

Desarrollo de un repositorio de información mediante integración de información superpuesta

Piedad Garrido¹, Jesús Tramullas²

¹ Dpto. Informática e Ingeniería de Sistemas, Escuela Universitaria Politécnica de Teruel, C/ Ciudad Escolar s/n, 44003 Teruel, Spain. piedad@unizar.es

² Dpto. Ciencias de la Documentación, Universidad de Zaragoza, C/ Pedro Cerbuna, 12, 50009 Zaragoza, España. tramullas@unizar.es

Resumen

Las bases de datos híbridas son un desarrollo mixto que permite combinar información estructurada y semiestructurada. El proyecto que se presenta desarrolla una aproximación de este tipo para un servicio de información digital, mediante la integración de diferentes esquemas de etiquetado de metadatos en una base de datos híbrida.

Palabras clave: Bases de datos híbridas, Esquemas de etiquetado, Integración de información, Metadatos.

Abstract

Hybrid databases are a mixed development which allows the structured and non-structured data combination. The project shows an approach of this kind for a digital information service by means of the integration of different labelled metadata schemes in a hybrid database.

Keywords: Hybrid databases, Information integration, Metadata, Tagging schemes.

1 Planteamiento

La herramienta de software libre *Potnia* (Mozilla Public License) (Garrido y Tramullas, 2005) es un sistema de directorio temático especializado (Navarro y Tramullas, 2005), desarrollado con el lenguaje de programación PHP, que utiliza como servidor de bases de datos el motor relacional MySQL. La estructura de descripción de los recursos de información corresponde a la norma ISO 15836 *Dublin Core*. Se distribuye a través de *Sourceforge* (<http://potnia.sourceforge.net>) y de *FreshMeat* (<http://freshmeat.net/projects/potnia>). Se encuentra en desarrollo una nueva versión, que pretende integrar un nuevo repositorio de

información, para lo cual se ha estudiado y evaluado cambiar la tipología de la base de datos hacia bases de datos nativas en XML, bases de datos orientadas a objetos o bases de datos híbridas. Esta contribución expone la evolución del repositorio de información, cuyos fundamentos se ceñían a la integración del estándar propuesto por la DCMI (*Dublín Core Metadata Initiative*) con el paradigma de los *Topic Maps*, y que se ha decidido ampliar a la integración de SKOS (*Simple Knowledge Organization System*) como lenguaje de etiquetado alternativo al Dublín Core y/o complementario. Esta integración ya ha sido probada por autores como Delcambre y Bowers (2000).

2 Fundamentos técnicos

El paradigma de los Topic Maps descrito en la norma ISO 13250:2003 surgió por los trabajos desarrollados por el grupo Davenport hacia el año 1991, con el objetivo de desarrollar un estándar para la elaboración de la documentación técnica del software, por su continuo cambio debido a la aparición en muy poco espacio de tiempo de versiones mejoradas del software desarrollado. Los esfuerzos, en sus comienzos se centraron en hacer posible la automatización de la semántica esencial de los índices analíticos. Posteriormente se generalizó al desarrollo paralelo de iniciativas hacia otros elementos como podían ser las tablas de contenidos, referencias cruzadas, glosarios y tesauros. Se desarrolló un primer borrador por el grupo de SGML de la ISO en 1996, para finalmente en el verano de 1999 aceptar y publicar la primera versión de la norma ISO/IEC 13250: 2000. A comienzos del 2000 se funda una organización independiente, TopicMaps.org, con el objetivo de adaptar esta norma a un lenguaje más adecuado para la Web, ya basándose en las recomendaciones XML. En el 2001, se publicó la primera versión de una DTD para expresar los Topic Maps que fue admitida por la ISO e incorporada como estándar mediante una enmienda técnica. En mayo del 2002 se aprobó y publicó la segunda edición de la norma, recogida finalmente como ISO/IEC 13250:2003, incorporando en su Anexo C, la sintaxis XML.

El modelo intenta proporcionar un esquema de representación de estructuras de conocimiento, en forma de red semántica, y asociarlas con recursos de información, por lo que los topic maps sólo tienen sentido en el medio digital, ellos mismos constituyen un hiperdocumento, que conecta información heterogénea, por lo que son ideales para aplicarlos a grandes colecciones de información en continuo crecimiento. No en vano, Charles Goldfarb definió este paradigma como “GPS del universo de la información”. De ahí que el desarrollo se haya centrado en él para la mejora del repositorio de información de Potnia.

