

# Estructura, semántica, extracción de información y XML legislativo: experiencias en la Universidad de Valladolid

**M.<sup>a</sup> Mercedes Martínez González**  
**Dámaso-Javier Vicente Blanco**  
**Pablo de la Fuente Redondo**  
**Joaquín Adiego Rodríguez**  
**Alma María Pisabarro Marrón**  
**José Manuel Sánchez Felipe**  
Universidad de Valladolid (España)

## Resumen

La extracción de la estructura implícita en el contenido de los textos legislativos y su posterior representación de forma explícita a través del marcado XML es uno de los pilares del trabajo desarrollado en la Universidad de Valladolid por el grupo de investigación que se ocupa de la manipulación electrónica de textos legislativos. Esta solución permite abordar una serie de tratamientos adicionales de extracción de información, manipulación, navegación, y búsquedas semánticas en los que la estructura juega un papel relevante. En este artículo se ofrece una perspectiva general del trabajo realizado por el grupo hasta el momento.

**Palabras clave:** XML legislativo. Manipulación de estructura. Extracción de información. Marcado automático.

## Abstract

The extraction of the structure implicit in legislative documents content, followed by its elicitation and representing as XML markup, is one of the bases of the work developed in the University of Valladolid by the team of researchers interested in the electronic management of legislative texts. An important advantage is that a set of additional processes of information extraction, management, navigation, and semantic searching, in which structure plays a significant role, can be tackled once this information is made available to automatic processes in the form of XML data. This paper offers an overview of the work made by the team until now.

**Keywords:** Legislative XML. Structure management. Information extraction. Automatic markup.

## 1. Introducción

Los sistemas de información legislativa son una fuente de información habitual para los profesionales del derecho. El grupo de Valladolid lleva unos años investigando sobre la utilización de XML para representar y manipular este tipo de documentos y su información. Nuestra propuesta está relacionada con la manipulación de las versiones en español de los textos de la Unión Europea relacionados con el conflicto de jurisdicciones en materia civil y mercantil. Este trabajo se enmarca dentro de un proyecto de investigación en colaboración con el Ilustre Colegio de Procuradores de Valladolid (1).

Nuestra propuesta incluye la utilización de técnicas de extracción automática de información del contenido de los textos. Se extrae su estructura y otras propiedades (metainformación, relaciones) que permitirán realizar búsquedas semánticas sobre las colecciones de textos legislativos. Para facilitar estos procesos, se aborda la organización y articulación de los textos legislativos que se pretende utilizar.

La propuesta es innovadora en el uso de técnicas de extracción de información aplicadas a los textos legislativos de la Unión Europea en su versión en español. Actualmente no nos consta ningún otro trabajo similar sobre estas mismas colecciones. Los trabajos realizados en el ámbito europeo han tenido una finalidad distinta, como sucede con la Red Judicial Europea ([http://ec.europa.eu/civiljustice/index\\_es.htm](http://ec.europa.eu/civiljustice/index_es.htm)). El resultado de los procesos de extracción de información es un conjunto de colecciones de documentos XML que sirven para representar el contenido y la estructura de los textos legislativos, y los metadatos que describen estos textos. La utilización final por parte de los usuarios se posibilita a través de un conjunto de aplicaciones que proporcionan interfaces de usuario amigables y personalizadas en función del perfil de los usuarios.

Nuestro enfoque se fundamenta en considerar la manipulación de la estructura, las referencias, y los metadatos como procesos diferentes, pero no aislados, en los que se explotan conjuntamente los textos y sus relaciones.

Las propuestas para la extracción de información se han probado con diferentes colecciones: una de textos legislativos españoles y una de textos legislativos de la Unión Europea sobre conflicto de jurisdicciones (véase Vicente, 2007). En cuanto a las aplicaciones de usuario, se han utilizado preferentemente con los textos comunitarios, pues son estos los que más interesan a los juristas que colaboran en esta investigación.

## 2. Fundamentos de la propuesta

Las soluciones que proponemos intentan sacar el máximo partido posible del conocimiento del entorno jurídico que tienen los propios especialistas (modo de or-

ganizar la información, formas preferidas de acceder a ella, de buscar, conceptos y relaciones relevantes, etcétera). Cuando iniciamos nuestra investigación tomamos dos decisiones fundamentales: por un lado, utilizar siempre código abierto, para facilitar el desarrollo de *software* adaptado a nuestras necesidades específicas con poco coste económico; por otro, realizar desarrollos con esfuerzos de implementación razonables, lo cual se consigue con desarrollos basados en estándares XML como XPath, y *software* XML, como *parsers* XML, procesadores XSLT, etcétera.

