así como información básica de cada programa.
- <<http://www.asindexing.org/thessoft.html>>, página de la American Society of Indexers dedicada a software de gestión de tesauros.
- <<http://www.aussi.org/resources/software/thesauri.htm>>, página de la Australian Society of Indexers, donde se recopila software de gestión de tesauros.

Además, hemos incorporado un programa experimental, SGAT 2.0, realizado por Juan A. Pastor Sánchez <pastor@um.es>, miembro del Grupo de Tecnologías de la Información de la Universidad de

Murcia. Esta aplicación todavía se encuentra en fase de mejora.

2) Establecimiento de los criterios de evaluación

Los criterios para evaluar los programas informáticos se suelen dividir en dos grandes bloques: uno dedicado a criterios que son aplicables a cualquier tipo de software (criterios generales), y otro conjunto compuesto por criterios adaptables al grupo de programas evaluados (criterios específicos). Este trabajo está dedicado solamente a evaluar los criterios específicos de programas para la gestión automática de tesauros. A continuación se presentan estos criterios separados en dos apartados, el primero hace referencia a la construcción y mantenimiento, mientras que el segundo, se dedica a las salidas de los tesauros.

Relativos a la construcción y mantenimiento de tesauros

Debido al gran número de *items* que se evalúan en este apartado se han dividido en tres subapartados:

- *Definición de los términos y sus atributos:* Se analizan las posibilidades que ofrece el software para que el usuario desarrolle la estructura necesaria para sus necesidades específicas.
- *Relaciones entre los términos:* Se evalúa la definición de relaciones (relaciones predefinidas y relaciones definidas por el usuario); los tipos de relaciones permitidas entre los términos; y la potencia para la construcción de tesauros multilingües.
- *Edición y mantenimiento del tesoro:* Se analizan los criterios relacionados con la entrada, la modificación y la supresión de datos. Asimismo, se incluyen los criterios referentes al control de consistencia del tesoro.

Relativos a las salidas de los programas de gestión de tesauros

- *Salida a pantalla:* Se comprueban los modos de visualización y presentación del tesoro, así como la visualización de términos individuales entre otros aspectos.
- *Salida a impresora y a archivo, o a ambos:* Se analiza la definición de los formatos de salida, las copias proporcionadas, los formatos de la página impresa o los formatos para la exportación a un fichero, entre otros elementos.

3) Preparación de los programas y evaluación

Con la finalidad de disponer los programas para su evaluación se han llevado a cabo una serie de pruebas:

- *Prueba 1:* Revisión de la documentación adjunta al programa de gestión de tesauros y ficheros de ayuda.
- *Prueba 2:* Introducción de un tesoro nuevo. Algunos de los programas cuentan con un tesoro a modo de ejemplo, pero hay otros que no, por lo que se consideró conveniente homogeneizar este punto introduciendo un pequeño tesoro.
- *Prueba 3:* Control de consistencia. Se borraron y modificaron descriptores, o ambos, y no descriptores y relaciones.
- *Prueba 4:* Tesoros multilingües. Se comprobó cómo trabajan al describir relaciones entre términos en distintos idiomas.
- *Prueba 5:* Salidas. Se estudiaron las salidas tanto impresa como en fichero, así como las posibilidades de exportación.
- *Prueba 6:* Importación de datos. Opciones implementadas para el intercambio de datos.

4) Elaboración de tablas comparativas con los resultados obtenidos

Se han realizado una serie de tablas comparativas en donde están agrupados los resultados obtenidos por cada programa con respecto a la evaluación de un determinado criterio. El objetivo de estas tablas es la presentación de la información de una manera esquemática y que permita una fácil comparación entre los distintos programas.

Resultados

Datos comunes al software evaluado

El software para la gestión de tesauros, lo podemos dividir en dos grandes grupos: programas independientes de cualquier otro tipo de software, o programas que sólo están disponibles como un módulo de un sistema de almacenamiento y recuperación de información. Aunque hay una serie de ventajas al utilizar un módulo de gestión de tesauros dentro de un sistema completo (integración entre las operaciones de catalogación, indización y búsqueda), hay que tener en cuenta que dichos

sistemas necesitan de muchos más recursos hardware (memoria principal, espacio de almacenamiento, velocidad de procesamiento, etc.) y software (sistemas gestores de bases de datos), que los sistemas de gestión de tesauros independientes. Estos últimos son más apropiados cuando no se necesita una interacción con el sistema de almacenamiento y recuperación de la información.

