

Modelos formales para la definición estructural y semántica en documentos XML. Comparación de posibilidades en un corpus textual de documentación jurisprudencial.

Bonifacio Martín Galán, J. Tomás Nogales Flores, M^a del Carmen Arellano Pardo

Departamento de Biblioteconomía y Documentación
Universidad Carlos III de Madrid
{bmartin, nogales, marellan}@bib.uc3m.es

Resumen: Se presenta una comparación de soluciones a la modelización formal de la estructura y la semántica de documentos textuales, en concreto a una colección de sentencias del Tribunal Constitucional español, que se ha utilizado como material de experimentación de tecnologías XML en una tesis doctoral ya defendida. Se analizan y aplican los dos modelos principales de la tecnología XML para la definición formal de los documentos electrónicos, el modelo inicial de la DTD y el modelo más avanzado del XML Schema. Este último se ha revelado como una tecnología más potente para la expresión de restricciones tanto en las estructuras lógicas de los tipos documentales jurídicos como en la definición de determinados datos presentes en los mismos. Igualmente, se introducen otros modelos de esquema con clara orientación a la definición de ontologías o lenguajes de marcado semántico capaces de ofrecer una respuesta a las necesidades existentes para la construcción de la denominada Web semántica en Internet, caso del RDF Schema o, más recientemente, del DAML+OIL.

Presentación

La presente comunicación pretende dar a conocer una parte de las investigaciones llevadas a cabo durante estos últimos años por varios profesores adscritos al grupo de *Tecnologías de la Información* del Departamento de Biblioteconomía y Documentación de la Universidad Carlos III de Madrid sobre la aplicación de las tecnologías XML (*Extensible Markup Language*) a diversos tipos de documentos, entre ellos los jurídicos, para su almacenamiento, gestión y difusión en la Web. Entre otros ámbitos de aplicación práctica, estas tecnologías se han empleado en la maqueta desarrollada como parte de la tesis doctoral elaborada por Bonifacio Martín Galán y dirigida por J. Tomás Nogales Flores bajo el título de *Tratamiento y Difusión en Internet de Información Jurisprudencial Mediante Tecnologías XML: Aplicación al Caso del Tribunal Constitucional*, defendida en febrero de 2002 [1]. En esta tesis se abordaron desde el plano teórico y experimental un gran número de estándares integrados dentro de la familia de tecnologías XML, desde las concebidas para la definición de estructuras y

vocabularios de marcado (DTD, XML Schemas y RDF Schema) o las propias para el establecimiento de enlaces hipertextuales avanzados (XLink y XPointer) hasta las orientadas a la búsqueda y recuperación documental (XPath y XQuery), pasando por las diseñadas para dar un formato de presentación a los documentos XML (CSS y XSL-FO) o transformarlos (XSLT).

En esta comunicación sólo se abordarán las tecnologías XML que dan respuesta a las necesidades tanto de modelización de las estructuras documentales existentes en los documentos jurisprudenciales (sentencias) emanados del Tribunal Constitucional español, como de generación de un lenguaje específico de marcado capaz de definir semánticamente los elementos esenciales de estos textos, analizando los resultados obtenidos de la anterior investigación así como de los obtenidos en posteriores actualizaciones de la misma.

Metalinguajes de marcado y la Definición de Tipo de Documento

La gran aportación de los denominados lenguajes de marcado generalizado, o *metalinguajes*, radica en la inclusión de mecanismos para la generación de gramáticas específicas de marcado adaptables a la tipología documental manejada. Este concepto de metalenguaje de marcado no tendría su esplendor definitivo hasta 1986 con la aprobación como estándar internacional del denominado *Standard Generalized Markup Language* (SGML) (norma ISO 8879:1986).

Este metalenguaje de alcance internacional contempla en sí un par de tipos de sintaxis (abstracta y concreta) para un lenguaje de marcado, proporcionando un mecanismo normalizado para generar lenguajes de marcado descriptivo que pueden usarse, principalmente, para la descripción de la estructura de múltiples tipos de documentos. La sintaxis abstracta, materia de interés de esta comunicación, se usa para declarar las reglas que definen la inserción de marcas descriptivas en los documentos electrónicos. A esta expresión formalizada se la conoce por el nombre de *Document Type Definition* (DTD o Definición de Tipo de Documento).