Desde la perspectiva investigadora nuestra propuesta se fundamenta en tres ideas:

En primer lugar, los textos jurídicos son documentos que tienen una estructura bien definida, implícita en su contenido. Esta estructura es la que reconoce cualquier persona durante una lectura ordenada desde el inicio del documento. Habitualmente siguen unas reglas que determinan cuáles son las divisiones permitidas en estos textos, la jerarquía de inclusión entre ellas, etcétera, aunque en ocasiones puede haber excepciones. Existen manuales de técnica jurídica que recogen estas reglas, de los cuales es un ejemplo el del Grupo de Estudios de Técnica Legislativa (1989). Estas reglas son las que utilizaremos con el fin de elaborar las propuestas que sustentan los procesos automáticos para reconocer estructura y marcar automáticamente textos, abordados en la sección 3.

En segundo lugar, las relaciones entre los textos legislativos son muy frecuentes. Puede tratarse de relaciones semánticas (documentos de la misma categoría, que tratan sobre aspectos similares...) o de otras, como las referencias cruzadas. En cualquier caso, es habitual que las relaciones presentes en los documentos no se limiten a las de los propios documentos entre sí, sino que sean más precisas, implicando elementos concretos dentro de un documento legislativo (véase el ejemplo de las referencias que se muestran en la tabla I). La estructura de los documentos sirve para precisar qué parte de un documento participa en una relación. Para un usuario, es un modo de delimitar el contenido que debe analizar y, por tanto, ser más eficaz en la localización de la información que le interesa. Esto es, introducir la manipulación de estructura en los procesos de extracción de información. Nuestro objetivo es ofrecer soluciones automáticas que permitan recoger ese nivel de granularidad de precisión en la información disponible. Es decir, que la detección de relaciones y su información sea automática y precisa.

En tercer lugar, es posible encontrar en el contenido de los textos metainformación que los describe. Por ejemplo, los textos normativos suelen incluir información sobre su ámbito de aplicación espacial y temporal. En algunos casos es posible caracterizar patrones, de texto y de localización (en la estructura), que pueden servir para abordar su reconocimiento automático y la consiguiente generación automática de la metainformación asociada.

Teniendo en cuenta estas premisas, nuestro modelo es un grafo de relaciones cuyos nodos son los documentos y sus divisiones (elementos de estructura). Los arcos son las relaciones entre ellos (de estructura, referencias, modificaciones...). Sobre este modelo se apoyan las soluciones que ofrecemos para procesos relacionados con la estructura: *a*) extracción de estructura y consiguiente generación de un documento XML cuyas marcas reflejen la estructura reconocida; *b*) extracción de referencias, basada en la estructura; *c*) recuperación de información estructurada; *d*) obtención de estructuras virtuales (consolidación, enlaces de hipertexto); *e*) anotación de elementos de estructura.

### 2.1. Desarrollo sobre XML

Actualmente la utilización de XML para representar textos jurídicos está generalizada. Existen varios esquemas creados específicamente para este propósito (Andrés y Gómez Skarmeta, 2007; Arellano, Nogales y Martín, 2003; Arellano, 2005; Arnold-Moore y Clemen, 2000; Boer et ál., 2002; Bolioli et ál., 2002; Finke, 1997; Gómez Skarmeta, 2000; Marchetti et ál., 2002; Martínez, Fuente y Derniame, 2003; Palmirani y Brighi, 2002; Magnusson-Sjöber, 1997a y 1997b; Vitali, 2007) y se siguen creando nuevos esquemas. Por otro lado, la manipulación de estructura es una propiedad inherente a XML, lo cual facilita la implementación de procesos en los que se explota la estructura de los documentos.

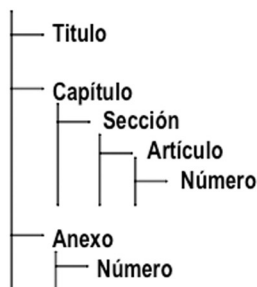
En nuestro caso utilizamos XML para almacenar los documentos y sus metadatos. Los esquemas correspondientes están representados como DTDs o XML Schemas, lo cual permite validar automáticamente su composición. Utilizamos también XSLT, el lenguaje de transformación asociado a XML y XPath para direccionar elementos de estructura.

