

Construcción de un tesoro de Ciencias de la Documentación aplicado a la docencia de las técnicas documentales

Blanca Rodríguez Bravo

dphbrb@unileon.es

M^a Luisa Alvite Díez

dphlad@unileon.es

Con la colaboración de:

Josefa Gallego Lorenzo

M^a Antonia Morán Suárez

Carmen Rodríguez López

Lourdes Santos de Paz

Universidad de León

España

Resumen

Se presenta el proyecto de construcción de un tesoro de Ciencias de la Documentación realizado por el Área de Biblioteconomía y Documentación de la Universidad de León, con ayuda del programa MultiTes, y gracias a la ayuda concedida en el año 2000 por la Junta de Castilla y León. Una nueva subvención otorgada por el mismo organismo en abril de 2002 servirá para su extensión semántica, traducción al idioma inglés y desarrollo de un CGI que posibilite superar la consulta navegacional del tesoro con la búsqueda de términos por medio de un formulario.

El proyecto de construcción de un tesoro de Ciencias de la Documentación, surgió como una iniciativa del Área de Biblioteconomía y Documentación de la Universidad de León, a la que ha contribuido decisivamente la Consejería de Educación y Cultura de la Junta de Castilla y León, con la concesión de una subvención, por Resolución de 16 de junio de 2000, que se enmarcaba dentro del apartado d) del punto 1.2. de las bases de la convocatoria para la elaboración de recursos de apoyo a la enseñanza universitaria.

Fue dirigido en una primera fase por el Profesor José Antonio Martín Fuertes, y tras su desgraciado fallecimiento por la Profesora Blanca Rodríguez Bravo. El equipo de trabajo lo han constituido las profesoras M^a Luisa Alvite Díez, Josefa Gallego Lorenzo, M^a Antonia Morán Suárez, Blanca Rodríguez

Bravo, M^a Carmen Rodríguez López y Lourdes Santos de Paz. Todos los miembros pertenecen al área de conocimiento de Biblioteconomía y Documentación, adscrita al Departamento de Patrimonio Histórico-Artístico y de la Cultura Escrita de la Universidad de León.

Como se expondrá más abajo, esta ayuda ha ido seguida de una nueva petición de subvención con el fin de ampliar el contenido del tesoro elaborado durante los años 2002-2003. El 4 de abril ha salido la resolución provisional favorable a dicha petición de Proyecto de Extensión Semántica del Tesoro de Ciencias de la Documentación en el seno de la convocatoria de elaboración de *Recursos de Apoyo y Experiencias Innovadoras de las Universidades de Castilla y León* que dirige la profesora Blanca Rodríguez Bravo y que amplía el equipo de colaboradores.

1 Objetivo y situación de partida

Como se justificaba en la petición de la ayuda, el objetivo ha sido la elaboración de un tesoro especializado en Documentación, necesario para la realización de las prácticas de la asignatura troncal "Lenguajes documentales" y de la optativa "Tesoros", ambas impartidas en el tercer curso de la Diplomatura en Biblioteconomía y Documentación de la Universidad de León.

Existen varios tesoros en lengua española elaborados para tratar diversas áreas del conocimiento, pero ninguno sobre las Ciencias de la Documentación. Por tanto, es de gran utilidad para la docencia el poder contar con un tesoro propio de estas Ciencias, que permita potenciar las prácticas de indización y

recuperación de la información evitando la rémora que supone el empleo de instrumentos inadecuados, que lejos de clarificar los contenidos docentes contribuyen a su enmarañamiento.

El estudio y reflexión sobre los conceptos de las distintas Ciencias de la Documentación, la elección de la terminología que mejor representa a esos conceptos y su estructuración semántica es también de enorme utilidad para el profesorado participante en la elaboración de este trabajo, máxime cuando al ser un conjunto de Ciencias jóvenes tienen todavía mucho por definir en este terreno terminológico, donde muchos conceptos reciben variedad de denominaciones siendo necesario lograr un grado mayor de normalización y univocidad.

La estructuración jerárquica de los distintos conceptos implica también una interesante reflexión sobre las distintas materias y el orden de impartición de sus principales contenidos, es decir, tiene beneficios didácticos importantes.