3 Las bases de datos híbridas

Cuando se plantea la utilización de una solución alternativa al modelo de datos relacional, la revisión de las herramientas disponibles muestra como posibilidades el uso de los servidores de ficheros planos, bases de datos documentales, bases de datos relacionales, bases de datos orientadas a objetos, bases de datos nativas en XML y bases de datos híbridas (Pardede, Rahayu y Taniar, 2004). Las características de la información hipermedia distribuida, cada vez más abundante en los contextos de trabajo de la mayoría de la disciplinas, no son adecuadamente tratadas, por limitaciones técnicas o metodológicas, en muchos de ellos, lo que obliga a buscar soluciones mixtas. Cuando se desarrollaron los sistemas de información hipertextual empezaron a surgir unos lenguajes de etiquetado que permitían aportar semántica a un conjunto de etiquetas que hasta el momento habían sido totalmente estáticas, por lo que surgió un nuevo concepto de repositorio de información: las

bases de datos nativas en XML. Una base de datos nativa en XML, ni tiene campos, ni almacena datos atómicos, en sentido tradicional, ya que lo que almacena son documentos XML, de ahí que se les denomine bases de datos centradas en documentos (data-centric databases).

Para migrar el repositorio de información relacional de Potnia, y aunque ninguno de los distintos tipos de bases de datos se ceñía exactamente a lo buscado, se observó que la aproximación que más cercana resultaba era la de las bases de datos nativas en XML. Los objetivos perseguidos con este rediseño del repositorio de información eran el de seguir manteniendo la integridad referencial ofrecida por el modelo relacional, aportar semántica, funcionalidad ofrecida hoy en día por las bases de datos nativas en XML y el último y no por ello menos importante ampliar el conjunto de características descriptivas de la información en tiempo real. Además, se espera poder incorporar prestaciones de las ofrecidas por los sistemas de gestión documental de cara a la recuperación de la información, una vez que la información estuviese organizada mediante un instrumento de organización y representación del conocimiento, por lo que se planteó finalmente hacer uso de una base de datos híbrida (Klemens, Aberer y Klas, 1999).

Las definiciones que se encuentran sobre bases de datos híbridas son dispares, pero el denominador común de la idea básica es el concepto de bases de datos en sí. Se trata de un conjunto de información relacionada que se encuentra agrupada o estructurada sirviendo al fin para el que ha sido diseñada. Hace algunos años se empezó a hablar de bases de datos híbridas como aquellas bases de datos que combinaban características de las bases de datos relacionales y las bases de datos orientadas a objetos, y que manejaban datos textuales y datos binarios.

El diseño de este tipo de bases de datos híbridas, a caballo entre las lexicográficas y terminográficas, trae a colación la noción de recurso multifuncional, ausente en otras perspectivas y que debe ser recuperada por todos aquellos recursos electrónicos que vayan acompañados de un lenguaje de etiquetado, ya que en el abanico de usuarios potenciales se dibuja una gran variedad de tipos de conocimiento, no siempre uniformes. La noción de recurso multifuncional no es nueva. Por ejemplo, Galinski (1988) predijo la utilidad de esta manera de almacenar información. Sager (1990) dibujó una noción para el modelado de datos bastante pragmática, al mismo tiempo que estableció unas normas de uso para implementar dichas bases de datos. Ahmad (1996) describió un sistema actualmente en uso, TransTerm (Dalphin, Brown y Stockwell, 1997) en el que se desarrolla esta misma idea de la base de datos a caballo entre la terminográfica y la lexicográfica. Otro claro ejemplo, se da en el campo de la información y la documentación, donde una base de datos híbrida podría ser una base de datos que mezclara características de una base de datos documental con características de una base de datos relacional. De hecho, numerosas soluciones de gestión documental están incorporando soluciones técnicas basadas en la combinación de ambos enfoques.

Para el trabajo de investigación que se plantea, se consideró necesario evaluar, en primer lugar, la posibilidad de implantar la solución mediante una base de datos nativa en XML, ya que se iba a trabajar con lenguajes de etiquetado basados en metadatos, y comprobar si se cubrían todas las expectativas anteriormente comentadas. De no ser así, se plantearía, como segunda opción, la construcción de un repositorio de información híbrido propio.