Se distinguen cuatro grandes bloques a la hora de exponer las tareas que debe soportar un software de gestión de tesauros: Construcción y mantenimiento; Salidas del tesoro; Indización y recuperación y finalmente Integración y compatibilidad de vocabularios. Veamos con mayor detenimiento cada uno de ellos.

Construcción y mantenimiento del tesoro

La construcción y el mantenimiento de un tesoro incluye una variedad de tareas y actividades, entre las que podemos destacar:

Los términos que constituyen el tesoro se seleccionan y almacenan en el mismo. Es aconsejable guardar junto a ellos, una serie de informaciones relativas a los mismos (atributos) como, por ejemplo, el idioma, la fuente del término, las notas de alcance, definiciones, etc. En ocasiones deben integrarse términos procedentes de diferentes tesauros previamente existentes.

Hay que definir las relaciones entre los términos. Generalmente, este tipo de software está diseñado para soportar las relaciones básicas:

- Relaciones de equivalencia: USE/UF.
- Relaciones jerárquicas: BT/NT.
- Relaciones asociativas RT/RT.

También están diseñados para mantener la consistencia en el tesoro, asegurando que cuando se crea o se elimina una de estas relaciones entre dos términos, la relación recíproca también se crea o se anula.

Por último, el software debe ofrecer facilidades para el mantenimiento del tesoro, puesto que estos deben actualizarse regularmente con respecto a las relaciones y terminología debido a los cambios en la terminología, en la indización y en los hábitos de recuperación de los usuarios.

Salidas del tesoro

Dentro de este apartado, podemos distinguir: visualización en la pantalla del tesoro, salidas impresas y salida o volcado a un archivo. En cuanto

a la visualización en pantalla del tesoro, o de una parte seleccionada previamente, todas las aplicaciones de gestión de tesoros permiten la visualización del mismo de forma alfabética y sistemática. Normalmente también ofrecen una visualización jerárquica y muchos de ellos incluso una permutada (KWIC/KWOC). Las salidas por la impresora suelen brindar las mismas alternativas que se dan para la visualización en pantalla. Por último, la mayoría de los programas presentan la salida a un archivo, que se puede utilizar para su posterior edición por parte de un procesador de textos, o incluso para la integración de los datos en otro tesoro.

Indización y recuperación

Los tesoros integrados en sistemas de almacenamiento y recuperación de información pueden interactuar con algunas de las tareas relacionadas con la indización y recuperación. Como hemos mencionado anteriormente, este tipo de software queda fuera de nuestro ámbito de estudio y por lo tanto no profundizaremos más en este apartado.

Integración y compatibilidad de vocabularios

Este aspecto está recibiendo bastante atención en los últimos tiempos, puesto que existe una creciente demanda por parte de los usuarios en un conjunto de prestaciones que hasta ahora no estaban muy extendidas. Las principales demandas son las siguientes:

- La entrada y salida de los datos contenidos en un tesoro en un formato de intercambio debe ser fácilmente transportable de un sistema gestor de tesoros a otro (importación/exportación).
- El uso de vocabularios compatibles en la indización y en la recuperación.
- Tipos de relaciones adicionales para explotar todo el potencial contenido en los tesoros multilingües.

En definitiva, y para concluir con este apartado, podríamos resumir todo lo anteriormente expuesto en que para satisfacer las necesidades de los usuarios, un buen programa de gestión de tesoros debe ser tan flexible como sea posible, dejando al usuario que defina libremente sus necesidades específicas en cuanto a los requisitos referidos al tamaño y complejidad del vocabulario, la definición de relaciones y la visualización del tesoro.

Datos específicos del software evaluado

BEAT Thesaurus Software 3.2

Desarrollado por el español Josep Sau, la versión analizada es la 3.2 que cuenta con versiones del programa en castellano, catalán e inglés. Este programa trabaja bajo el sistema operativo DOS.

Principales características:

- Verifica la coherencia de cualquier modificación en los términos o relaciones y crea automáticamente las relaciones inversas.
- Permite la asociación de notas de texto a cada término para comentar su aplicación (notas explicativas), la historia (notas históricas) o la fuente terminológica (notas de fuente).
- Mantiene la indización de los términos automáticamente. No todas las palabras que forman un término se consideran para la búsqueda. El usuario puede crear una lista propia de palabras vacías que se recogen en un fichero llamado Beat.STP en el momento de la creación de un nuevo tesoro o en la reindización de uno ya existente. El fichero Beat.STP es un fichero de texto que puede ser editado por el usuario.
- Genera listados del tesoro en diferentes formatos que se envían a ficheros de texto Ascii y se imprimen o editan posteriormente con un procesador de textos.
- Proporciona la posibilidad de incorporar datos (términos, relaciones, códigos, o notas) desde ficheros de texto con un proceso de carga masiva.
- Mantiene un fichero de diario (extensión .LOG) con el registro de las operaciones, modificaciones o errores generados durante las sesiones de trabajo. Este fichero de texto puede visualizarse, editarse o borrarse fuera del programa. El fichero .LOG al generarse con un formato que sirve como entrada para una carga masiva, puede aplicarse para rehacer en el mismo orden todas las operaciones de una sesión de trabajo sobre una copia del estado anterior del mismo tesoro.

