

Resumen

El esquema de catalogación de una colección digital sirve para catalogar (describir y clasificar) adecuadamente los objetos que la conforman. Dicha catalogación es esencial para habilitar la explotación efectiva de la colección, tanto por parte del sistema de gestión de la misma (repositorio digital), como por parte de las herramientas externas encargadas de recuperar, reproducir y agregar los objetos. Efectivamente, funcionalidades básicas como la recuperación de recursos a partir de consultas, o la navegación guiada, dependen todas ellas de la adecuada catalogación de los recursos.

De esta forma, en el dominio de las bibliotecas digitales se han propuesto múltiples estándares para la catalogación de objetos digitales, cuyo propósito último es garantizar la interoperabilidad entre los distintos repositorios y aplicaciones que manipulan los objetos. No obstante, la adopción de esquemas de catalogación preestablecidos puede no ser una solución satisfactoria en aquellos escenarios donde los esquemas de catalogación, en lugar de establecerse a priori, son artefactos que cambian y evolucionan a lo largo de todo el ciclo de vida de la colección, con el fin de adaptarse a requisitos y necesidades de catalogación cambiantes, que no se conocen al inicio de la producción de la colección, sino que aparecen y maduran al mismo tiempo que la colección crece y evoluciona. Este ha sido el caso en nuestras experiencias colaborando con distintos grupos de Humanidades en distintos dominios (arqueología, literatura convencional, literatura digital, escritura creativa, etc.), experiencias que han mostrado que la definición inductiva de esquemas de catalogación (es decir, la evolución concurrente de esquemas y colecciones), lejos de ser una excepción es la realidad habitual en la creación de colecciones muy específicas, frecuentemente orientadas a la investigación y docencia.

Con el fin de permitir que los esquemas de catalogación evolucionen y se adapten a las necesidades cambiantes que surgen durante el ciclo de vida de una colección, es necesario proporcionar mecanismos que permitan reconfigurar dichos esquemas conforme la colección cambia. Esta tesis aborda este problema de la gestión de colecciones de objetos digitales con esquemas de catalogación que pueden reconfigurarse dinámicamente.

De esta forma, el primer aspecto abordado en esta tesis tiene que ver con la dimensión lógica del problema de la reconfiguración de esquemas: ¿cómo se puede expresar de manera efectiva dicha reconfiguración? Para ello, la tesis propone enfocar el problema desde el nivel estructural o sintáctico de los esquemas de catalogación, entendiendo la reconfiguración como la aplicación de un conjunto básico de operaciones de edición sobre la estructura de los esquemas (de forma similar a como, por ejemplo, el procesamiento de un lenguaje informático se centra en la sintaxis abstracta del lenguaje). Más concretamente, la tesis propone centrarse en representaciones arborescentes de los esquemas, y en utilizar un conjunto básico de operaciones de edición sobre dichas representaciones (añadido, renombrado y eliminación de nodos, cambio de filiación de nodos en la jerarquía, y fusión de nodos para resolver problemas de sinonimia). La reconfiguración en sí será llevada a cabo por los expertos en el dominio de la colección, utilizando editores adecuados, expertos que cuidarán de que las reconfiguraciones tengan sentido a niveles semánticos y pragmáticos (de forma similar a como ocurre con el marcado descriptivo de un documento, utilizando, por ejemplo, un lenguaje de marcado descriptivo definido mediante SGML o XML). La tesis muestra la factibilidad de esta propuesta mediante la aplicación del enfoque a distintos tipos de esquemas de

catalogación, tanto esquemas basados en metadatos como basados en vocabularios controlados (listas de términos, taxonomías y tesauros facetados).

El segundo aspecto abordado por la tesis tiene que ver con la dimensión humana del proceso de reconfiguración: ¿cómo guiar al experto que realiza la reconfiguración del esquema en esta actividad? Para ello, la tesis propone aplicar un enfoque basado en análisis de datos, según el cual la colección catalogada se analiza para inducir una estructura comparable con el esquema original. El experto, entonces, puede comparar la estructura inferida a partir de la colección con el esquema inicialmente propuesto, y utilizar las diferencias encontradas para, bien diagnosticar usos potencialmente erróneos del esquema durante el proceso de catalogación, bien para reconfigurar el esquema con el fin de adaptar el mismo a la realidad de su uso. Este enfoque se ha aplicado al caso particular de la catalogación de anotaciones sobre textos literarios clasificadas mediante taxonomías. La técnica de análisis de datos utilizada ha sido el análisis de conceptos formales, lo que ha permitido inducir organizaciones reticulares de las anotaciones. Dichos retículos pueden compararse de manera significativa con las taxonomías originales, ofreciendo a los expertos, de esta forma, una guía valiosa de cara a la reconfiguración.

Por último, el tercer aspecto abordado en esta tesis se refiere al factor tecnológico del proceso de reconfiguración. Efectivamente, para que el sistema de gestión de colecciones pueda llevar a cabo funcionalidades básicas como la navegación o la búsqueda a las que se ha hecho referencia anteriormente, es necesario disponer de índices apropiados de los objetos. Dichos índices se construyen a partir de la catalogación de los mismos. Por tanto, si el esquema de catalogación se reconfigura, dichos índices pueden invalidarse parcial, o incluso totalmente. Para evitar este efecto es necesario proponer índices que permanezcan invariantes a la reconfiguración de los esquemas. No obstante, esto puede acarrear, a su vez, una merma considerable en el rendimiento, frente a índices dependientes de un esquema específico. En esta tesis se proponen propuestas de indexado que permiten minimizar el impacto del rendimiento producido por las reconfiguraciones en el caso de una funcionalidad crítica del sistema de gestión de colecciones: la navegación guiada por los esquemas de catalogación. Para ello, la tesis comienza modelizando dicha navegación mediante un autómata finito determinista, el autómata de navegación, y muestra cómo el enfoque puede aplicarse con los distintos tipos de esquemas de catalogación contemplados (tanto esquemas de metadatos, como los basados en vocabularios controlados). Mientras que dicho autómata puede proporcionar directamente el índice requerido, se muestra cómo, en el peor de los casos, su tamaño puede crecer exponencialmente con el tamaño de la colección. Por tanto, se proponen alternativas de indexado que permiten recrear dinámicamente las partes relevantes de dicho autómata durante la navegación: una alternativa clásica, basada en índices invertidos, y una alternativa más sofisticada, basada en dendrogramas. La tesis analiza el comportamiento en la práctica de ambas alternativas, y muestra, por último, cómo la alternativa basada en dendrogramas puede mejorar sustancialmente la basada en índices invertidos.