2 Contenido y estructura

El objetivo era la construcción de un tesoro de Ciencias de la Documentación que englobara la terminología necesaria para el tratamiento de la información de los distintos campos abarcados por el concepto científico de "documentación" en un sentido amplio. Se partió de una división en campos semánticos en función de las asignaturas obligatorias que en la titulación se imparten, como un medio de distribución del trabajo entre los participantes en el proyecto, sin embargo los campos semánticos resultantes se alejan de esta primera división pues se constató que existían muchas interconexiones entre las disciplinas documentales y, por tanto, muchos conceptos se repetían.

El tesoro se ha estructurado finalmente en siete campos semánticos:

- 01 Ciencias de la Documentación. Historia. Teorías. Sistemas.
- 02 Investigación y Metodología documental.
- 03 Información. Documentos. Fuentes de información.
- 04 Tratamiento y Recuperación de la información.
- 05 Sistemas de Información.
- 06 Tecnologías de la Información.
- 07 Profesionales y Usuarios.

En ellos se organizan los conceptos de forma sistemática, yendo de los conceptos generales a los más específicos. De esta forma se ha

elaborado un índice jerárquico que ha servido de base para la elaboración de los otros dos apartados del tesoro: el índice alfabético y el índice permutado.

El índice alfabético relaciona todos los descriptores y no descriptores indicando bajo cada uno las relaciones preferenciales que remitan de Descriptores a No descriptores por medio del operador U.F. (Used for = Usado Por) y de los no Descriptores a los Descriptores con el operador USE.

Las relaciones jerárquicas establecidas en el índice arbóreo: B.T. (Broader Term = Término Genérico) y N.T. (Narrower Term = Término Específico), y las relaciones asociativas indicadas por el operador R.T. (Related Term = Término Relacionado), convierten al tesoro en un instrumento que induce y guía al documentalista y al usuario en la utilización de conceptos relacionados con el que había buscado en primer lugar.

Se ha elaborado, por último, un índice permutado tipo KWOC que permite localizar los descriptores sintagmáticos o precoordina- dos por el segundo o tercer término significativo.

La elaboración de este tesoro se ha llevado a cabo con ayuda del programa americano MultiTes de ahí que los operadores que indican las relaciones entre el vocabulario aparezcan en inglés.

Paralelamente se elaboró una sede web con contenidos y recursos relacionados con los tesoros que creemos ha proporcionado valor añadido al instrumento terminológico construido.

3 Metodología y Fases de la actividad realizada.

En el trabajo realizado ha habido varias fases de actividad paralelas.

3.1 Fases de la elaboración intelectual del tesoro:

Como se ha dicho, se realizó un primer reparto por asignaturas entre los participantes con vistas a la extracción y estudio del vocabulario. La primera fase consistió, por tanto, en la recopilación del vocabulario. Cada participante en el proyecto se ha ocupado de un campo temático según los distintos perfiles del profesorado del Área.

Los procedimientos seguidos han sido dos:

- 1) Método sintético o inductivo, consistente en la extracción del vocabulario utilizando

Diccionarios, Terminologías, Tesoro de la UNESCO, Listas de encabezamientos de materia, y todos aquellos instrumentos léxico-semánticos que nos proporcionen terminología ya asentada.

- 2) Método analítico o deductivo, consistente en la indización de artículos y monografías de los últimos años para extraer un vocabulario actualizado.

El resultado ha sido una recopilación de más de mil términos, entre Descriptores y No descriptores.

La segunda fase consistió en la puesta en común y en el estudio de la terminología extraída. Se eligió entre los distintos sinónimos el Descriptor o término preferente, evitando los homógrafos o términos polisémicos, redactando notas aclaratorias para la utilización de los descriptores que pudieran resultar ambiguos. La terminología de las Ciencias de la Documentación dista mucho de estar bien fijada, por lo que se ha introducido en el tesoro mucha sinonimia y en muchos casos optar por el Descriptor ha resultado problemático.

La tercera fase consistió en el establecimiento de los campos semánticos, eligiendo los macrodescriptores o cabezas de jerarquía, procediendo a continuación, al desarrollo de esas jerarquías, ubicando cada descriptor en su lugar idóneo conceptualmente. En muchos casos hubo que optar por una ubicación en detrimento de otra que también podía ser apropiada pues, como se ha dicho, la misma terminología se repite en distintos ámbitos.