MultiTes 6.0 para Windows

Principales características:

- Validación de relaciones conflictivas.

- Generación automática de relaciones recíprocas.
- Soporte de tesauros monolingües y multilingües.
- Soporte para tesauros polijerárquicos.
- Más de 100 millones de términos por tesoro.
- Soporte de las relaciones normalizadas más las establecidas por el usuario. Cuando se crea un nuevo tesoro, se definen automáticamente las relaciones y notas normalizadas. El sistema permite el cambio de estas relaciones y la definición de las que el usuario considere necesarias. Puede tener cuatro pares de relaciones de cada tipo (USE/UF sería un par).
- Definición de hasta cuatro campos de comentarios compuestos por texto libre con 65 000 caracteres.
- Permite la importación de un fichero de texto con los términos y sus relaciones.
- Genera de informes. MultiTes permite la selección de los términos que se desean imprimir, así como la información que acompaña a cada término. También posee un generador de ficheros HTML.
- Puede reconstruir de índices (*rebuild*), que permite rehacer los ficheros índices que son necesarios para el sistema.

Stride 6.2.1

Características:

- Puede trabajar en entornos multiusuario y está diseñado de forma que se pueda operar tanto de manera independiente como dentro de un sistema de información.
- Disponible para plataformas Unix y como un API de desarrollo de software.
- Permite el uso de las relaciones estándar y la definición de otras relaciones que necesite el usuario en su ámbito de trabajo. El tipo de relación puede ser: jerárquico, asociativo, equivalencia y clasificación.
- Cada término del tesoro se clasifica dentro de un grupo. Si se especifican los grupos, es posible la definición de listas de grupos. Si no se ha especificado ningún grupo, todos los términos están agrupado bajo el grupo 0 (predefinido). En un tesoro políglo, los

grupos se utilizan para asignar los términos en sus respectivos idiomas.

- Para que puedan importarse ficheros deben de estar en un formato de texto Ascii denominado *STRIDE Fetch File Format*. Estos ficheros se crean desde cualquier procesador de texto que grabe archivos en formato Ascii.

TCS-Y2K (*Thesaurus Construction Systems*)

Liu-Palmer Thesaurus Construction System, versión 4 para Microsoft Windows® 95 ó 98 o NT 4.0.

Características básicas:

- Propociona varios formatos de visualización de los términos.
- Asegura la consistencia con la generación automática de las relaciones inversas.
- Facilita el mantenimiento de la base de datos del tesoro con la actualización automática de los índices necesarios.
- Proporciona listados estructurados que pueden ir a la impresora o a un archivo para su edición en un procesador de textos.
- Una peculiaridad de este sistema es que todo término debe estar dentro de una jerarquía. No es de extrañar que la vista principal de un tesoro sea por lo tanto la jerárquica.
- Puede soportar miles de jerarquías en un único tesoro. Cada jerarquía puede contener un número ilimitado de términos. El nombre de la jerarquía es un título o frase descriptiva apropiada que no se trata como un término dentro del sistema, es decir, el nombre de una jerarquía no es un término del tesoro.

SGAT 2.0

El Sistema Gestor para la Automatización de Tesauros (SGAT) ha sido desarrollado por Juan Antonio Pastor Sánchez. En las figuras 1, 2, y 3 se muestran las ventanas de gestión del tesoro y la presentación alfabética que muestra el programa.



Fig. 1. Sistema Gestor para la Automatización de Tesoros (SGAT).