### 3. Extracción de estructura y marcado automático

Conocer la estructura de un documento aporta ventajas, como la posibilidad de utilizarla para acceder a elementos concretos o en las búsquedas de información (recuperación de información estructurada). La estructura que nos planteamos reconocer y representar de modo automático es la implícita en el contenido, esto es, la misma que reconoce un usuario lector. Los textos legislativos son un buen ejemplo de documentos muy estructurados cuyas divisiones se denominan *elementos del articulado*. La figura 1 muestra un fragmento de uno de estos documentos, extraído del Reglamento (CE) n.º 44/2001 de 22 de diciembre de 2000 “on jurisdiction and the recognition and enforcement of judgments in civil and commercial matters”. En la parte izquierda de la figura se puede ver la estructura del documento, y el contenido del mismo en la derecha.

Nuestro enfoque consiste en imitar el método que permite a un lector construir una imagen mental de la estructura del documento mientras lo lee desde el inicio hasta el final. Un elemento empieza donde termina otro, y el lector lo reconoce por la pre-

LEY / DECRETO / ...



**Council Regulation (EC) No 44/2001 of 22 December 2000 on jurisdiction and the recognition and enforcement of judgments in civil and commercial matters**

...

**Chapter II**  
**JURISDICTION**  
**Section 1**  
 General provisions  
**Article 2.**  
 1. Subject to this Regulation, persons domiciled in a Member State shall, whatever their nationality, be sued in the courts of that Member State.  
 2. Persons who are not nationals of the Member State ...

**Article 3**  
 1. Persons domiciled in a Member State may be sued in the courts of another Member State only by virtue of the rules set out in Sections 2 to 7 of this chapter.  
 2. In particular the rules of national jurisdiction set out in Annex I shall not be applicable as against them.

...

Figura 1. Fragmento de un texto legislativo, y su estructura.

sencia de algunas frases o expresiones características (es preciso fijarse en que esto es independiente del modo o formato con que se haya representado el documento).

La extracción de estructura es, por tanto, un análisis del contenido. Opera sobre documentos electrónicos que, en caso de tener algún tipo de marcado, no se corresponde con la estructura implícita en su contenido (documentos cuyas marcas delimitan elementos XML que no coinciden con los elementos del articulado). Al igual que hace un lector, el proceso automático de análisis reconoce el inicio de un nuevo elemento del articulado (división) por la aparición de determinados patrones de texto. El proceso no se limita al reconocimiento, sino que también consiste en un marcado automático (normalización automática) de los documentos, proceso que permite evolucionar de una copia electrónica cuya estructura lógica no se corresponde con la estructura interna del texto legislativo a una nueva copia, también XML, cuya estructura lógica representa la estructura interna del texto legislativo. Las nuevas marcas, que delimitan el inicio y el fin de elementos XML cuyo contenido coincide con los elementos del articulado, se generan automáticamente a medida que se van reconociendo divisiones durante el análisis. En nuestro trabajo de 2003 (Martínez, Fuente y Derniame, 2003) se pueden encontrar más detalles sobre este proceso. Las reglas que rigen la estructura de estos documentos están representadas en un esquema al que haremos referencia de nuevo en la sección 4: el *esquema pivote*. Las marcas que se generan son las que se corresponden con los tipos de elementos recogidos en este esquema.

Este proceso se probó en varios experimentos, organizados en etapas. Las colecciones de prueba contienen textos normativos españoles y europeos. Las etapas, ordenadas según su secuencia temporal, son:

*Primera etapa.* Se somete a los documentos a un preproceso, previo al proceso de análisis y normalización (marcado) propiamente dicho. En él se incluyen varias tareas: todos los documentos se pasan a texto plano, sea cual sea su formato (Word u otros); los textos legislativos que estaban fragmentados en varios documentos se unifican en un uno solo; se eliminan elementos de información que pueden perturbar el análisis automático (notas, definiciones, índices...); se genera un marcado elemental, que únicamente contiene la división en párrafos del documento (<p>,</p>) u otros elementos de presentación de bajo nivel, no recogidos en los manuales de redacción legislativa. Esta fase (que requiere un importante esfuerzo manual) produjo 1665 documentos XML bien formados.