La cuarta fase correspondió a la nivelación de los campos semánticos, profundizando o abreviando las distintas jerarquías para buscar un equilibrio entre los distintos apartados.

La quinta fase, una vez elaborado el índice jerárquico, consistió en desarrollar un índice alfabético, estudiando, fundamentalmente, las posibles asociaciones entre los descriptores.

3.2 Fases de la construcción automatizada del tesoro:

Desde la puesta en marcha del proyecto y, tal y como se señalaba en la petición de ayuda, consideramos imprescindible contar con una aplicación informática específica para el tratamiento y almacenamiento de los términos de indización, toda vez que las fases intelectuales previas hubieran sido

cumplimentadas y, obviamente, para la edición del tesoro en formato impreso y electrónico.

Consideramos que, dado que la web se ha convertido indiscutiblemente en interfaz única, el proyecto debía desembocar en un producto publicable y consultable a través de la World Wide Web.

Dado que conocíamos previamente el tesoro elaborado para la consulta por materias al catálogo de la Universidad de Barcelona, nos dirigimos a la biblioteca de la referida entidad con el fin de conocer el software empleado. El programa utilizado es *Beat thesaurus*, disponible gratuitamente en <http://www.aussi.org/resources/software/Beat.zip>. Se trata de un programa en versión DOS y que no cuenta con interfaz web, su uso, sin embargo, resultaba sencillo y permitía el control y almacenamiento de los términos adecuadamente.

El paso siguiente fue dirigirnos a las principales empresas consultoras y distribuidoras de software documental en nuestro país con el fin de recabar información sobre posibles productos ya elaborados para propósitos semejantes. Todas ellas nos ofrecían como única posibilidad la adquisición de sistemas de gestión de bases de datos documentales dentro de los cuales podíamos utilizar un módulo específico de tesoro o control de autoridades.

El problema común a todas las soluciones planteadas consistía, de un lado en el precio elevado de los correspondientes softwares y, de otro, la infrutilización de los mismos para el proyecto tan concreto que pretendíamos desarrollar.

De este modo, Doc6 (<http://www.do6.com>) nos propuso la utilización del programa Inmagic DB/Text Works para el tratamiento de los términos y DB/TextWeb Publisher para la fase de publicación y consulta en la web. Su solución resultaba económicamente desproporcionada por estar centrada en la implantación de una aplicación documental completa ideada para el tratamiento y consulta de bases de datos tanto en la web como en intranets.

Igualmente contactamos con Micronet (<http://www.micronet.com>) y su oferta consistía en la adquisición completa del programa Knosys para Windows.

Acudimos, por último, a la empresa Baratz, cuya propuesta se cimentó en la posible creación ad hoc de un programa específico

partiendo de Microsoft Access, sin ninguna garantía de control o validación sobre los términos empleados, el rango o las relaciones de los mismos.

Considerando en este punto que las soluciones aportadas por las consultoras de más arraigo en nuestro país no nos podían proporcionar software específico y, sus soluciones eran económicamente inviables para el desarrollo del proyecto decidimos ponernos en contacto con la empresa americana Multites (<http://www.multites.com>) Para conocer su programa de creación, mantenimiento y publicación de tesauros: MultiTes.

El programa está disponible para Windows, precisa unos 50 MB de disco duro para un tesoro de aproximadamente 10.000 términos y 32 MB de RAM. Solicitamos una versión demo para observar que se adecuaba exactamente a nuestras necesidades. La versión sólo existe en inglés por lo que era preciso confirmar que el programa reconocía acentos y grafías no inglesas (ñ).

El programa cuenta además con una aplicación específica para la publicación del tesoro en Internet: MultiCGI, un script CGI que corre únicamente en servidores NT y que resulta incompatible con el servidor de la Universidad de León que corre sobre LINUX con disponibilidad para ejecutar programas en Perl. Esta opción resultó impracticable a pesar de solicitar al Secretariado de Información y Comunicaciones un servidor NT para poner en marcha el proyecto o bien la ayuda de un programador de la Universidad para crear un CGI o tecnología ASP.

Por tanto, nos hemos visto en la obligación de emplear solamente la versión de HTML disponible genéricamente en MultiTes y que posibilita la publicación y consulta navegacional del tesoro pero no la búsqueda de términos del mismo en un formulario de consulta para cuya optimización era preciso el CGI ya comentado. Finalmente, el factor decisivo fue la adecuada relación coste/calidad.