De forma general, la DTD define las reglas de marcado para un tipo concreto de documento, estableciendo los nombres de los elementos empleados para crear las etiquetas de marcado (identificadores genéricos), los atributos que éstos pueden tomar, así como las reglas de subordinación, secuencia y frecuencia de aparición que se establecen para dichos elementos.

El metalenguaje XML (Recomendación del W3C desde 1998)¹, reducción y simplificación del SGML para la Web, hereda este mismo mecanismo de la DTD, si bien en este nuevo contexto se introduce el concepto de *documento bien formado* (correcto sintácticamente pero carente de una DTD) en oposición al de *documento válido* (que además es conforme a las reglas establecidas en una DTD), único posible en SGML y en el que nos centraremos en esta comunicación.

XML se ha venido aplicando desde su origen, y de forma general, bajo dos prismas distintos (aunque, a menudo, complementarios): aplicaciones en que los procesos de

¹ Toda la información oficial se encuentra disponible en <http://www.w3.org/XML/>

manipulación de la información contemplan al documento como un todo (*document applications*) siendo, por tanto, el ser humano su principal consumidor, y aquellas orientadas a la gestión de los datos contenidos (*data applications*), dirigiéndose la información tratada al consumo por parte de programas informáticos [2].

El modelo de la DTD de XML ha venido satisfaciendo las necesidades de la edición electrónica para construir lenguajes específicos de marcado. Si bien es cierto que pocos dudan de los beneficios que proporciona este modelo en el entorno tecnológico del XML, más aún desde la óptica de los profesionales de la información y la documentación científica [3], no es menos cierto que resulta insuficiente, como expondremos, para lo que esta tecnología pretende: XML está concebido de modo que ha de ser capaz de manipular tanto el texto de los documentos como cualquier otra forma de estructuración de datos. De este modo, surgen los denominados *XML Schemas*.

Modelización de la estructura lógica de los documentos jurisprudenciales mediante el modelo de la DTD

Como señalábamos, una de las líneas de investigación aplicada del citado equipo de trabajo incide en el análisis de la viabilidad de aplicación las tecnologías derivadas de la Web, especialmente sus lenguajes de marcado de textos electrónicos (desde el HTML hasta el metalenguaje XML), a los documentos jurídicos². Esta línea de investigación se inició en 1998 y desde entonces hasta la actualidad su continuidad ha sido permanente, teniendo como uno de sus últimos hitos más destacados en el campo de aplicación práctica de tecnologías puramente XML el desarrollo de la maqueta elaborada para la tesis doctoral ya citada³.

En esta comunicación nos centraremos en uno de los puntos capitales de este desarrollo práctico realizado mediante el uso de las tecnologías XML: la modelización de la estructura lógica de los documentos jurisprudenciales.

Los documentos jurídicos, y en especial los judiciales, se caracterizan por incluir un contenido netamente textual en el que la estructuración de sus bloques principales, sin llegar a existir una norma que los regule, suele ser bastante homogénea entre las distintas jurisdicciones (provinciales, autonómicas, nacionales y supranacionales). Además, en el caso del Tribunal Constitucional español, al tratarse de un tribunal de reciente creación (1978), la estructuración de los contenidos de sus resoluciones no ha sufrido grandes variaciones en el tiempo. En la investigación sólo se contempló el tratamiento de las Sentencias, pues sólo éstas pueden derivar en jurisprudencia, no así los Autos a pesar de que en su contenido pueda existir una fundamentación jurídica.

Estamos, pues, ante lo que anteriormente denominábamos aplicaciones del XML orientadas al documento (o al texto, si se prefiere) por lo que el empleo de las técnicas

² La bibliografía producida por este equipo de investigación ha sido considerable a lo largo de estos años. Véase, por ejemplo, uno de los últimos artículos elaborados, resumen en cierta medida de todo el trabajo emprendido en estos años, en la referencia bibliográfica [4].

