

Sistema de consulta vía Web para el Instituto Andaluz de Patrimonio Histórico

Nieves R. Brisaboa¹, M. José Escalona², Manuel Mejías²,
Ángeles S. Places¹, Francisco J. Rodríguez¹, Jesús Torres²

¹Departamento de Computación
Facultad de Informática
Universidade da Coruña

²Departamento de Lenguajes y Sistemas
Facultad de Informática y Estadística
Universidad de Sevilla

brisaboa@udc.es, {escalona, risoto}@lsi.us.es,
{asplaces, franjrm}@mail2.udc.es, jtorres@lsi.us.es

Resumen. En este artículo se presenta un sistema de consulta vía Web que está siendo desarrollado para el Instituto Andaluz de Patrimonio Histórico. Dicho sistema utiliza una arquitectura, basada en ontologías, para la consulta vía Web de Bibliotecas Digitales. Dicha arquitectura proporciona independencia física y lógica a la aplicación de consulta frente a la base de datos.

La interfaz de usuario utiliza la estrategia del Lenguaje Natural Acotado (BNL), por lo que se garantiza la facilidad de su uso.

Palabras Clave. Ontologías, Interfaces de Usuario, Web.

1 Introducción

En este artículo se describe un sistema de información que estamos desarrollando para el Instituto Andaluz de Patrimonio Histórico. Dicho sistema permitirá gestionar, explotar y difundir el rico patrimonio histórico del que dispone la comunidad andaluza. En este proyecto estamos colaborando un grupo de investigación de la Universidad de Sevilla y el Laboratorio de Bases de Datos de la Universidad de A Coruña.

El objetivo de este proyecto es, no sólo mejorar la funcionalidad de las actuales aplicaciones con las que trabajan los arqueólogos, arquitectos y etnólogos del Instituto Andaluz de Patrimonio Histórico, sino, además, proporcionar al público en general un medio de acceso vía Web para que sea posible consultar toda la información catalogada por el instituto. Este artículo se centra en este segundo aspecto, es decir, en

Trabajo parcialmente subvencionado por la CICYT (TEL99-0335-C04-02), por la Xunta de Galicia (PGIDT99PXI10502A) y por la CICYT y los fondos FEDER (TIC2000-1673-C06-03).

el desarrollo de un sistema de información que permita el acceso a través de Internet a la información del Instituto Andaluz de Patrimonio Histórico.

La arquitectura que se va a usar en el desarrollo del sistema de información del Instituto Andaluz de Patrimonio Histórico es la propia de una biblioteca digital ya que la información que se va a manejar incluye no sólo datos estructurados sino también información no estructurada (por ejemplo, gráficos, fotografías o documentos).

En el Laboratorio de Bases de Datos de la Universidad de A Coruña, se diseñó una arquitectura basada en ontologías para la consulta y recuperación de información de bibliotecas digitales. Dicha arquitectura permite que una biblioteca digital cumpla dos requisitos de diseño que consideramos fundamentales para asegurar el éxito de cualquier sistema de información documental que vaya a ser accedido vía Web: (1) Presentar una interfaz de usuario intuitiva y fácil de usar y (2) Proporcionar independencia de la base de datos subyacente. Se decidió, por tanto, utilizar esta arquitectura para desarrollar el nuevo sistema de información del Instituto Andaluz de Patrimonio Histórico. Se detalla a continuación el interés de las dos características antes citadas:

1. Interfaz de usuario intuitiva: La interfaz de usuario debe ser lo suficientemente intuitiva como para que un usuario pueda usarla, desde su primer acceso, sin tener que leer, previamente, complicadas instrucciones de uso y sin tener que asistir a ningún tipo de curso de formación. Este es un aspecto crucial para el éxito de la biblioteca digital ya que sólo si los usuarios encuentran la interfaz sencilla, flexible y fácil de usar van a estar cómodos con ella y, por lo tanto, van a acceder a la biblioteca digital.

Como se explicará a lo largo del trabajo, nuestra arquitectura permite construir una interfaz de estas características. Para ello se usa la llamada técnica del *Lenguaje Natural Acotado (BNL)* [6]. Esta técnica consiste, básicamente, en presentar una colección de frases con huecos que el usuario debe rellenar para expresar su consulta. El éxito de esta técnica está avalado por varios años de uso de la misma en el marco del proyecto CICYT TEL96-1390-C02-02 [1].

2. Independencia lógica y física de la interfaz de usuario con respecto a la base de datos: La independencia asegura que modificaciones en el esquema de la base de datos o el cambio del Sistema de Gestión de la Base de Datos provocarán un impacto mínimo en el sistema.

En nuestra arquitectura, la independencia se consigue con la utilización de ontologías. Tanto la generación de la interfaz de usuario como la generación de las consultas finales para la base de datos son procesos que usan la ontología para guiar su ejecución. De esta forma, es posible cambiar el comportamiento de ambos procesos sin necesidad de describir su código. Simplemente habrá modificar la ontología.

El resto del artículo se estructura como sigue. En la sección 2 se describe la información que actualmente gestiona el *Instituto Andaluz de Patrimonio Histórico*. En la sección 3 se describen las fases del proyecto global en el que está enmarcado este artículo. En la sección 4 se describe la arquitectura que se va a usar para el desarrollo del sistema de información del *Instituto Andaluz de Patrimonio Histórico*. En la sección 5 se describen las ontologías del sistema. En la sección 6 se describen los esqueletos de frase que permiten construir las frases en *Lenguaje Natural Acotado*. En la sección 7 se describe, en general, el proceso de consulta a la base de datos y, en particular, el algoritmo que usa el sistema para construir la interfaz de usuario.

2 Ámbito del problema

El *Instituto Andaluz de Patrimonio Histórico (IAPH)* es un organismo dependiente de la *Dirección General de Bienes Culturales* de la *Consejería de Cultura* de la *Junta de Andalucía*. Desde hace más de diez años en el IAPH se cataloga, gestiona y difunde toda la información referente al patrimonio histórico andaluz.