*Segunda etapa.* La extracción de estructura y normalización recibe como entrada los documentos resultado de la primera etapa. Contiene el analizador que reconoce patrones de texto y la generación automática de las nuevas marcas XML. En esta etapa se generan las marcas asociadas a elementos estándar que sí están recogidos en los manuales de redacción legislativa y, por tanto, en el esquema que describe las reglas de estructura. Funcionó con 1583 documentos. Dado que está implementado sobre un *parser* (analizador) XML, como indicamos en la sección 1, este componente *normalizador* (creador del marcado XML) no pudo procesar algunos documentos que tenían errores que el *parser* XML no acepta (&icute;, &ordm;, etcétera), tablas que no estaban bien formadas o documentos que no tenían un elemento raíz.

#### 4. Extracción de información estructurada

A la hora de extraer información del contenido de los documentos aprovechamos la estructura para delimitar las zonas del documento donde pueden encontrarse los patrones de texto que buscamos. Asimismo, una vez localizados, utilizamos también la estructura para indicar en qué puntos del documento se han encontrado los ítems de información recogidos. Los dos tipos de información que extraemos son referencias y metadatos sobre la aplicabilidad de los documentos. En adelante nos centramos en las referencias.

Las referencias cruzadas entre documentos normativos son muy frecuentes. Sirven para incluir el contenido de otros textos (Grupo de Estudios de Técnica Legislativa, 1989), para redirigir hacia documentos donde se puede ampliar la información sobre el tema o para modificar otros documentos (algunos textos normativos son una colección de modificaciones a otros previos, de modo que cada modificación implica una referencia que precisa cuáles son los fragmentos afectados por dicha modificación).

En muchos casos las referencias lo son a porciones de los documentos, esto es, a determinados elementos del articulado dentro de un texto legislativo. En estos casos la posición de los fragmentos respecto a la estructura interna de los docu-

mentos sirve para localizarlos e identificarlos. Esto es aplicable tanto a las referencias internas (en el mismo documento) como a las externas (a otro documento distinto). Interesa, pues, reconocer e interpretar de manera precisa las referencias encontradas, para transferir al sistema que manipule esta información posteriormente una precisión (granularidad) en su tratamiento equivalente a la que se utilizó en su redacción.

El proceso de extracción de las referencias se basa en un análisis del contenido de los textos. Combina tres subprocesos: la *detección* de patrones, la *resolución* de las referencias y su *almacenamiento* en una base de datos.

Estas tareas son responsabilidad de las herramientas que se muestran en la figura 2. El *analizador* recibe el documento en el que se buscan las referencias. Colabora con un *reconocedor de patrones*, que detecta patrones en los fragmentos de texto que recibe del analizador; este componente maneja la información sobre el vocabulario utilizado para referenciar elementos del articulado y sobre las gramáticas que describen los patrones asociados a cada tipo de referencia. La organización y definición de estas gramáticas está asociada a una taxonomía, que presentamos en el apartado 4.1. El resultado de este análisis es un conjunto de referencias; para cada una de ellas se obtiene la cadena encontrada y los datos que la describen (documento en el que se encontró, fragmento donde se halló, documento referenciado, fragmento o fragmentos referenciados).

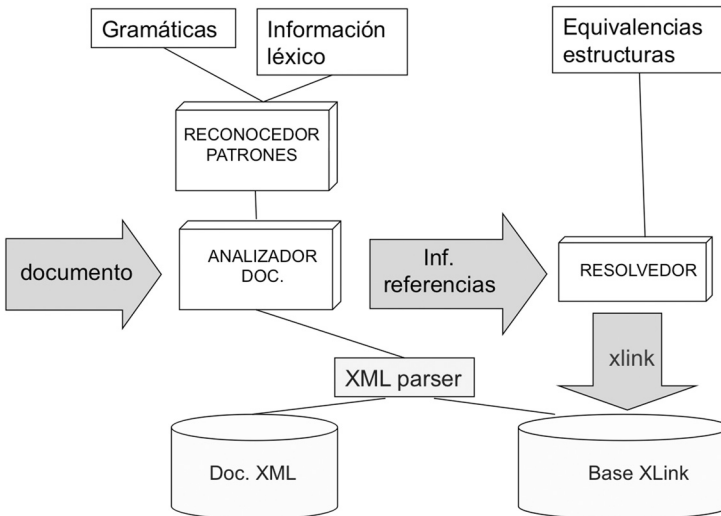


Figura 2. Software que se encarga de la extracción de referencias.