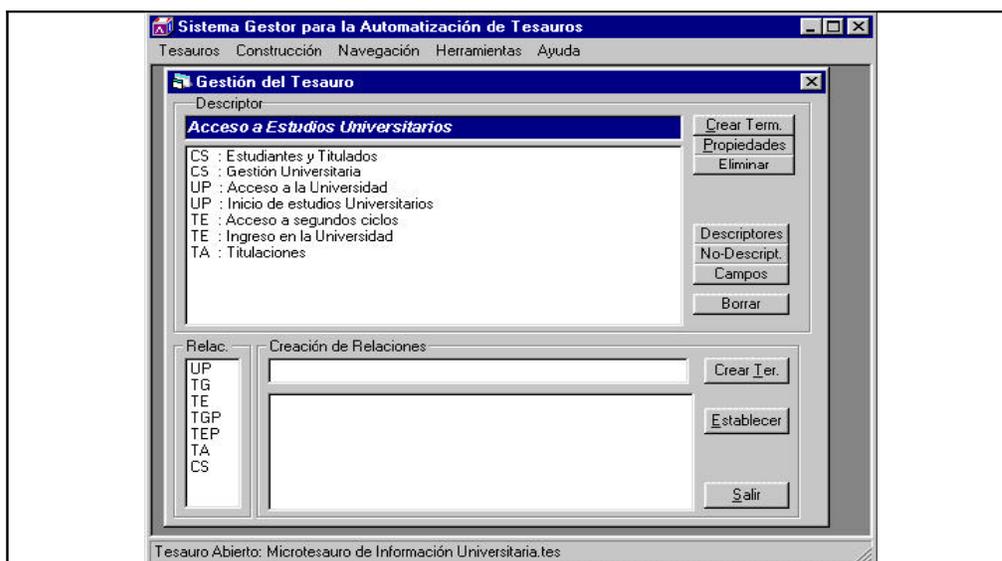


Fig. 2. SGAT. Gestión del Tesoro. Acceso a estudios universitarios.

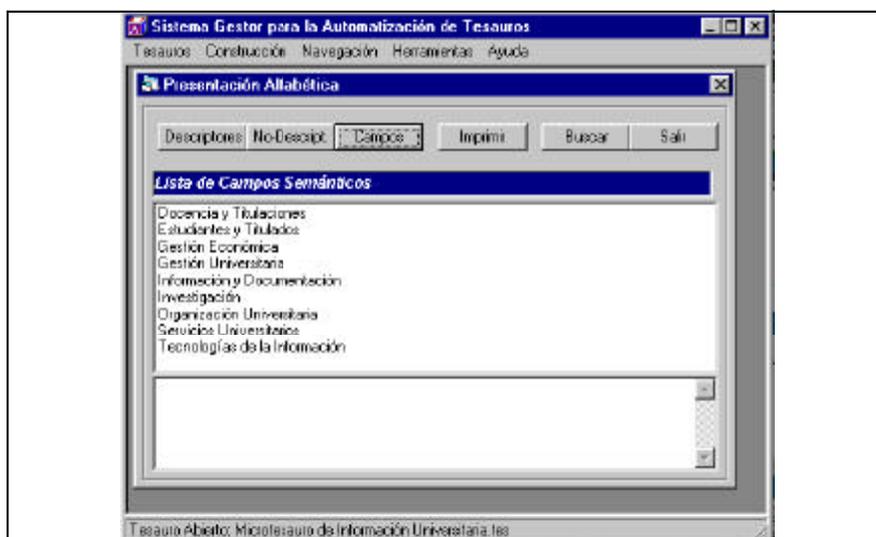


Fig.3. SGAT. Presentación alfabética. Lista de campos semánticos.

Tabla comparativa del software evaluado

A pesar de que todos los sistemas han sido sometidos a las pruebas que hemos descrito anteriormente, en algunos casos no ha sido posible determinar los límites para un determinado criterio de evaluación. En estos casos, hemos recurrido a la documentación o a la ayuda que proporciona dicha aplicación. Si se especificaba dicha cuestión, la hemos reflejado. En caso de que no aparecieran estas limitaciones recogidas en la documentación, lo hemos indicado como indeterminado. Por ejemplo, en algunos sistemas no se ha podido determinar el número máximo de términos que es capaz de albergar, si bien en todos ellos se introdujo un tesoro formado por 3 000 descriptores, unos 700 no descriptores y unas 8 000 relaciones.

En las tablas 1 y 2 se muestran los resultados de la evaluación de dichos programas, agrupados por los distintos criterios que se han evaluado.

Conclusiones

Del análisis de los programas informáticos para la gestión de tesauros se concluye que:

- 1) En cuanto a las capacidades de almacenamientos de términos, se ha comprobado que el número máximo de términos que se pueden introducir, por lo general, no tiene limitaciones (salvo el espacio de almacenamiento disponible). En lo que se refiere a la longitud máxima de cada término, en casi todos los programas suele ser más que suficiente, si bien en algunos casos está limitada.
- 2) Se debería mejorar la gestión de las notas puesto que algunos programas sólo admiten uno o dos tipos de notas, por lo que sería recomendable más flexibilidad para que el usuario tuviera la posibilidad de definir las notas que estimara convenientes en cada caso. En lo que se refiere a la longitud de las notas se ha observado que todos cuentan con un espacio suficiente.
- 3) Ninguno de los programas estudiados permite el establecimiento del estado del término excepto uno.
- 4) En la mayoría del software solamente se pueden establecer las relaciones normalizadas entre los términos, aspecto que se debería mejorar, bien definiendo más relaciones de cada tipo o bien dejando al usuario la definición de las relaciones necesarias.
- 5) Aunque la mayoría de los programas reconocen los términos superiores (*Top terms*), muy pocos son los que presentan la posibilidad de identificarlos mediante algún tipo de indicador.
- 6) El control de la consistencia es un tema que por lo general suele estar bien resuelto, lo que ayuda a mantener la integridad y coherencia de los datos. No obstante, algunos de los programas presentan lagunas como la formación de bucles en las relaciones jerárquicas.
- 7) Las posibilidades de presentación de los datos en pantalla es suficiente, si bien un inconveniente es que no todos los programas ofrecen la posibilidad de establecer criterios para la visualización de subconjuntos.
- 8) En cuanto a la impresión, deberían dar una mayor libertad al usuario para la definición de los formatos de impresión que le sean necesarios y por otro lado, la incorporación de un generador de informes que estableciera el esquema de formato para la impresión.
- 9) La mayoría de los programas evaluados son capaces de preparar ficheros para la exportación de datos a otros sistemas. En algunos casos, permiten la posibilidad de enviar a la impresora la visualización en pantalla que se está realizando.
- 10) Por último, y desde una perspectiva de conjunto, se debería mejorar la flexibilidad de este tipo de software y permitir al usuario la adaptación del sistema a las necesidades de cada tesoro, dejando mayor libertad en la definición de las notas, en las relaciones específicas, en la definición de formatos de impresión, etc.

Referencias

- 1) Milstead, J. Specification for thesaurus software. *Information Processing and Management* 27(2-3):165-175, 1991.
- 2) Interián, M. G. Confección de un tesoro por medios automatizados. *Ciencias de la Información* 24(2):108-109, 1993.
- 3) Wu, M. M. Criteria issue for the thesaurus management software. *Journal of Library and Information Science* 20(2):75-90, 1994.
- 4) Iljon, A. Le logiciel Astute: un outil informatique pour la creation, la mise a jour, l'edition et l'impression de thesauri mono et multilingues. *Documentaliste* 14(5/6):25-33, 1977.
- 5) Burkart, M. Proterm a software package for construction, updating and handling of thesauri and other kinds of vocabularies. *Nachrichten fur Dokumentation* 39(4):249-252, 1988.
- 6) Miranda, L. M., M. B. B. Medeiros y M. K. Sujii. Elaboração de tesouros utilizando-se o programa de elaboração de tesaurtos em microcomputador (Tecer). *Revista de Biblioteconomia de Brasília* 18(2):185-192, 1990.
- 7) López Alonso, M. A. La gestión automatizada de tesouros: estado de la cuestión. *Revista Interamericana de Bibliotecología* 21(1):49-69, 1998.
- 8) Ganzmann, J. Criteria for the evaluation of thesaurus software. *International Classification* 17(3/4):148-157, 1990.
- 9) Milstead, J. Thesaurus software packages for personal computers. *Database* 13(6):61-65, 1990.
- 10) Rohou, C. La gestion automatisee des thesaurus: etude comparative de logiciels. *Documentaliste* 24(3):103-108, 1987.
- 11) Ritzler, C. Comparative study of PC-supported thesaurus software. *International Classification* 17(3-4):138-147, 1990.
- 12) Schonfeldt, R. Thesaurus software: a comparison. *Nachrichten fur Dokumentation* 46(3):177-186, 1995.

Bibliografía

Correa Uribe, G. Un modelo para la evaluación integral de tesouros. *Revista nteramericana de Bibliotecología* 22 (1):139-146, 1999.

Negrini, G. Clasthes: a thesaurifacet creator. *Library Science with a Slant to Documentation* 31(1):1-12, 1994.

Rodríguez Muñoz, J. V. Construcción del esquema conceptual del tesoro mediante un modelo de datos [Microforma]. Murcia, Universidad, 1994.

Rosa, Antonio de la. Tesouros, Tesouros automáticos y Tesouros online. *IWE. El profesional de la información* 8(7-8):4-15, 1999.

Recibido: 23 de julio del 2001.

Aprobado: 3 de agosto del 2001.

Gregorio Moya Martínez

Departamento de Información y Documentación,
Facultad de Ciencias de la Documentación,
Universidad de Murcia
30071. Murcia
España
Correo electrónico: <goyoma@um.es>.