³ Esta maqueta está disponible en Internet en la dirección <http://damocles.uc3m.es/TC/>

clásicas procedentes del SGML en el campo del análisis y reconocimiento del documento para la definición de tipos documentales resultan idóneas.

Así, partiendo de las premisas generales establecidas por diversos especialistas, como J. André [5] o E. Maler [6], para el desarrollo y confección de DTDs, se fue acotando esta primera fase de la investigación emprendida con los estudios y conclusiones de otros investigadores, como A. Salminen [7], C. Magnusson [8] o Y. Marcoux [9], que habían abordado con anterioridad, y desde la aplicación del SGML, la modelización de las estructuras de los documentos jurídicos. De igual modo, se analizaron las experiencias internacionales más recientes con aplicación de las tecnologías XML a este tipo especial de documento jurídico, siendo destacables por la similitud con las investigaciones de nuestro grupo los trabajos de normalización abordados por la organización internacional *Legal XML*⁴ y, fundamentalmente, los trabajos emprendidos por el Consejo Constitucional de Francia enmarcados dentro del ambicioso proyecto nacional SPAD (*Service Public d'Accès au Droit*)⁵.

Con ello se procedió a modelizar la estructura lógica de las Sentencias del Tribunal Constitucional español, y a desarrollar una DTD basada en XML válida para definir formalmente su estructura y su vocabulario de marcado. Los resultados finales más significativos pueden resumirse en los siguientes puntos⁶:

- La estructura formal o de composición del texto de las sentencias analizadas es homogénea pues en todas se mencionan en primer término los datos identificativos de la resolución (Sala o el Pleno que la dicte, número de Sentencia, fecha de emisión, tipo de recurso que la motiva y el acto), seguidos de antecedentes de hecho, fundamentación jurídica, decisión adoptada, fecha, y en su caso, votos particulares de uno o varios magistrados.
- También la redacción de las sentencias analizadas es bastante homogéneo, pero no hay una normalización adecuada de algunos aspectos como las citas de normas legislativas, resoluciones de tribunales inferiores o las propias del Tribunal Constitucional.
- Para la confección de la DTD ha sido necesario incluir un bloque estructural inicial (la cabecera de las sentencias) para dar cabida a elementos que no figuran directamente en el texto de los documentos y convenía incluir para su control tanto documental (autor, título, fecha de publicación en el BOE, descriptores de materia, resumen del recurso, resumen del fallo, etc.) como informático (nombres de los ficheros electrónicos completo y abreviado, número de identificación, etc.).
- La DTD incluye, además de la citada *Cabecera*, otros bloques principales o de primer nivel: *Preámbulo* (con los datos relativos a las partes en litigio, el nombre del magistrado ponente, el asunto y la fecha de la causa tratada), *Antecedentes* de hecho (asunto tratado e la historia del hecho judicial hasta su revisión por el tribunal), *Fundamentos jurídicos* (principios doctrinales y normativa jurídica aplicada en la resolución), *Fallo* del Tribunal (decisión del tribunal y su correspondiente

⁴ Información disponible en <http://www.legalxml.org/>

⁵ Información disponible en http://www.atika.pm.gouv.fr/XML/repertoire_de_schema/spad.shtml

⁶ Esta extensa DTD o lenguaje específico de marcado de sentencias del Tribunal Constitucional se encuentra disponible en <http://damocles.uc3m.es/TC/docsDTD/SentenciaTC.dtd>

disposición) y *Votos* particulares (si algún o algunos magistrados disienten del fallo), definido como opcional.