Estos son los datos que procesa el componente *resolvedor*, que se encarga de asociar cada referencia encontrada con la dirección de un ítem legislativo y su correspondiente representación digital. Si no es capaz de hacerlo, el *resolvedor* genera un mensaje de error, que se asociará a la referencia, de modo que los restantes servicios de la biblioteca que posteriormente puedan utilizar esta información sepan que no ha sido posible resolverla y, por tanto, que no podrán acceder en la biblioteca a un fragmento de texto que procesar. En caso de que sí se pueda resolver, se genera un enlace con toda su información, que se almacena en una base de enlaces.

Un problema importante es la heterogeneidad en los esquemas de los documentos analizados y/o referenciados. Si el marcado difiere (por ejemplo, si provienen de distintas fuentes), sería preciso tener en cuenta tantos tipos de estructura como esquemas haya en las colecciones, lo cual afectaría a los procesos de extracción y resolución. Nuestra solución es distinguir dos tipos de manipulación: *a*) de la estructura, de la cual se encarga el analizador, que en última instancia se apoya en un *parser* XML y, por tanto, tiene incorporado el tratamiento de la estructura del documento XML analizado; *b*) del contenido, que corre a cargo del reconocedor de patrones; este componente desconoce por completo la estructura del documento analizado, ya que siempre recibe fragmentos de texto; es aquí donde ocurre la extracción de información, en su concepción tradicional.

En los prototipos que hemos desarrollado hemos optado por abordar la manipulación de estructuras heterogéneas como un problema de integración de esquemas, tal como se entiende en el mundo de las bases de datos. Todos los documentos se transforman a un *esquema pivote*, seleccionado por su correspondencia con las reglas generales que definen la estructura interna de los textos legislativos. Esto simplifica enormemente los procesos de resolución de las referencias: el vocabulario de los esquemas y el encontrado en las referencias es similar, y los elementos de la estructura interna de un texto legislativo se corresponden unívocamente con elementos del documento XML (lo que de otro modo no estaría garantizado y podría complicar enormemente las posibilidades de abordar con éxito una resolución automática). Este esquema pivote es el mismo al que se ajustan los documentos resultado de la extracción de estructura y marcación automática presentadas en la sección 2.

#### 4.1. Detección de referencias

La detección de patrones se basa en un conjunto de gramáticas, asociadas a una taxonomía, que clasifica las referencias en función de cuatro criterios: tipo, rango, número y uso del contexto.

*Tipo.* Distingue si el ítem referenciado es un documento completo o un fragmento. En la primera fila de la tabla I se muestra un ejemplo de una referencia a un *documento*. En la segunda fila, sin embargo, se trata de una referencia a varios *fragmentos*.



*Rango.* Diferencia entre referencias *internas*, al mismo documento, y *externas*, a otros documentos. La referencia de la primera fila de la tabla I es externa, mientras que la referencia interna de la tercera fila redirige a otro artículo del mismo documento en que se encuentra (Reglamento 44/2001).

*Número.* En este caso se considera si la referencia atañe a un único elemento de información (referencia *simple*) o a varios (referencia *múltiple*).

*Contexto.* Este criterio recoge si es necesario tener algún conocimiento del contexto para resolver la referencia, como documentos referenciados previamente cerca de la referencia, etcétera. Si no es necesario conocer el contexto, se trata de referencias *directas*. En otro caso, se trata de referencias *relativas*. En la tabla I se muestran dos referencias directas y otras dos relativas, donde se necesita el contexto para saber en qué documento se encuentran los artículos mencionados.

Referencia	Tipo	Rango	Número	Contexto
Decisión 1999/468/CE del Consejo, de 28 de junio de 1999	Documento	Externa	Singular	Directa
Artículos 68, 72, 73 y 74 de la Ley de Enjuiciamiento Civil	Fragmento	Externa	Múltiple	Directa
Sin perjuicio de lo dispuesto en el artículo 4	Fragmento	Interna	Singular	Relativa
Convenios y Tratado mencionados en el artículo 69	Documento	Externa	Múltiple	Relativa

*Tabla I. Algunas referencias de ejemplo y su clasificación según la taxonomía presentada en la sección 4.1 (extraídas del Reglamento (CE) n.º 44/2001 del Consejo de 22 de diciembre de 2000 relativo a la competencia judicial, el reconocimiento y la ejecución de resoluciones judiciales en materia civil y mercantil).*