4 XTM: uso e integración con SKOS and/or Dublin Core

El paradigma de los topic maps, y en concreto la especificación XTM, es muy amplio, por lo que para llevar a cabo el diseño del nuevo repositorio de información se ha centrado su desarrollo en base a sus tres componentes fundamentales: topic, occurrence and association. Esta tripleta de conceptos fue propuesta por Steve Pepper, en un artículo clásico sobre la cuestión (Pepper, 2002).

Un topic map no es sino la representación material o concreta de la percepción humana o abstracta de la realidad. Puede representar cualquier cosa: personas, entidades individuales o colectivas, conceptos, etc. Cada topic, a su vez es una instancia de una o más clases de topics que pueden indicarse o no de forma explícita. Son portables, por lo que pueden aplicarse a distintos elencos de recursos de información.

Una Ocurrence es cualquier información que es especificada como relevante para un subject dado. En concreto, se trata de recursos externos o internos de información, enlazados mediante una referencia que sirve para su localización y que aclaran o ejemplifican el significado del topic. Las referencias a recursos se realizan mediante URI's según la especificación correspondiente IETF. Estos recursos, como se puede observar en el esquema conceptual, a diferencia de características como el topic name, no se almacenan habitualmente en la descripción básica del topic sino que forman capas separadas, ya que las ocurrences, por sí mismas, pueden constituir un conjunto de información.

Cierra el círculo el concepto de Association. "An association is a relationship between one or more topics, each of which plays a role as a member of that association". El rol que un topic juega en la asociación indica la forma en que participa en ella. Al igual que ocurre con sus anteriores compañeros, las associations se pueden clasificar en clases o association types. La association facilita el poder agrupar topics que tienen la misma relación, por lo que este aspecto afecta directamente a poder implementar interfaces más intuitivas y amigables para navegar por conjuntos grandes de información. Por otra parte, el modelo incorpora otros conceptos que hacen aumentar su potencial, tales como: scope, PSI y facet. El primero de ellos marca el límite de validez de las características asignadas a un topic, minimizando así su ambigüedad (Rath, 2001). El segundo aparece debido a la capacidad de fusión (merge) de topic maps, circunstancia que se da por ejemplo en situaciones multilingües, un ejemplo claro es el que muestra Marc de Graauw (2002):

```
Topic: "Last Name"; PSI: familyname  
Topic: "Achternaam"; PSI: familyname
```

El concepto de facet no fue incluido finalmente en la especificación XTM, y permitía, según su definición original, asignar metadatos a los recursos dentro del propio Topic Map . Esta decisión que fue muy discutida, justificando finalmente su exclusión aduciendo que los recursos podían ser asimismo topics y serles asignadas sus propiedades a través de sus características. Durante el desarrollo de este trabajo se ha concluido que la utilización de subjects direccionables, una solución sencilla y usable desde el punto de vista de diseño de las bibliotecas digitales, no cubría todos los requerimientos planteados, de ahí que acompañando a la descripción de los recursos con los topic maps, se haya decidido completar la información con un lenguaje de etiquetado que permita su intercambio y fusión con información similar que pueda encontrarse distribuida en otros repositorios. Los lenguajes de etiquetado seleccionados son Dublin Core y SKOS.

Tanto XTM como DC y SKOS contemplan el concepto de URI, con distinta notación, los tres reflejan el concepto de que trabajar con subjects y trabajar con una solución combinada de ellos aumenta la semántica de la información representada. La solución combinada hace que no se tenga que pensar en ellos como lenguajes de etiquetado de representación independientes, sino totalmente complementarios, y con un único fin: que existan el mínimo número de barreras posibles en el intercambio de información y la interoperabilidad de sistemas. Todas estas características son ventajosas, ya que los topic maps son un instrumento independiente de los recursos, que se puede utilizar frente a distintas colecciones documentales, heterogéneas y con ubicaciones distintas. Por último, se ha observado que el nivel de independencia que introduce este paradigma entre los recursos de información y el topic map que los relaciona y organiza tiene la importante ventaja de aumentar y enriquecer sus entornos de aplicación y utilidades.

5 Potnia Hybrid DataBase

Antes de desarrollar un diseño ad-hoc, se decidió poner a prueba XMLdb, una base de datos nativa en XML open-source. XMLdb utiliza como sistema gestor de base de datos MySQL, que era el sistema gestor de bases de datos que soportaba la primera versión de Potnia. Resultaba necesario ponerlo a prueba porque así se comprobaría si sería capaz de soportar las etiquetas XTM, SKOS y Dublín Core, ya que uno de los objetivos del nuevo diseño de base de datos es el de conseguir que la información pueda ser etiquetada no sólo en DC sino también en SKOS, o incluso con una mezcla de ambos, partiendo de la premisa de poder usar, para todas las posibles soluciones de etiquetado el mismo repositorio de información.