2.1 Información que integrará el sistema de información del IAPH

Toda la información que se gestiona en el IAPH está referida al concepto *bien*. Los bienes se clasifican en dos grandes grupos:

- **Bienes muebles:** una vasija, una pintura o una escultura son ejemplos de bienes muebles.
- **Bienes inmuebles:** una catedral o un yacimiento concreto son ejemplos de bienes inmuebles. Existen tres áreas de investigación dedicadas al estudio de los bienes inmuebles: la Arquitectura, la Arqueología y la Etnología.

Aunque existen ciertos datos específicos para los bienes inmuebles y otros exclusivos para los bienes muebles, en general, para todos ellos se almacena la misma información. Dependiendo de su naturaleza, esta información se agrupa en varios bloques:

- **Información de identificación y localización:** este bloque de información está compuesto por datos como el código que impone al bien la *Dirección General de Bienes Culturales*, la denominación general que se le da al bien o el municipio y la provincia en la que se encuentra ubicado el bien.
- **Información de descripción:** son datos que describen el bien de forma detallada. Algunos ejemplos son los siguientes: periodo histórico, estilo, escuela, iconografía, etc.
- **Información de análisis:** recoge información derivada del análisis químico y físico del bien: material de composición, material utilizado en su construcción, etc.
- **Información sobre la protección:** los organismos públicos asignan a cada bien un grado de protección que aumentan a medida que el interés cultural del bien crece. En este bloque se recoge información sobre los estados de protección por los que pasa un bien, el estado de protección actual y las referencias a las resoluciones en el *Boletín Oficial del Estado* o el *Boletín Oficial de la Junta de Andalucía* que dictan cada cambio de estado de protección.
- **Información sobre la conservación:** cada bien es sometido a diferentes estudios sobre el estado de conservación en el que se encuentra y sobre los deterioros potenciales que puede sufrir.
- **Información sobre las restauraciones:** este bloque de información almacena datos relativos a las restauraciones por las que pasa un bien.
- **Información documental:** está formada por las referencias de los documentos relativos a un bien.
- **Información bibliográfica:** contiene los datos bibliográficos de los libros que tratan sobre un bien.
- **Información gráfica:** se trata de documentos gráficos en los que aparece el bien (fotos, vídeos, etc.).

2.2 Sistema actual del IAPH

Aunque, como hemos comentado, básicamente se necesitará el mismo conjunto de atributos para describir todos los bienes independientemente del tipo que sean (Bienes muebles o Bienes inmuebles), el sistema actual del IAPH está compuesto por ocho bases de datos independientes con aplicaciones específicas que las manejan. En la Fig. 1 se presenta un esquema del sistema actual del IAPH y a continuación se describen las bases de datos y las aplicaciones que lo componen.

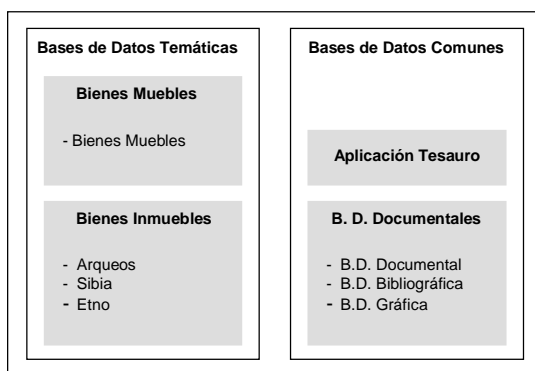


Fig. 1. Sistema actual del IAPH

- **Base de datos de *Bienes muebles*:** base de datos que almacena toda la información recogida sobre bienes muebles, excepto la información que hemos denominado bibliográfica, documental y gráfica en la clasificación anterior.
- **Bases de datos para bienes inmuebles:** Existen actualmente tres bases de datos. Cada una de ellas almacena igualmente toda la información recogida sobre bienes inmuebles, excepto, igual que en el caso anterior, la información que hemos denominado bibliográfica, documental y gráfica. Las tres bases de datos y sus aplicaciones son las siguientes: (1) *ARQUEOS*: usada por los investigadores de arqueología del IAPH para almacenar información relativa a aquellos bienes inmuebles con valor arqueológico. (2) *SIBIA*: usada por los investigadores de arquitectura del IAPH para almacenar información relativa a aquellos bienes inmuebles que presentan valor arquitectónico. (3) *ETNO*: usada por los investigadores de etnología del IAPH para almacenar información relativa a aquellos bienes inmuebles que poseen cierto valor etnológico.

Es importante destacar que existen bienes inmuebles que son importantes para más de un área de investigación (arqueológica, arquitectónica y etnológica). Esto provoca que, muchas veces, la misma información esté repetida en las tres bases de datos, generándose de esta forma redundancia y los consecuentes problemas de consistencia y dificultades de mantenimiento.

Estas cuatro bases de datos se han llamado bases de datos temáticas aún cuando los atributos que almacenan sobre los bienes son, básicamente, los mismos independientemente de si el bien es mueble o inmueble e independientemente, también, de quién lo haya estudiado (arqueólogos, arquitectos o etnólogos).

Además de estas cuatro bases de datos, existen otras cuatro que se han llamado bases de datos comunes:

- **Base de datos documental:** almacena la información documental de todos los bienes.
- **Base de datos bibliográfica:** almacena la información bibliográfica de todos los bienes.
- **Base de datos gráfica:** almacena la información gráfica de todos los bienes.
- **Base de datos de tesauros:** en cada bloque de información existen campos que el usuario puede rellenar libremente, pero hay atributos cuyos posibles valores son términos que están normalizados y almacenados en un tesauro.