Debemos hacer constar que el programa MultiTes nos ha facilitado sobremanera las tareas mecánicas de construcción del tesoro, ha resultado imprescindible para la validación de relaciones y la consistencia en la asignación de términos, asimismo ha generado automáticamente los correspondientes índices.

3.3 Elaboración de una sede web para albergar el tesoro

Paralelamente se creyó conveniente trabajar en la creación de una sede web que albergase el tesoro que se estaba elaborando y le proporcionase valor añadido. La realización de este trabajo ha sido responsabilidad de las profesoras M^a Luisa Alvite Díez y Blanca Rodríguez Bravo.

Dado que el proyecto se concedió como apoyo a la elaboración de material docente, nos pareció adecuado construir un sitio web para introducir el tesoro en línea y, asimismo insertar contenidos relacionados con los lenguajes documentales que ayuden al alumno a entender el tesoro dentro del más amplio contexto de las herramientas terminológicas.

Atendiendo a esta doble intención se han introducido los siguientes contenidos en la sede web:

- Conceptos sobre lenguajes documentales
- Tesoros en la red organizados en categorías temáticas y evaluados
- Softwares de construcción de tesauros
- Softwares de indización automática
- Lenguajes de etiquetas y su aplicación en los tesauros
- Recursos sobre organización del conocimiento: bases de datos, directorios, instituciones, publicaciones y bibliografía

3.4 Perspectivas de futuro

La última fase del proyecto, noviembre-diciembre de 2001, ha sido la dedicada a incluir esta sede web sobre herramientas terminológicas dentro de la sede oficial del Área de Biblioteconomía y Documentación hospedada en el servidor de la Universidad de León. Sin duda, es esta última fase, la dedicada a la difusión del tesoro, la que proporcionará realmente sentido a este proyecto.

Durante el presente curso académico y en el marco de la docencia de las asignaturas de Lenguajes documentales y Tesoros, se está usando el instrumento terminológico diseñado, y se está evaluando su utilidad en la indización. Ello permitirá apreciar la necesidad de profundizar en los distintos campos, y, por tanto, la conveniencia de incluir descriptores más específicos, o, por el contrario, de eliminar descriptores demasiado concretos e ir a un nivel menor de precisión. El uso puede plantear igualmente la necesidad de intercambiar sinónimos y sustituir un Descriptor por un No

descriptor si se advierte que el último tiene más uso que el primero.

Hay que hacer notar que un trabajo de base terminológica, como el que nos ocupa, y que refleja contenidos científicos, tiene que evolucionar paralelamente con las ciencias cuyos conceptos representa y con sus denominaciones, en nuestro caso las ciencias documentales. Un tesoro exige siempre, por tanto, un trabajo de mantenimiento.

Es por ello que hemos solicitado otra ayuda para el bienio 2002-2003. Nuestra petición se enmarca dentro de los apartados b) y d) del apartado 1.2. de las bases de la convocatoria pública de concesión de ayudas para la *elaboración de recursos de apoyo y experiencias innovadoras en la enseñanza de las universidades* de Castilla y León (Orden de 27 de agosto de 2001, B.O.C. y L. de 7 de noviembre). En el momento actual conocemos su provisional resolución favorable.

4 Objetivos de la actividad a realizar

Los objetivos perseguidos son:

- Realizar los ajustes convenientes al tesoro elaborado y que hemos presentado a la Consejería de Educación y Cultura de la Junta de Castilla y León.
- Ampliar los campos semánticos con contenidos científicos colaterales a las Ciencias de la Documentación: Codicología, Bibliotecnia, Conservación y Restauración e Informática.
- Traducción al inglés con el fin de que el tesoro sea bilingüe, lo que ampliará sus usuarios potenciales.
- Diseño "ad hoc" de un CGI para la consulta del tesoro en Internet, o la adquisición de la aplicación específica que el programa MultiTes tiene para ello: MultiCGI y de un servidor NT donde ésta corra con el fin de permitir la formulación de consultas además de la navegación por sus índices.