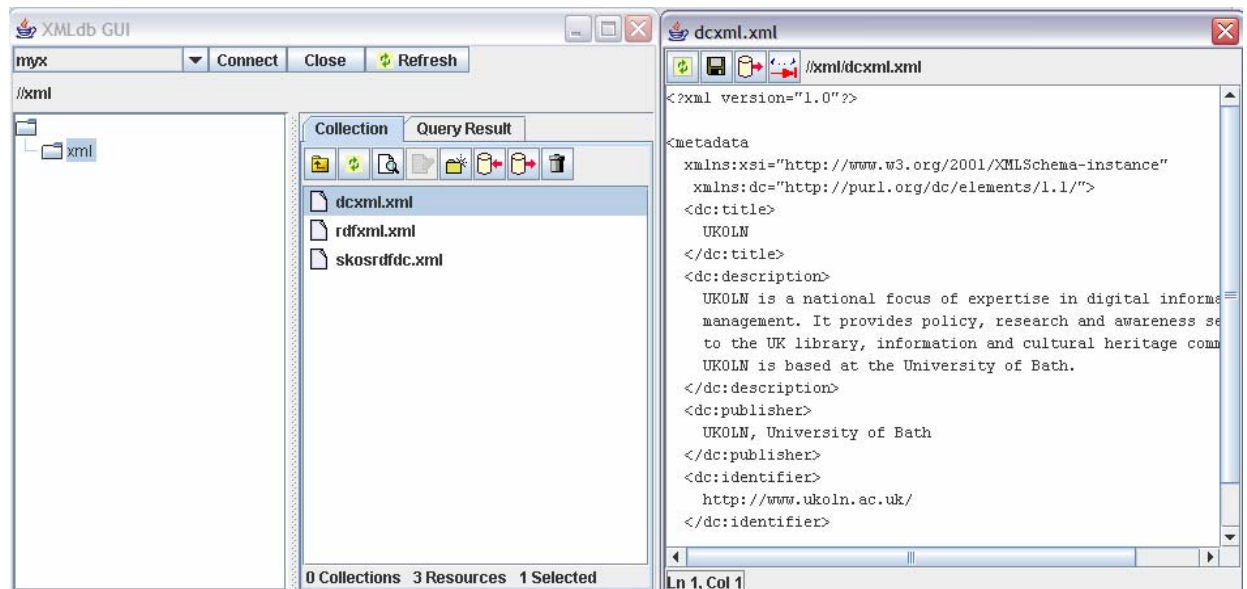


Figura 1. Carga de ficheros en la base de datos myx

Para ello se procedió a la carga de tres tipos de ficheros, tal y como se muestra en la figura 1. Los ficheros xml contenían, respectivamente, la siguiente información:

- dcxml.xml: Información etiquetada con Dublín Core embebido en un esquema XML.

- rdfxml.xml: Información etiquetada en RDF embebido en un esquema XML.
- skosrdfdc.xml: Información etiquetada en RDF y SKOS embebido en un esquema XML.



Figura 2. Comprobación del volcado de información realizado con XMLdb

Una vez comprobado que, tal y como se muestra en la figura 2, era posible soportar SKOS, Dublin Core y RDF, se analizó qué función desempeñaba esta base de datos nativa en XML. XMLdb realmente es una capa que se encuentra por encima del sistema gestor de bases de datos relacional MySQL, y lo que hace es volcar los ficheros con extensión xml que incluyen información etiquetada con metadatos.

El siguiente paso consistía en analizar el proceso de recuperación de información. La información en la primera versión de la herramienta era recuperada analizando los strings mediante sentencias de selección SQL. Para contemplar toda la casuística y la combinación de las distintas situaciones a resolver, se tenían que construir grandes consultas que hacían que el código de la aplicación creciera de forma ingente, y la recuperación de la información ofreciera mucho ruido, debido al amplio margen dejado por los patrones de búsqueda de un modelo de recuperación meramente binario.