A la vista de los graves problemas de diseño de las bases de datos y aplicaciones que actualmente se usan en el IAPH, se decidió abordar un proyecto global de reingeniería para diseñar un sistema de información centralizado y robusto que sostenga la información actualmente gestionada por el instituto, que evite la redundancia y que sea la base unificada de las actuales aplicaciones de gestión (*Arqueos, Sibia, Etno y Bienes Muebles*) que verán también mejoradas algunas de sus funcionalidades.

Otro de los objetivos marcados en este proyecto global es el diseño y desarrollo de un sistema de consulta vía Web que ponga a disposición del público en general toda la información sobre el patrimonio histórico andaluz catalogada por el IAPH.

3 Fases del proyecto

La ejecución del proyecto global se va a llevar a cabo en cinco fases. A continuación se describen de forma muy básica estas cinco fases:

– **Fase 1: Análisis del sistema actual:**

Durante esta fase, que ha sido ya llevada a cabo, se ha estudiado el sistema actual del IAPH. En el apartado anterior se ha descrito de forma general la información gestionada por el IAPH, así como las bases de datos y las aplicaciones existentes actualmente.

– **Fase 2: Diseño del nuevo modelo de datos**

Se ha diseñado un modelo de datos en donde se podrá almacenar de forma conjunta y no redundante toda la información que actualmente gestiona el IAPH sobre los bienes.

– **Fase 3: Volcado de los datos al nuevo modelo de datos**

Todos los datos que actualmente están almacenados en bases de datos diferentes tienen que ser volcados a la nueva base de datos que sigue el modelo conceptual diseñado durante la fase anterior.

Esta fase no es objeto de este artículo, por lo que no se describe aquí.

– **Fase 4: Reingeniería de las aplicaciones actuales**

Durante esta fase, se adaptarán las aplicaciones informáticas actuales (*Bienes Muebles, ARQUEOS, SIBIA y ETNO*) dedicadas a la gestión (altas, bajas y modificaciones) de los bienes al nuevo esquema de la base de datos. Con el nuevo esquema conceptual, estas aplicaciones aumentarán su rendimiento y mejorarán algunas de sus funcionalidades.

Al igual que la fase anterior, esta fase no es objeto de este artículo, por lo que no se describe aquí.

– **Fase 5: Diseño de las interfaces vía Web.**

El objetivo de esta fase es diseñar e implementar las aplicaciones necesarias para permitir el acceso a través de Internet a la información gestionada por el IAPH. De esta forma, se pondrá a disposición del público en general toda la información sobre el rico patrimonio histórico que posee la comunidad andaluza.

En esta fase, que se está llevando a cabo en paralelo con las fases tres y cuatro, se centra el artículo que se presenta.

4 Arquitectura del Sistema

La arquitectura que está siendo aplicada es la diseñada por el Laboratorio de Bases de Datos. Esta arquitectura está completamente descrita en [4, 5]. En este trabajo se presenta la utilización concreta de la arquitectura general al sistema de información del Instituto Andaluz de Patrimonio Histórico.

La arquitectura del sistema, mostrada en la Fig. 2, está formada por tres capas diferentes, que se comunican mediante tres lenguajes de intercambio de datos definidos a tal efecto. Estas tres capas se describen a continuación:

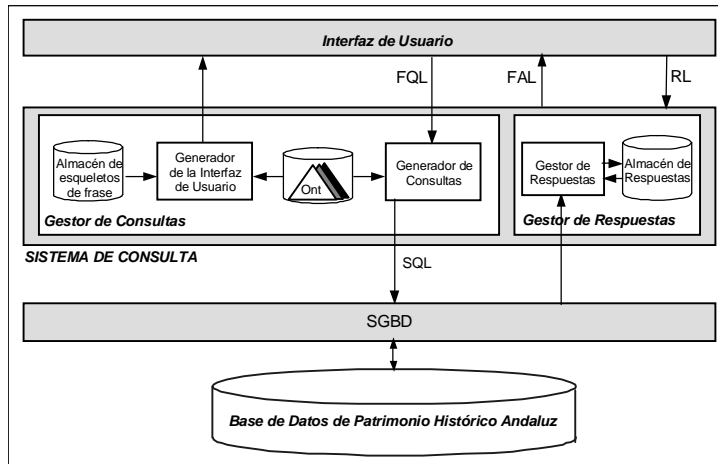


Fig. 2. Arquitectura del Sistema

- **CAPA 1. Interfaz de Usuario:** La interfaz de usuario, sencilla, fácil de usar y basada en la técnica *Lenguaje Natural Acotado*, se genera automáticamente en cada acceso del usuario. Como se explicará en secciones posteriores, el *Generador de la Interfaz de Usuario*, que es un módulo situado en la capa 2, genera la interfaz de usuario guiado por la ontología.
- **CAPA 2. Sistema de Consulta:** El *Sistema de Consulta* está formado por dos grandes módulos. Por un lado, está el *Gestor de Consultas*, encargado de los procesos necesarios para construir y enviar la consulta a la base de datos y, por otro lado, el *Gestor de Respuestas* encargado de recibir las respuestas de la base de datos y presentárselas al usuario de forma conveniente facilitando la navegación del usuario por el conjunto de datos recuperados. A continuación se describen estos

dos módulos. Como este trabajo está centrado en la parte de consulta del sistema, el *Gestor de Respuestas* sólo está descrito de forma básica.

- **Gestor de Consultas:** El *Gestor de Consultas* está formado por dos módulos principales: el *Generador de la Interfaz de Usuario* que se encarga de construir la interfaz de usuario y el *Generador de Consultas*, encargado de recibir la consulta del usuario y traducirla a SQL, lenguaje que usa la base de datos del IAPH.

La ejecución de ambos módulos del *Gestor de Consultas* está guiada por las ontologías del sistema y, además, el *Generador de la Interfaz de Usuario* usa los esqueletos de frases para construir la interfaz de usuario. Por esta razón, el funcionamiento de estos dos módulos será explicado en detalle en la sección 8, después de haber descrito completamente cómo son, tanto las ontologías de nuestro sistema, como la estructura de los esqueletos de frases.