## 5. Aplicaciones de usuario

Varias aplicaciones de usuario facilitan la utilización de las herramientas de extracción de información, la supervisión de sus resultados o el acceso a los documentos de los cuales se ha extraído esa información. Las aplicaciones de las que disponemos actualmente permiten:

- Navegar por los documentos, utilizando su estructura interna y la red hipertexto construida a partir de las referencias (véase Martínez, Fuente y Derniame, 2003). La interfaz es web. Los documentos HTML se construyen automáticamente a partir los datos XML, usando XSLT.
- Acceder a la colección de documentos sobre conflictos de jurisdicción en la Unión Europea. Existen dos versiones de esta aplicación. La primera está diseñada para permitir a un usuario trabajar de forma autónoma con su propia colección de documentos y datos. En la segunda, un servidor mantiene

las colecciones de documentos y ofrece una interfaz multiusuario. Esta última se ha usado en algunos experimentos docentes, presentados por Vicente et ál. (2007). Actualmente existe una versión mejorada de esta, donde se han introducido algunas funcionalidades que se revelaron necesarias tras la experiencia, como la posibilidad de visualizar simultáneamente varios documentos para compararlos. La figura 3 contiene una captura de pantalla de la versión monousuario. La figura 4 corresponde a la comparación de documentos en la versión más reciente de la opción multiusuario.

- Revisar los resultados de la extracción y corregirlos si es necesario. La figura 5 muestra una instantánea de esta aplicación. En este caso el usuario ve simultáneamente el texto y las referencias encontradas, y puede seleccionar *editar* cada referencia para modificar alguno de sus atributos en caso de que haya errores.



Figura 3. Aplicación para manejar documentos y anotarlos teniendo en cuenta su estructura.

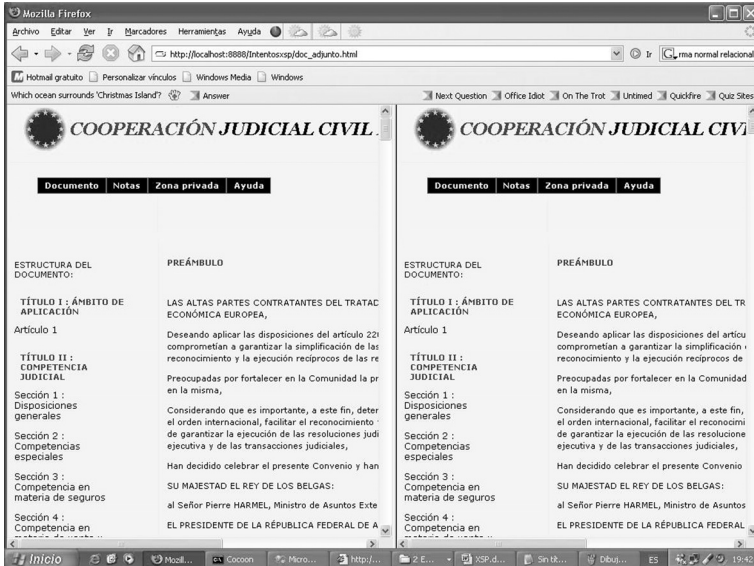


Figura 4. Comparación de documentos en la versión multiusuario de la aplicación que manipula documentos sobre conflictos de jurisdicción.

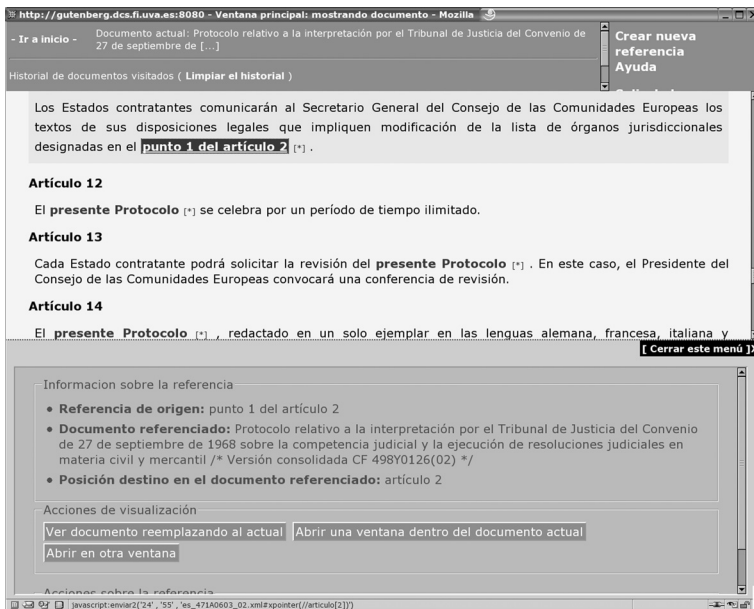


Figura 5. Supervisión de las referencias extraídas automáticamente mediante una aplicación.