Una vez volcados los archivos XML haciendo uso de XMLdb en MySQL se encuentran los siguientes problemas:

- *Almacenamiento*: Una base de datos de estas características almacena la información en formato XML, y en este mismo repositorio se almacenan los índices que se generan por cada documento XML almacenado. La unidad atómica de información con la que se trabaja es el documento en sí, lo que entorpece un nivel específico y totalmente dirigido como el buscado.
- *Procesamiento de datos*: El procesamiento de datos en este tipo de bases de datos es jerárquico, típico de documentos de esta tipología. Esto se debe a que aún no existe un

lenguaje estándar que permita la actualización, inserción o eliminación de elementos de un documento XML. Existe un lenguaje, Xupdate, que permite realizar actualizaciones en un documento XML pero aún no es un estándar y muchos gestores de este tipo no lo soportan.

- *Búsquedas:* Este tipo de bases de datos no utiliza SQL como lenguaje de consulta sino que hacen uso de Xpath. El problema de este tipo de búsquedas es que no permiten construir consultas un poco más elaboradas como por ejemplo, productos cartesianos, concatenaciones u ordenaciones, debido a que Xpath no fue creado realmente para búsquedas en bases de datos, sino simplemente para búsquedas en una sola unidad documental.

En consecuencia, la utilización del concepto de bases de datos nativas en XML no daba respuesta a las necesidades planteadas, ya que la información contenida entre etiqueta de apertura y etiqueta de cierre, no podía ser buscada con el nuevo tipo de base de datos y su lenguaje de consulta alternativo al SQL, el XPath. Se optó, por lo tanto, por diseñar un repositorio de información híbrido (Nicola y Van der Linden, 2005), cuya estructura conceptual se incluye en el siguiente apartado.

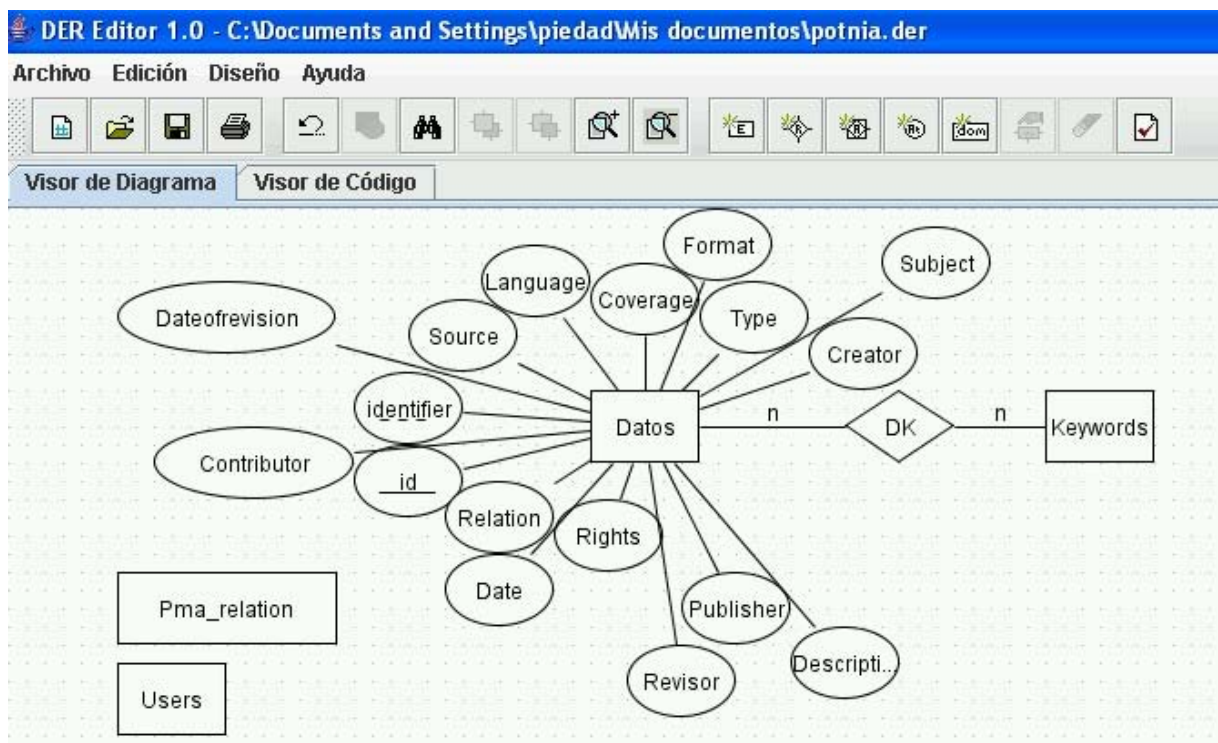


Figura 3. Diseño Inicial con el esquema básico DC (Tramullas y Garrido, 2006)