- **Gestor de Respuestas:** Este módulo es el encargado de presentar al usuario las respuestas obtenidas desde la base de datos. Básicamente, el *Gestor de Respuestas* funciona como sigue.

Una vez ejecutada la consulta, la base de datos devuelve, como primera respuesta, los códigos DGBC (código asignado a los bienes por la *Dirección General de Bienes Culturales*) y la denominación general de aquellos bienes que cumplen las restricciones expresadas por el usuario. Estos códigos y sus denominaciones son almacenados por el *Gestor de Respuestas* en el *Almacén de Respuestas*, y presentados al usuario en la primera página de respuesta. Los demás datos, imágenes, documentos, etc. asociados a cada bien son presentados al usuario bajo demanda. De esta forma, se evitan los costes implicados en el envío de enormes cantidades de información, que en muchas ocasiones no interesa al usuario, evitando la saturación del sistema. Por lo tanto, a partir de la primera página de respuesta, el sistema pone a disposición del usuario elementos gráficos que le permitan solicitar la información que necesite sobre un bien determinado, que le permitan consultar el siguiente bien recuperado, etc. Es decir, elementos gráficos que le permitan navegar a través de los bienes recuperados y de toda la información asociada a los mismos, de una manera fácil e intuitiva. Para cada petición del usuario, el *Gestor de Respuestas* obtendrá de la base de datos la información que necesite a través el código del bien almacenado en el *Almacén de Respuestas*.

Como hemos comentado, este es el funcionamiento básico del *Gestor de Respuestas*. Un estudio más detallado de este módulo debería contemplar temas como tiempo límite de espera por las respuestas de la base de datos, mantenimiento de la sesión del usuario, etc. Sin embargo, el *Gestor de Respuestas* no es objeto de este artículo, por lo que no se detalla aquí.

- **CAPA 3. Base de Datos:** Nos referimos a la nueva base de datos de patrimonio histórico andaluz una vez que se termine la migración desde las actuales bases de datos. El mantenimiento de esta base de datos será llevado a cabo por las nuevas aplicaciones de gestión del IAPH, por lo tanto no es responsabilidad del sistema de consulta vía Web que se describe en este artículo.

La comunicación entre la *Interfaz de Usuario* y el *Sistema de Consulta* se realiza mediante tres lenguajes de intercambio de datos que se han definido para este fin. Todos ellos han sido creados en XML y la función de cada uno de ellos es la siguiente:

- **FQL** (*Lenguaje Formal de Consulta*): Es un lenguaje para dar formato a las restricciones de consulta expresadas por el usuario. El usuario construye la consulta expresando restricciones sobre los conceptos de la ontología, por lo tanto, el lenguaje FQL depende de la ontología.
- **FAL** (*Lenguaje Formal de Respuesta*): El lenguaje FAL proporciona el formato en el que se van representar los datos asociados a un bien recuperado por el sistema. El *Gestor de Respuestas* almacena en este formato la información recuperada de la base de datos para un bien concreto, antes de enviarla a la *Interfaz de Usuario*. De esta forma, puede ser interpretada y presentada al usuario de forma conveniente.
- **RL** (*Lenguaje de Peticiones*): El lenguaje de formato RL es el formato en el que se representan las peticiones que efectúa el usuario para ver el siguiente bien recuperado o más datos sobre el bien que esté visualizando en ese momento.

5 Ontologías en nuestro sistema

Para entender el funcionamiento del *Generador de la Interfaz de Usuario* y del *Generador de Consultas* es necesario comprender claramente cómo son las ontologías del sistema.

Existen diferentes definiciones del término ontología [7, 8, 9] pero, básicamente, todas ellas coinciden en que una ontología es una especificación de una conceptualización, es decir, un conjunto de conceptos y las relaciones entre ellos. De este modo, una ontología describe un dominio de interés.

En nuestro sistema, las ontologías son una conceptualización del esquema de la base de datos de patrimonio histórico, pero no otra representación formal del modelo Entidad-Relación de la base de datos. En las ontologías aparecen únicamente aquellos conceptos sobre los que el usuario puede expresar las restricciones de su consulta. La elección de estos conceptos responde, como ya hemos comentado anteriormente, a un documento que recoge las necesidades de consulta de los usuarios y que ha sido obtenido durante fase número uno del proyecto global, la *Fase de Análisis*.

Los conceptos de la ontología son términos usados en el mundo real. Por eso, la ontología proporciona un claro y comprensible modelo de datos para los usuarios. Existen dos tipos de relaciones entre los conceptos que forman nuestras ontologías: (1) Relaciones de *Generalización/Especialización* (relaciones “*is a*”): por ejemplo, una obra es un libro, un recopilatorio, una revista, una tesis, etc. (2) Relaciones de *Descripción* (relaciones “*tiene*”): por ejemplo, una persona tiene un nombre, una o más direcciones, uno o más números de teléfono, etc.

Cada concepto hoja de las ontologías, es decir, cada concepto de las ontologías que no presente ninguna de las relaciones anteriores (*is a* o *has*), tiene asociados dos tipos de información:

1. **El *Id* del esqueleto de frase**, necesario para generar la frase en *Lenguaje Natural Acotado* que permitirá usuario expresar restricciones sobre dicho concepto. La sintaxis de los esqueletos se describe en la sección 6. Solamente los conceptos hoja tienen asociados esqueletos de frase. Los conceptos no hoja son tratados por el algoritmo que genera la interfaz de usuario de forma diferente, como se explicará en la sección 7.
2. **Expresión para recuperar los datos de la base de datos**. Es decir, la expresión necesaria para construir la parte de la sentencia SQL en la que se indican las

restricciones que tiene que cumplir el bien, en relación a ese concepto, para ser recuperado. Habitualmente, la expresión asociada a un concepto es el nombre del atributo y el nombre de la tabla en donde están almacenados los valores de ese concepto, pero podría ser una expresión en álgebra relacional más compleja, por ejemplo, cuando el valor del concepto debe calcularse sobre columnas procedentes de tablas diferentes, sobre las que se hay que hacer antes una operación de *join*.