- Cada uno de estos bloques se ha descompuesto en bloques estructurales inferiores (por ejemplo, el Preámbulo se descompone en un apartado de *Composición* de la Sala y en un apartado de *Introducción* a la Sentencia), hasta el nivel de detalle preciso para el marcado final de las sentencias (como, por ejemplo, magistrados que componen la Sala o Pleno, tipo de recurso interpuesto, sus promotores, representantes y defensores de éstos, personas físicas o jurídicas que comparecen, numeración de antecedentes y fundamentos jurídicos, normas y sentencias citadas, magistrados que formulan un voto particular, etc.).
- La DTD incorpora, no sin ciertos problemas, el vocabulario propio de XLink para reflejar en el marcado los correspondientes enlaces hipertextuales (en este caso, enlaces simples) a normas y sentencias citadas; incluye asimismo diversos atributos en muchos de los elementos para normalizar la forma de entrada de información (tipo de recurso, Sala, fechas, nombres, etc.) o para incluir información necesaria para el control, representación y recuperación de los documentos (gran número de códigos de identificación, de descripción y de formato).

Limitaciones del modelo de la DTD y migración al modelo de esquema XML

A pesar de las innegables aportaciones del modelo de la DTD de XML para la definición de estructuras formales y vocabularios de marcado de documentos con contenido principalmente textual, éste se reveló insuficiente. Si bien la DTD obtenida permitió establecer un primer marcado de las sentencias del TC seleccionadas algunas de sus limitaciones pronto se hicieron patentes. Por ejemplo, la ambigüedad del contenido de elementos definidos en la DTD como de contenido mixto (texto mezclado con otros subelementos, que pueden aparecer o no, en cualquier orden y en un número de veces indeterminado), potencial fuente de errores durante el marcado del texto. O las limitaciones en la definición de los tipos de datos que deberían contener elementos y atributos, que, simplificando, sólo es posible definirlos como datos de carácter o, sólo para algunos atributos, una lista de valores predeterminados. Igualmente, el modelo de la DTD no está pensado para su integración con el estándar del *XML Namespaces*, creado para posibilitar la incorporación al vocabulario desarrollado otros vocabularios externos. De este modo, y para poder imponer restricciones en el modelo de contenido mixto así como en la tipología de los datos (números enteros, rangos de valores, fechas, códigos alfanuméricos, etc.), especialmente en los valores que habían de tomar muchos de los atributos, e incorporar otros elementos externos (los propios del lenguaje XLink) se hacía inevitable usar el modelo más evolucionado del *XML Schema* del W3C⁷.

⁷ La información oficial relativa al modelo de esquema XML del W3C se encuentra disponible en <http://www.w3.org/XML/Schema>

Los esquemas XML dan un paso más allá en la evolución del XML permitiendo superar las limitaciones del modelo de la DTD al ofrecer un mecanismo mucho más potente y expresivo para que las aplicaciones web puedan intercambiar datos de forma más consistente sin tener que renunciar a mecanismos *ad hoc* de validación de los mismos. De este modo, frente a un modelo de tipos de datos orientado al texto, como son los soportados por la DTD de XML, se contraponen un modelo con fuerte orientación a objetos, influenciado enormemente por los informáticos desarrolladores de bases de datos [10]. Además de esta diferencia de base, otras características hacen del *XML Schema* un modelo más moderno y robusto para esta modelización y definición formal de las estructuras y vocabularios para documentos XML; así, a diferencia de la DTD, que utiliza una sintaxis propia, los esquemas utilizan la misma sintaxis de los documentos XML; permiten definir con mayor riqueza y complejidad estructuras internas; contemplan los tipos de datos soportados por las DTD pero los amplían con tipos de datos específicos (decimal, entero, fecha y hora, etc.); permiten hacer uso de otros estándares XML acompañantes como XPath, XSL, XML Namespaces, XLink, etc., y establecer restricciones sobre los mismos de interés para el usuario, ampliando sus capacidades de definición y actuación [11].