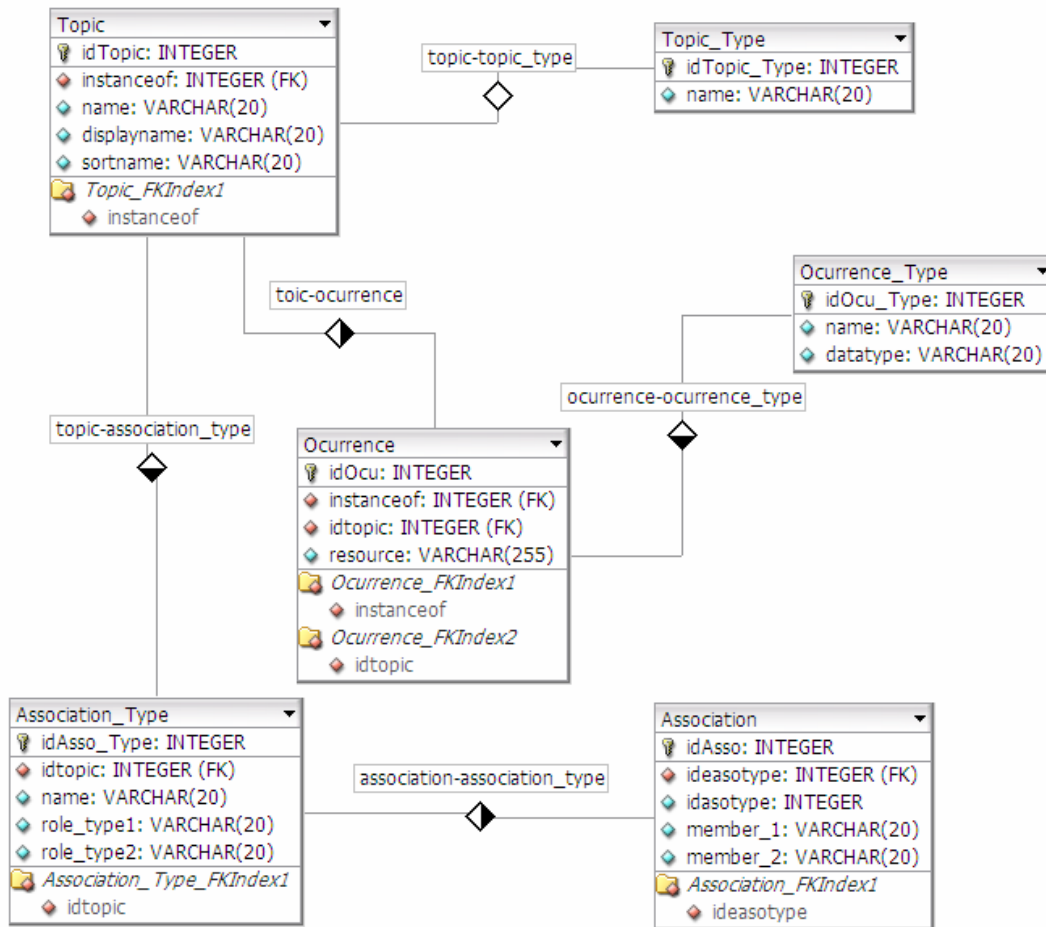


Figura 4. Diagrama E-R extendido

En la versión inicial de la aplicación (figura 3), se planteó el almacenamiento de las quince etiquetas básicas del Dublin Core Simple para describir un recurso de información. Como el planteamiento de la segunda versión parte de la base del poder suministrar valor añadido a la información estática almacenada en la base de datos, se ha optado por ampliar el conjunto de relaciones con las mostradas en la figura 4, pudiendo ampliar la descripción de los recursos informativos en tiempo real, sin tener que modificar el diseño. Esta manera de producir automáticamente topic map desde fuentes de información estructurada ya ha sido comentada de forma teórica por autores como Garshol (2005). Se obtiene así un modelo de datos relacional extendido y renormalizado (Balmin y Papakonstantinou, 2005). Se tiene garantizada una estructura base, que puede ser ampliada muy fácilmente y al momento si se necesitara ampliar el DC simple al DC qualified, o si se quisiera cambiar a SKOS, incluir FOAF, etc. e incluso obtener una solución en la que aparezcan varios de estos lenguajes de etiquetado mezclados (Jiang, et al., 2002).