Las ontologías están definidas en XML. XML permite definir formatos y es un lenguaje universal para dar estructura a los documentos y los datos. Entre las posibilidades de representar la ontología hemos elegido XML porque este lenguaje nos permite definir formatos, para intercambio de datos y metadatos, que es fácilmente legible por humanos.

En nuestro sistema disponemos de dos ontologías: una de las ontologías, presentada en la Fig. 3, está orientada a consultas en las que se busca un bien concreto cuya denominación o código ya es conocido por el usuario, mientras que la segunda ontología, que aparece en la Fig. 4, está orientada a búsquedas de conjuntos de bienes que tengan ciertas características en común.

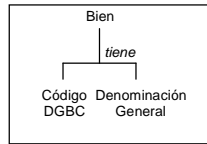


Fig 3. Ontología para la búsqueda de un bien concreto

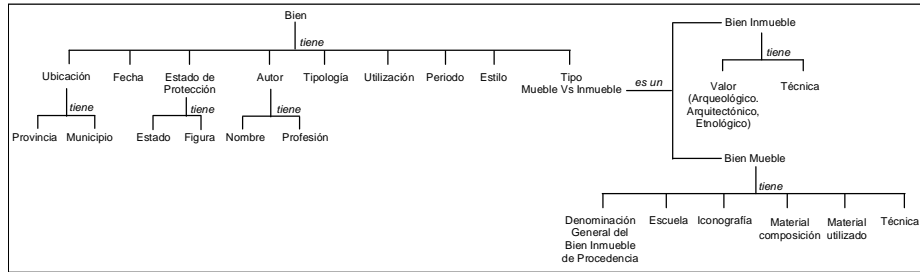


Fig. 4. Ontología para la búsqueda de un conjunto de bienes con características comunes

6 Esqueletos de frase

En esta sección, se describen los esqueletos de frase que proponemos para nuestro sistema. Todos los esqueletos de frase contienen la etiqueta <CONCEPTO>, que el sistema sustituye por el nombre del concepto de la ontología que el usuario va a restringir en un momento dado del proceso de consulta.

El esqueleto de frase se convierte, de esta forma, en una frase en *Lenguaje Natural Acotado* (BNL) lista para presentar al usuario. La frase en *Lenguaje Natural Acotado* contiene huecos que se tienen que rellenar por el usuario. Algunas veces se rellenan marcando una posibilidad de entre varias, otras veces marcando varias posibilidades,

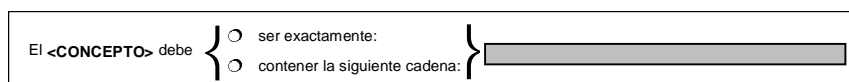
algunas veces escribiendo un valor en un campo de edición (representados en las figuras por rectángulos) y, en otras ocasiones, seleccionando un valor de una lista de posibles valores que ofrece el sistema (para aquellos atributos que toman valores del tesauo).

Como se ha comentado en la sección 5, cada concepto hoja de la ontología tiene asociado el identificador del esqueleto de frase adecuado para permitir al usuario expresar restricciones sobre dicho concepto. A continuación se presentan los esqueletos que se usarán inicialmente en el sistema.

6.1 Esqueletos para expresar restricciones

Los conceptos de tipo de dato *Cadena de caracteres* tendrán asociado el esqueleto presentado en la Fig. 5 o el esqueleto presentado en la Fig. 6. El que un concepto tenga asociado un esqueleto u otro depende de si el atributo correspondiente toma, únicamente, valores que aparecen en el tesauo o cualquier otra cadena de caracteres.

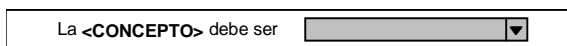
El esqueleto de la Fig. 5 está asociado a los conceptos *Código DGBC* y *Denominación General* de la ontología de la Fig. 3 y a los conceptos *Nombre*, *Profesión*, *Denominación General del Bien Inmueble de Procedencia* de la ontología de la Fig. 4.



El <CONCEPTO> debe ser exactamente: contener la siguiente cadena:

Fig. 5. Esqueleto de frase para el tipo de dato *Cadena de caracteres*

Los conceptos *Provincia*, *Municipio*, *Estado*, *Figura*, *Utilización*, *Periodo*, *Estilo*, *Técnica*, *Material de composición* y *Material utilizado* de la ontología de la Fig. 4 tienen asociado el esqueleto que se presenta en la Fig. 6. En este caso, además de sustituir la etiqueta <CONCEPTO> por el nombre del concepto que se esté tratando, también rellenará la lista desplegable con los valores del tesauo que puede tomar dicho concepto.



La <CONCEPTO> debe ser

Fig. 6. Esqueleto de frase para el tipo de dato *Cadena* que toma valores del tesauo.

Por ejemplo, si el concepto actual fuese *Periodo*, el esqueleto de la Fig. 6 (que es el que tiene asociado este concepto) se convertiría en la frase en *Lenguaje Natural Acotado* que aparece en la Fig.7.