5 Metodología y Fases de la actividad por realizar

5.1 Fases de la elaboración intelectual del tesoro:

Se realizará un primer reparto por áreas semánticas entre los participantes con vistas a la extracción y estudio del vocabulario. La primera fase consistirá, por tanto, en la

recopilación del vocabulario. Cada participante en el proyecto se ocupará de un campo temático según los distintos perfiles del profesorado incluido en este proyecto, ya no exclusivo del área de Biblioteconomía y Documentación. Hemos ampliado los miembros participantes con los profesores que imparten asignaturas optativas en la titulación de Biblioteconomía y Documentación, especialistas en los campos nuevos que queremos incluir.

Los procedimientos seguidos para ello serán los mismos utilizados en la construcción del tesoro del que ya partimos. Habrá que prestar especial atención a la correcta integración de los nuevos campos temáticos y su vocabulario con los ya existentes. La utilización del tesoro ya elaborado en la docencia durante el curso 2001-2002 nos orientará sobre las modificaciones más convenientes.

Se añadirán ahora dos nuevas fases: la de traducción de los descriptores al inglés con la ayuda de léxicos y diccionarios especializados, y la de actualización de la versión en línea y la mejora en su consulta con el CGI. La implementación del CGI estará supeditado a la cuantía de la ayuda recibida.

5.2 Fases de la construcción automatizada del tesoro:

En esta segunda fase nos plantearíamos como objetivo el diseño de un CGI para la consulta del tesoro en Internet, o la adquisición de un servidor NT que nos permita utilizar la aplicación específica que el programa MultiTes tiene para ello: MultiCGI. Ello permitiría formular búsquedas en el tesoro. Para asesorarnos con este fin hemos pedido la colaboración de una profesora, Ángela Díez Díez, ingeniera superior informática perteneciente al área de conocimiento de Ingeniería de Sistemas y Automática.

Los resultados se integrarían en la sede web elaborada sobre contenidos relacionados con los lenguajes documentales que ayuden al alumno a entender el tesoro dentro del más amplio contexto de las herramientas terminológicas.

6 Medios personales y materiales

6.1 Medios personales

Formarán parte del equipo de trabajo los profesores del Área de Biblioteconomía y

Documentación y una profesora del Área de Ingeniería de Sistemas y Automática, coordinados y dirigidos por la Profesora Blanca Rodríguez Bravo.

Los profesores que se incorporan por primera vez al proyecto se encargarán de la recopilación del vocabulario de los nuevos ámbitos temáticos. J. C. Pajares de Bibliotecnia, A. López García de Codicología, I. Álvarez Marcos de Conservación y restauración y A. Diez Diez de la Informática.

Los profesores que han realizado la elaboración de la primera fase del tesoro realizarán la supervisión y estudiarán la conveniencia de creación de nuevos campos temáticos o la incorporación del nuevo vocabulario en los ya existentes. Igualmente estudiarán el establecimiento de relaciones entre los descriptores nuevos y los ya introducidos. C. Rodríguez López coordinará el campo temático de Bibliotecnia, J. Gallego Lorenzo el de Codicología, L. Santos de Paz el de Conservación y Restauración y M^a L. Alvite Díez el de Informática.

M^a A. Morán Suárez coordinará la traducción de los descriptores al inglés, y junto con L. Santos de Paz y M^a Luisa Alvite Díez probarán el tesoro elaborado para determinar su grado de efectividad y realizar los ajustes necesarios.

A. Diez Diez estudiará el desarrollo de un CGI o la instalación de un servidor donde corra la aplicación MultiCGI del software MultiTes.

6.2 Medios materiales

La biblioteca del Área de Biblioteconomía y la del Departamento de Ingeniería Eléctrica y Electrónica nos proporcionarán la bibliografía de consulta y de indización para la extracción del vocabulario. Se utilizará también la sección de referencia de la biblioteca central, fundamentalmente los léxicos y diccionarios. Emplearemos igualmente tesoros impresos y accesibles en línea, al igual que bases de datos documentales: LISA, ISA, ERIC, ISOC, Datathèque, Compludoc, etc. y bancos de datos terminológicos: Erodicauton, CORDE, CREA, etc.

Los tesoros en línea y los principales recursos bibliográficos y documentales relacionados con el tema que nos ocupa han quedado ya recogidos en la sede web construida en la primera fase, 2000-2001, de este proyecto.