Una vez definido el esquema de etiquetado, se decidió probar la nueva base de datos híbrida, frente a un directorio temático especializado que almacenara la información relativa al Maestrazgo (Sales, 2002). Para ello el primer paso es acompañar el esquema conceptual anteriormente planteado con la base de datos relacional que se planteó para el desarrollo de este proyecto. En términos de los topic maps, Maestrazgo sería un topic type, como topics se tendrían: AguaViva, Alcorisa, Aliaga, etc. Si se quisiera asignar recursos electrónicos a estos tópicos para construir a specialized subject-gateway, se dispone del mecanismo de las asociaciones, que no son sino relaciones entre dos o más topics, además de operadores de

conjunto del SQL como la UNION que permitirían obtener una completa vista del resultado. Ello no hubiera sido posible si el lenguaje de recuperación de la información hubiera sido XPath. Debe indicarse que los roles son propios de las asociaciones pero se ha considerado necesario representar esta característica en otra relación por claridad en el diseño y facilitar la construcción de una mejor interfaz de usuario.

Associationtype	Role1	Member1	Role2	Member2
Alojamiento	pueblo	AguaViva	hotel	
Alojamiento	pueblo	AguaViva	hostal	
dk	data	title	keywords	teruel

Aunque en el diseño sólo se han reflejado asociaciones binarias (obvio por el fundamento básico del MDR), destacar que los TM permiten asociar miembros con un grado de aridad n. Por último, comentar que para enlazar los topics con recursos externos o internos se hace uso de la información aportada por la relación ocurrencias. La diferenciación se lleva a cabo gracias al atributo datatype, ya que si su valor es URI, automáticamente se identifica como un recurso externo.

El prototipo desarrollado ha permitido dar un nuevo enfoque a la información almacenada en un repositorio relacional (Garrido y Tramullas, 2006). Las mejoras obtenidas con respecto a la versión anterior son las siguientes:

- Mayor expresividad, ya que no se limita la clasificación de la información a un conjunto de atributos estáticos, ni se ciñe la cardinalidad a las combinaciones básicas y estándares ofrecidas por el modelo de datos relacional. Se aprovecha la funcionalidad que ofrece la integridad referencial, que era una de las características básicas que interesaba preservar de este tipo de base de datos.
- Independencia con respecto al lenguaje de etiquetado: La información no se encuentra almacenada en base a un determinado lenguaje de etiquetado sino que puede ser mostrada en varios formatos distintos de manera independiente.
- Mejora en las búsquedas. La búsqueda se puede centrar en el concepto de topic, lo que era impensable desde el punto de vista relacional de la primera versión. Se puede así diseñar un buscador mucho más completo que el de la primera versión, que sin necesidad de hacer una búsqueda secuencial dirigida a texto completo, recupere la información de manera más eficaz, pudiendo incorporar características de valor añadido como pueden ser técnicas de soft-computing.

La solución, entre otros objetivos, plantea acompañar la información básica con un lenguaje de etiquetado que permita un proceso alternativo de búsqueda más rápido y eficiente. Además, pone los fundamentos para los siguientes desarrollos de la herramienta software:

- Interfaz gráfica de representación de la información dimensional: En la primera versión, esta representación estaba limitada a ser textual. Si la información es representada en base a unos topic, se puede llegar a hacer una representación tridimensional de la información.
- Permitir almacenar nueva información descriptiva de un recurso de información en tiempo real sin tener que modificar el diseño inicial del repositorio de información, lo que lo convierte en un sistema flexible y multidisciplinar. Al ser XTM independiente de XML, realmente se trata de un add-in, un complemento que añade un valor que

Xml no puede dar. Por lo tanto no se está obligado a representar este tipo de información en una base de datos nativa en XML, y resulta mucho más adecuado optar por una solución híbrida.