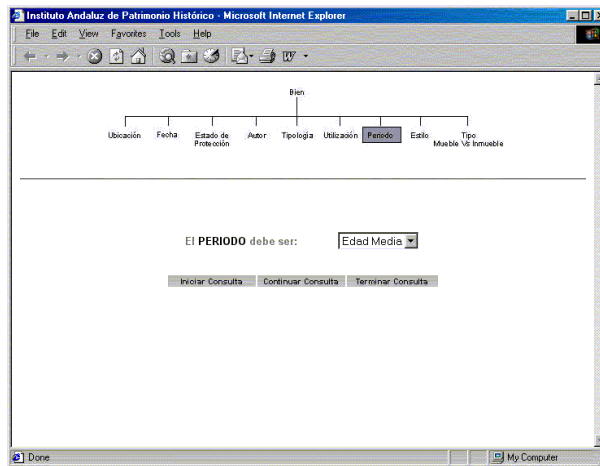


Fig. 7. Frase en *Lenguaje Natural Acotado* para el concepto *Periodo*

Para expresar restricciones sobre conceptos asociados a atributos de tipo *Fecha*, se propone el esqueleto de la Fig. 8. El concepto *Fecha* de la ontología tiene asociado el identificador de este esqueleto de frase. La frase en *Lenguaje Natural Acotado* que generaría el sistema para el concepto *Fecha*, es la que aparece en la Fig. 9.

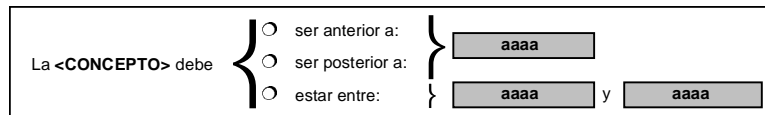


Fig. 8. Esqueleto de frase para el tipo de dato *Fecha*

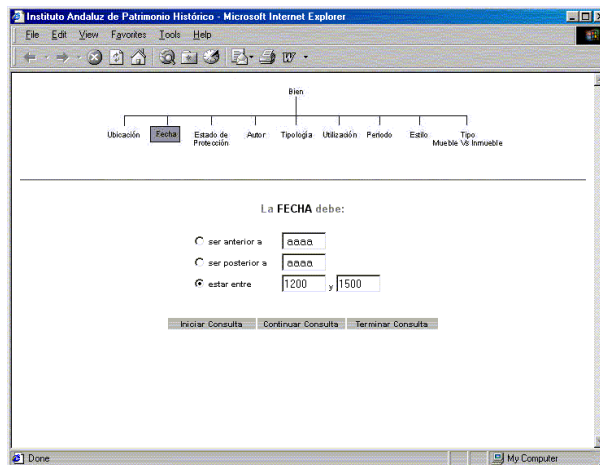


Fig. 9. Frase en *Lenguaje Natural Acotado* para el concepto *Fecha*

El conjunto de esqueletos de frase que hemos presentado no es estático, es decir, en cualquier momento es posible variar la sintaxis de un esqueleto para adecuarlo, aún más, al concepto al que esté asociado. Asimismo, es posible generar tantos esqueletos

de frase como sea necesario para contemplar todos los tipos de datos diferentes que se den entre los conceptos de las ontologías. Por ejemplo, si hubiese que incluir un concepto de tipo *Número* en la ontología, el esqueleto de frase que habría que asociar a dicho concepto podría ser el de la Fig. 10.

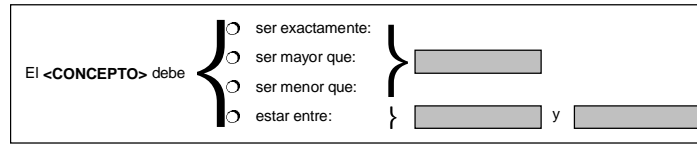


Fig. 10. Esqueleto de frase para el tipo de dato *Número*

Podemos asegurar, por tanto, que sea cual sea el tipo de dato asociado al concepto siempre podremos construir un esqueleto de frase adecuado para que el usuario pueda expresar restricciones sobre el mismo.

6.2 Esqueletos de navegación

Además de este tipo de esqueletos de la sección anterior, en nuestro sistema tenemos esqueletos de frase que usa el *Generador de la Interfaz de Usuario* para moverse a través de los conceptos de la ontología y permitir así al usuario construir la consulta completa.

Existen dos tipos de esqueletos de navegación: *Esqueletos de Especialización* y *Esqueletos de Descripción*.

6.2.1 Esqueletos de Especialización

Este tipo de esqueleto, que se presenta en la Fig. 11, se utiliza en conceptos que tengan una relación de tipo *is a* y permite que el concepto bajo proceso pase a ser una de las especializaciones.

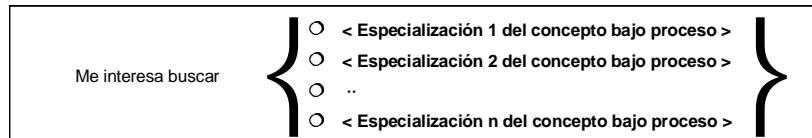


Fig. 11. Esqueleto de Especialización

Por ejemplo, si el concepto actual fuese *Tipo Mueble Vs Inmueble* de la ontología de la Fig. 4, el esqueleto de la Fig. 11 se convertiría en la frase en *Lenguaje Natural Acotado* que aparece en la Fig. 12.

El usuario seleccionará una de las dos opciones que le presenta la interfaz dependiendo del tipo de bien que esté buscando: *Bienes Inmuebles* o *Bienes Muebles*, respectivamente, y el concepto actual pasará a ser el seleccionado.