## 6. Conclusiones

Se ha ofrecido una perspectiva general del trabajo desarrollado hasta ahora por el grupo de investigación de la Universidad de Valladolid en el campo de XML Legislativo.

El trabajo incluye una colaboración con el Ilustre Colegio de Procuradores de Valladolid, gestionada a través de un proyecto de investigación regional, VA1010B, financiado por la Consejería de Educación de Castilla y León (1). Actualmente prosigue la investigación, en la ampliación de las colecciones de trabajo, en el análisis de sus características y en la aplicación de nuevas técnicas con las que se pueda mejorar la funcionalidad de los sistemas de información jurídicos. Se incluyen aquí avances en la detección y resolución automática de referencias, y las búsquedas semánticas en varias de sus vertientes. Una de ellas son las búsquedas basadas en anotaciones semánticas y similitud semántica (véase Al Hulou y Napoli, 2004; Nauer y Napoli, 2006), utilizando las posibilidades que RDF, RDF Schema y OWL ofrecen. Además se plantea una búsqueda basada en conceptos, donde se utilizarán tesauros y se aprovecharán algunos resultados obtenidos por miembros del equipo en su investigación en el campo de la recuperación de información (Adiego, Navarro y Fuente, 2007).

## Notas

- (1) Creación de una aplicación de Web Semántica para el tratamiento y la manipulación electrónicos de los textos jurídicos de la Unión Europea y el Espacio Económico Europeo (en su versión en español) en materia de conflictos internacionales de jurisdicción. Proyecto de investigación VA1010B06. Junta de Castilla y León, 2006-2007.

## Agradecimientos

El *software* que implementa la extracción de referencias fue desarrollado por Sandra Muñoz Mínguez, Nuria Serrano y Yoana Martín. Víctor Manuel Merayo implementó la aplicación monousuario presentada en la figura 4. Matías del Hoyo y Pablo Encinas han desarrollado el prototipo que aparece en la figura 5. Jesús Requejo se encargó de la aplicación que facilita la supervisión de referencias. Todos ellos son o han sido alumnos de la Universidad de Valladolid.

## Referencias

- Adiego, Joaquín; Navarro, Gonzalo; Fuente, Pablo de la (2007). Using structural context to compress semistructured text collections. // *Information Processing and Management*. 43:3 (mayo 2007) 769-790.
- Al Hulou, Rim; Napoli, Amedeo (2004). Utilisation des connaissances pour l'aide à la recherche documentaire fondée sur le contenu. // *Extraction et Gestion des Connaissances, ECG 2004* (Clermont-Ferrand, Francia, enero de 2004). *Revue des Nouvelles Technologies de l'Information*. (2004) 503.