Bibliografía citada

- AHMAD, K. [et al.]. A case for a linguistically-informed terminology database. En: *TKE '96. Terminology and knowledge engineering*. Frankfurt: Indeks Verlag, 1996.
- BALMIN, A.; PAPAKONSTANTINOY, Y. Storing and querying XML data using denormalized relational database. *The VLDB Journal: The International Journal on Very Large Data Bases*, 2005, vol. 14, n. 1, p. 30-49.
- BOWERS, S.; DELCAMBRE, L. *A generic approach for representing model-based superimposed information* [recurso electrónico]. Beaverton, OR: Oregon Health and Science University, 2000. <<http://www.cse.ogi.edu/tech-reports/2000/00-008.pdf&ei=MIWkQ4rAJKKQiALG56ydBw&sig2=qycTahRJnfO2HwaZ5mGX3g>>. [Consultado: 12 dic. 2005]
- DALPHIN, M. [et al.]. The translational signal database, TransTerm, is now a relational database. *Oxford Journals*, 1997, vol. 26, n. 1, p. 335-337.
- GALINSKI, C. Advanced terminology banks supporting knowledge-based MT: some reflections on the costs for setting up and operating a terminological data bank (TDB). En: *Conference of New Directions in Machine Translation: proceedings*. Dordrecht: Foris, 1988, p. 1-15.
- GARRIDO, P.; TRAMULLAS, J. Potnia: una herramienta para directorios temáticos basada en Dublín Core y Topic Maps. En: ISKO España (7º: Barcelona: 2005). *La dimensión humana de la organización del conocimiento: actas*, p. 238-248.
- GARRIDO, P.; TRAMULLAS, J. Semantic information stored in an extended denormalized database. En: *International Conference on Multidisciplinary Information Sciences & Technologies: proceedings (1ª: Mérida: 2006)*, p. 485-489.
- GARSHOL, L. M. *SKOS in Topic Maps*, 2005 [recurso electrónico] <<http://www.garshol.priv.no/blog/10.html>>. [Consultado: 4 jun. 2006]
- GRAAUW, M. *Bussiness Maps: Topic Maps Go B2B*. XML.com, 2002 [recurso electrónico] <<http://www.xml.com/pub/a/2002/08/21/topicmapb2b.html>> [Consultado: 13 ago. 2004]
- JIANG, H. [et al.]. XParent: an efficient RDBMS-Based XML Database System. En: *International Conference on Data Engineering (ICDE'02): proceedings (18th: 2002)*, p. 335-336.
- KLEMENS, B.; ABERER, K.; KLAS, W. Building a hybrid database application for structured documents. *Multimedia Tools and Applications*, 1999, vol. 8, n. 1, p. 65-90.
- NAVARRO, D.; TRAMULLAS, J. Directorios temáticos especializados: definición, características y perspectivas de desarrollo. *Revista Española de Documentación Científica*, 2005, vol. 28, n. 1, p. 49-61.
- NICOLA, M.; VAN DER LINDEN, B. Native XML support in DB2 universal database. Industrial session: XML support in relational system. En: *International Conference on*

Very Large Data Bases VLDB '05. VLDB Endowment: proceedings (31st: Trondheim, Norway: 2005).

- PARDEDE, E.; RAHAYU, W.; TANIAR, D. Database theory, technology and applications (DTTA): on using collection for aggregation and association relationships in XML object-relational storage. En: *ACM Symposium on Applied Computing: proceedings (Nicosia, Cyprus: 2004)*, p. 703-710.
- PEPPER, S. *The TAO of Topic Maps: finding the way in the age of infoglut*, 2002 [recurso electrónico] <<http://www.ontopia.net/topicmaps/materials/tao.html>>. [Consultado: 3 may. 2005]
- RATH, H. H. Semantic resource exploitation with Topic Maps. En: LOBIN, H. (ed.) *Sprach- und Texttechnologie in digitalen Medien. Frühjahrstagung der Gesellschaft für linguistische Datenverarbeitung (GLDV)*. Norderstedt, Germany: Books on Demand, 2001. p. 3 -15.
- SAGER, J. C. *A practical course in terminology processing*. Amsterdam: John Benjamins, 1990.
- SALES, E.; GARRIDO, P.; MARTÍNEZ, F. J. Servicios y recursos turísticos del Maestrazgo on-line: patrimonio cultural al alcance de todos. En: *Conferencia Europea de Empleo y Patrimonio Cultural, Promoción Económica y Nuevas Tecnologías en la Sociedad de la Información y del Conocimiento: actas (3ª: Madrid: 2002)*. Madrid: Universidad Complutense de Madrid, 2002.
- TRAMULLAS, J.; GARRIDO, P. Constructing Web subject gateways using Dublin Core, RDF and Topic Maps [recurso electrónico]. *Information Research*, 2006, vol. 11, n. 2 <<http://informationr.net/ir/11-2/paper248.html>>. [Consultado: 4 mar. 2006]