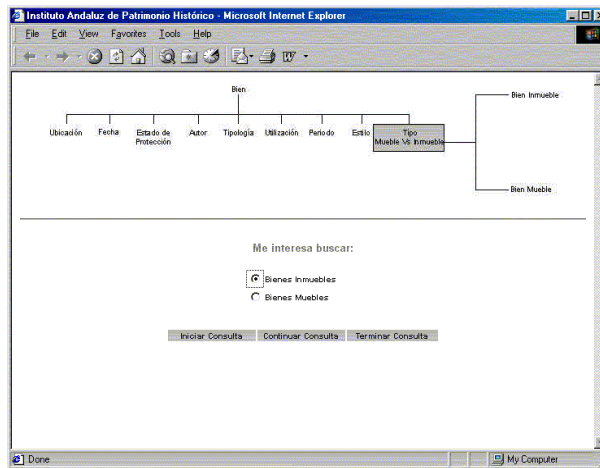


Fig. 12. Frase en *Lenguaje Natural Acotado* para el concepto *Tipo Mueble Vs Inmueble*

6.2.2 Esqueletos de Descripción

Este tipo de esqueleto de navegación se usa para preguntar al usuario en qué conceptos, de entre aquellos que describen al concepto que se está procesando, está interesado para expresar restricciones. Este esqueleto permite que el usuario especifique cuáles son las características de un concepto sobre las que quiere expresar restricciones.

Este esqueleto se presenta en la Fig. 13. Hay que destacar que aparecen en primer lugar los conceptos hoja y en segundo lugar aparece la lista de los conceptos no hoja.

Con respecto a <CONCEPTO>, me interesa especificar restricciones sobre su

<input type="checkbox"/>	< Concepto hoja 1 en la relación <i>has</i> >	}
<input type="checkbox"/>	..	
<input type="checkbox"/>	< Concepto hoja <i>n</i> en la relación <i>has</i> >	
<input type="checkbox"/>	< Concepto no-hoja 1 en la relación <i>has</i> >	
<input type="checkbox"/>	..	
<input type="checkbox"/>	< Concepto no-hoja <i>n</i> en la relación <i>has</i> >	}

Fig. 13. Esqueleto de Descripción

Por ejemplo, para el concepto *Bien* de la ontología de la Fig. 4, el *Esqueleto de Descripción* de la Fig. 13 se convertiría, una vez que el *Generador de la Interfaz de Usuario* sustituyese la etiqueta <CONCEPTO> por "*Bien*", en la frase en *Lenguaje Natural Acotado* que aparece en la Fig. 14. El usuario deberá seleccionar, de entre los conceptos que describen al concepto *Bien*, aquellos sobre los que le interesa expresar restricciones, para caracterizar así los *Bienes* que está buscando.

En la Fig. 14 se observa que el usuario ha seleccionado {Fecha, Periodo y Ubicación} como aspectos relevantes que le permitirán caracterizar los bienes que está buscando.

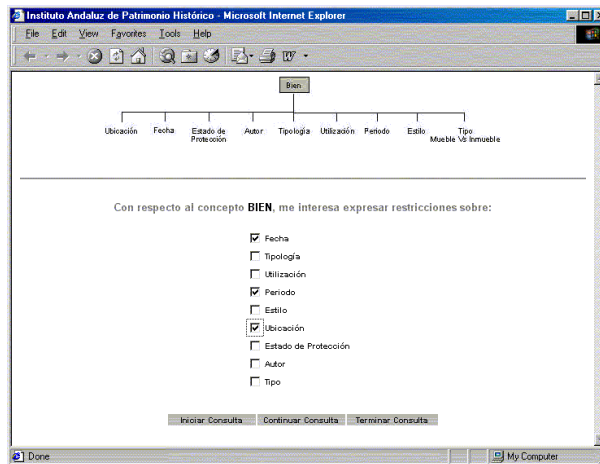


Fig. 14. Frase en *Lenguaje Natural Acotado* para el concepto *Bien*

7 Proceso de consulta

En el proceso de consulta a la base de datos de patrimonio histórico intervienen, como se ha visto en la sección 4, dos módulos principales: El *Generador de la Interfaz de Usuario* y el *Generador de Consultas*. A continuación se explica el funcionamiento de ambos módulos.

7.1 Generador de la Interfaz de Usuario

Como se ha adelantado durante la descripción de la arquitectura del sistema, el *Generador de la Interfaz de Usuario* construye la *Interfaz de Usuario* usando la ontología y los esqueletos de frase.

El proceso de construcción de la interfaz de usuario consiste en ir presentando al usuario las frases en *Lenguaje Natural Acotado* en el orden adecuado, para que el usuario exprese su consulta completa. Para llevar a cabo este proceso, el *Generador de la Interfaz de Usuario* usa un algoritmo que denominamos *Algoritmo de la Interfaz de Usuario* (AIU) y que vamos a explicar a continuación.

El algoritmo comienza preguntando al usuario si está interesado en consultar un bien concreto (del que ya conoce su denominación general o su código DGBC, asignado por la Dirección General de Bienes Culturales) o si busca un conjunto de bienes con unas características comunes. Según la respuesta del usuario, el sistema utilizará la ontología de la Fig. 3 o la ontología presentada en la Fig. 4, respectivamente.

El primer concepto que procesa el AIU es el concepto raíz de la ontología que haya seleccionado.

El AIU comprueba, en primer lugar, si el concepto bajo proceso tiene una relación de tipo *is a* (relación de *Generalización/Especialización*). En ese caso, el AIU instancia el *Esqueleto de Especialización* usando el concepto bajo proceso y los

conceptos de su especialización, recoge la opción que el usuario elige y vuelve a comenzar su ejecución usando como parámetro de entrada el concepto de la especialización elegido por el usuario.

Si el concepto bajo proceso no tiene especialización, el algoritmo comprueba si tiene una relación de tipo *has* (relación de *Descripción*), en cuyo caso instancia el *Esqueleto de Descripción* para dicho concepto. Una vez el usuario selecciona aquellos conceptos sobre los que quiere expresar restricciones, el algoritmo crea una lista, que denominamos *ConceptosHojaSeleccionados*, con los conceptos hoja seleccionados por el usuario y otra lista, que denominamos *ConceptosNoHojaSeleccionados*, que contiene los conceptos no hoja seleccionados por el usuario.