- Andrés Rivero, J. de; Gómez Skarmeta, A. F. (2007). CRNOLEX: A system for a dynamic representation of laws. // Biagioli, C.; Francesconi, E.; Sartor, G. (eds.). Proceedings of the V Legislative XML Workshop. Florencia: European Press Academic Publishing, 2007. 99-106.
- Arellano Pardo, M. C.; Nogales Flores, J. T.; Martín Galán, B. (2003). La organización hipertextual del ordenamiento jurídico. Posibilidades de XML y estándares relacionados. // Revista General de Información y Documentación. 13:2 (2003) 181-191.
- Arellano Pardo, M. C. (2005). Aportaciones de la técnica legislativa y XML a la informática jurídica documental. Madrid: Universidad Carlos III, Boletín Oficial del Estado, 2005.
- Arnold-Moore, Timothy; Clemes, Jane (2000). Connected to the Law: Tasmanian legislation using EnAct. // Journal of Information, Law and Technology. 1 (2000). [http://www2.warwick.ac.uk/fac/soc/law/elj/jilt/2000\\_1](http://www2.warwick.ac.uk/fac/soc/law/elj/jilt/2000_1) (2008-02-03).
- Boer, Alexander; Hoesktra, Rinker; Winkels, Radboud; Engers, Tom M. van; Willaert, Frederik (2002). Proposal for a Dutch legal XML standard. // Lecture Notes in Computer Science (LNCS). 2456 (2002) 3-12.
- Bolioli, A.; Dini, L.; Mercatali, P.; Romano, F. (2002). For the automated mark-up of Italian legislative texts in XML. // Bench-Capon, T. J. M. (ed.). Legal knowledge and information systems. JURIX 2002. 15.<sup>th</sup> annual conference (Dec., 2002). Ámsterdam: IOS Press, 2002. 21-30.
- Finke, N. (1997). TEI extensions for legal texts. // Text Encoding Initiative Tenth Anniversary User Conference. Providence: Brown University, 1997.
- Gómez Skarmeta, A. F. (2000). Una arquitectura de metadatos para la gestión de información en el web. // Novática. 146 (2000) 9-11.
- Grupo de Estudios de Técnica Legislativa (1989). Curso de Técnica Legislativa GRETEL. // Madrid: Centro de Estudios Constitucionales ("Serie de Técnica Legislativa", I), 1989.
- Marchetti, Andrea; Megale, Fabricio; Seta, Enrico; Vitali, Fabio (2002). Using XML as a means to access legislative documents: Italian and foreign experiences. // ACM SIGAPP Applied Computing Review. 10:1 (2002) 54-62.
- Martínez González, M.<sup>a</sup> Mercedes; Fuente, Pablo de la; Derniame, Jean-Claude (2003). XML as means to support information extraction from legal documents. // International Journal of Computer Systems, Science and Engineering. 18:5 (2003) 263-277.
- Martínez González, M.<sup>a</sup> Mercedes; Fuente, Pablo de la; Vicente Blanco, Dámaso-Javier (2005). Reference extraction and resolution for legal texts. // Lecture Notes in Computer Sciences (LNCS). 3776 (2005) 218-221.
- Martínez González, M.<sup>a</sup> Mercedes; Fuente, Pablo de la (2007). Introducing structure management in automatic reference resolution: an XML-based approach. // Information Processing and Management. 43:6 (2007) 1808-1832.
- Nauer, Emmanuel; Napoli, Amedeo (2006). A proposal for annotation, semantic similarity and classification of textual documents. // Artificial Intelligence: Methodology, Systems, and Applications, 12.<sup>th</sup> International Conference (AIMSA 2006). Varna, Bulgaria. Proceedings. Lecture Notes in Computer Science (LNCS). 4183 (2006) 201-212.
- Palmirani, Monica; Brighi, Rafaella (2002). An XML approach for Italian legislation acts. // JURIX 2002 workshop on e-government. Londres: Institute of Advanced Legal Studies, 2002.

- Magnusson-Sjöber, Cecilia (1997a). Corpus legis: a legal document management project. // *International Journal of Law and Information Technology*. 5:1 (1997) 83-99.
- Magnusson-Sjöber, Cecilia (1997b). DTD development for the legal domain. // *SGML Sweden'97: Estocolmo*, 1997.
- Vicente Blanco, Dámaso-Javier (2007). Medios electrónicos de pago y jurisdicción competente en supuestos de contratos transfronterizos en Europa: Los criterios de competencia judicial del derecho comunitario europeo y su aplicación a las relaciones contractuales involucradas en los medios electrónicos de pago. // Mata Martín, Ricardo; Jabato, Antonio (eds.). *Los medios electrónicos de pago. Problemas jurídicos*. Granada: Comares, 2007. 269-318.
- Vicente Blanco, Dámaso-Javier; Martínez González, M.<sup>a</sup> Mercedes; Sánchez Felipe, Jose Manuel; Adiego Rodríguez, Joaquín (2007). Experiences on teaching International Private Law with the support of an e-learning tool. // *The LEFIS Virtual Campus Design. Albarracín meeting (18-19 May 2007)*. [http://www.lefis.org/meetings/workshops/2007/albarracin\\_2007/index.htm](http://www.lefis.org/meetings/workshops/2007/albarracin_2007/index.htm) (2007-07-11).
- Vitali, Fabio; Zeni, F. (2007). Towards a country-independent data format: the Akoma Ntoso experience. // Biagoli, Carlo; Francesconi, Enrico; Sartor, Giovanni (eds.). *Proceedings of the V Legislative XML Workshop*. Florencia: European Press Academic Publishing, 2007. 67-86.

Recibido: 2008-02-20. Aceptado: 2008-04-14