Por ejemplo, si el concepto bajo proceso es *Bien* de la ontología de la Fig. 4, la frase en *Lenguaje Natural Acotado* que genera el algoritmo es la presentada en la Fig. 14. Como se puede ver en la Fig. 14 el usuario ha seleccionado tres conceptos. Estos tres conceptos, *Fecha*, *Periodo* y *Ubicación*, son las características del concepto *Bien* que el usuario quiere restringir para describir los bienes que está buscando. En este caso, la lista *ConceptosHojaSeleccionados* contendría {Fecha, Periodo} y la lista *ConceptosNoHojaSeleccionados* contendría únicamente {Ubicación}.

El *Algoritmo de la Interfaz de Usuario (AIU)* usa estas dos listas para continuar su ejecución. Para cada concepto que aparezca en la lista *ConceptosHojaSeleccionados*, el algoritmo instancia el esqueleto de frase que tenga asociado dicho concepto en la ontología y le presenta al usuario la frase en *Lenguaje Natural Acotado* resultado de la instanciación de dicho esqueleto.

Una vez procesados todos los conceptos de la lista *ConceptosHojaSeleccionados*, para cada concepto de la lista *ConceptosNoHojaSeleccionados*, el AIU vuelve a comenzar su ejecución usando como parámetro de entrada dicho concepto.

Para facilitar al usuario la comprensión de los conceptos sobre los que puede establecer restricciones, la parte superior de la pantalla le presenta la parte del árbol de la ontología por la que ya se ha pasado en el proceso de consulta. Además, las restricciones ya impuestas por el usuario mediante frases en *Lenguaje Natural Acotado* irán apareciendo en la pantalla escritas ya en lenguaje natural.

Las restricciones que expresa el usuario sobre cada concepto se traducen a una sentencia FQL. Cada sentencia FQL formará parte del documento global en FQL que representará la consulta completa del usuario.

Es necesario destacar, a la vista del *Algoritmo de la Interfaz de Usuario (AIU)*, que el *Generador de la Interfaz de Usuario* es totalmente independiente de la base de datos. Es decir, no es necesario reescribir el código de este módulo cuando se modifica el esquema de la base de datos (es decir, cuando se añaden o eliminan atributos, tablas, etc.). Sólo será necesario modificar la ontología y, quizás, añadir algún esqueleto de frase.

7.2 Generador de Consultas

Obtenido el documento en FQL que describe la consulta formulada por el usuario en la *Interfaz de Usuario*, el proceso de consulta a la base de datos pasa a ser responsabilidad de otro de los módulos del *Gestor de Consultas*: el *Generador de Consultas*.

La función del *Generador de Consultas* es traducir la consulta en FQL a SQL, lenguaje que usa la base de datos de patrimonio histórico.

Para llevar a cabo esta tarea, el *Generador de Consultas* usa la información que tiene asociado cada concepto en la ontología. Es decir, lee de la ontología la expresión NombreTabla.NombreAtributo asociada a cada concepto, para poder construir la parte de la sentencia SQL correspondiente.

Es necesario destacar que, igual que sucede con el módulo *Generador de la Interfaz de Usuario*, el *Generador de Consultas* es totalmente independiente de la base de datos ya que, otra vez, es la ontología la que hace transparentes para el *Generador de Consultas* todos los cambios que se produzcan en la base de datos. Es decir, no es necesario reescribir el código de este módulo cuando se modifica el esquema de la base de datos ni tampoco cuando cambia el sistema gestor de la base de datos. Únicamente será necesario cambiar la expresión asociada a cada concepto en la ontología.

8 Conclusiones

La arquitectura general en la que está basada el sistema que aquí se presenta está ya diseñada.

De su aplicación al sistema de consulta vía Web que se está desarrollando para el Instituto Andaluz de Patrimonio Histórico esperamos obtener interesantes resultados.

Referencias

1. Bases de Datos y Emblemática Hispánica bajo Internet. <http://rosalia.dc.fi.udc.es/cicyt>
2. Brisaboa, N.R., Durán, M.J., Lalín, C., Iglesias, E.L, López, J.R., Paramá, J., Places, A.S. y Penabad, M.. Integrating the access to documental databases on the web. *The Fifth World Conference on Integrated Design and Process Technology IDPT 2000*. Proceedings of The Fifth World Conference on Integrated Design and Process Technology IDPT 2000. Dallas, Texas. June 2000.
3. Brisaboa, N.R.; Durán, M. J.; Lalín, C.; López, J. R.; Paramá, J. R.; Penabad, M.; Places, A. S. Propuesta de Arquitectura para un Sistema de Bases de Datos Documentales. *IV Jornadas de Ingeniería del Software y Bases de Datos JISBD'99*. Actas de las *IV Jornadas de Ingeniería del Software y Bases de Datos JISBD'99*. Cáceres, Noviembre 1999.
4. Brisaboa, N.R., Places, A.S. y Rodríguez, F. J. "Arquitectura para Federación de Bases de Datos Documentales basada en Ontologías". En actas del *4º Workshop Iberoamericano de Ingeniería de Requisitos y Ambientes Software (IDEAS 2001)*. Págs. 252-262. San José, Costa Rica, Abril 2001.
5. Brisaboa, N. R., Penabad, M. R., Places, A. S., Rodríguez, F. J. Using ontologies for federation of Web accessible databases. In proceedings of the *13th international conference on software engineering & knowledge engineering (SEKE 2001)*. Págs. 87-94. Buenos Aires, Argentina, Junio 2001.
6. Brisaboa, N. R., Penabad, M. R., Places, A. S., Rodríguez, F. J. A Documental Database Query Language. *String Processing and Information Retrieval (SPIRE 2001)*. Aceptado y pendiente de publicación.
7. Gruber, T. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *IJHCS*, 43 (5/6): 907-928. 1994.
8. Gruber, T. <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>
9. Guarino, N. (ed.), Formal Ontology in Information Systems. *Proceedings of FOIS'98*. Amsterdam, IOS Press, pp. 3-15. , Trento, Italy, 6-8 June 1998.