Con todo ello, en una segunda fase de la investigación se trasladó la definición de tipo documental elaborada según el modelo de la DTD al modelo del *XML Schema*⁸, estableciéndose convenientemente los modelos de contenido de ciertos elementos que quedaban demasiado abiertos y ambiguos para un marcado riguroso de las sentencias (en especial, los denominados “modelos de contenido mixto”). Igualmente se ajustó adecuadamente la tipología de datos que deberían llevar ciertos elementos y atributos (por ejemplo, `xs:date` para los atributos que debían contener fechas o `xs:integer` para los que debían contener un número entero). Por último, se establecieron restricciones a ciertos tipos de datos cuando sus posibles valores debían ajustarse a normas estrictas de construcción (por citar un ejemplo, cada voto particular quedaba diferenciado dentro de un atributo identificador cuyo valor estaría formado por “VP” seguido del dígito que le correspondiese, empleándose para ello una restricción al tipo de dato `xs:ID` con el patrón `xs:pattern value="VP[1-9]{1}`).

⁸ Este esquema XML se elaboró en un principio siguiendo la especificación de la Recomendación Candidata del *XML Schema* del W3C, de octubre de 2000. Este primer esquema XML se encuentra disponible en <http://damocles.uc3m.es/TC/docsXSD/SentenciaTC.xsd>. Con la aparición en mayo de 2001 de la Recomendación final del W3C, este esquema ha sido mejorado y adaptado finalmente a dicho estándar. El esquema definitivo se encuentra disponible para su consulta en [http://damocles.uc3m.es/TC/docsXSD/SentenciaTC\(new\).xsd](http://damocles.uc3m.es/TC/docsXSD/SentenciaTC(new).xsd). Para una mejor comprensión del modelo desarrollado existe una representación gráfica del mismo en <http://damocles.uc3m.es/TC/docsXSD/esquema.png>, así como la explicación detallada de cada uno de los elementos, atributos, grupos y tipos de datos creados, accesible en <http://damocles.uc3m.es/TC/docsXSD/esquema.html>

Aproximación a los lenguajes de marcado semántico

La tercera y última fase de este apartado de la investigación hacía especial referencia a la asignación de un valor semántico universal al lenguaje de marcado desarrollado. Ya que este aspecto concreto de la asignación de metadatos a los documentos XML es otra línea de investigación dentro de nuestro Departamento, no ha sido tratado con el mismo detalle de profundidad que los anteriores, por lo que aquí se expone únicamente un acercamiento a esta materia.

Se trata éste de uno de los aspectos que más polémica ha suscitado entre los investigadores y especialistas de diversos campos de aplicación de las tecnologías XML. Si bien es cierto que los lenguajes de marcado descriptivo de textos electrónicos proporcionan una doble definición estructural, una *estructura abstracta*, que especifica cómo se ajustan las diferentes piezas que conforman el documento, y una *estructura semántica*, interesada en el significado del texto, tanto del conjunto como de cada una de las piezas que lo componen, algunos autores plantean dudas sobre la validez de dicha descripción semántica en el contexto de producción de los metalenguajes de marcado que han de servir de soporte tecnológico para la Web que se desea construir. Para el XML, algunos investigadores han venido matizando su doble funcionalidad para la descripción estructural y semántica de los documentos electrónicos. Así, R. Cover ya señaló en los inicios del XML que la marca descriptiva es en sí misma una forma de “metadatos” pues nos dice lo que el elemento es a través del nombre elegido y cómo los objetos informativos existentes dentro del documento se estructuran en un conjunto coherente y jerarquizado de bloques, pero matizando que XML sólo gobierna sobre la sintaxis, no existiendo mecanismos formales capaces de dar soporte a declaraciones restrictivas sobre la semántica de los elementos establecidos [12]. Más recientemente T. Bray, uno de los padres de XML, matizó que este metalenguaje no es capaz por sí solo de representar de forma adecuada los metadatos de un documento debido a los problemas de escalabilidad y de interoperabilidad de la metainformación en un entorno de trabajo adecuado para la Web [13]. En resumen, XML *per se* tan sólo permite definir semánticas parciales o particulares, válidas para la persona o colectivo que desarrolla o se adscribe a dicho lenguaje.

Bajo esta idea surge en el seno de la W3C un grupo de investigación para el desarrollo de la denominada *Web Semántica*⁹, encabezado por el propio Tim Berners-Lee, inventor de la Web y del HTML, donde el estándar RDF (*Resource Description Framework*)¹⁰ y sus desarrollos paralelos ocupan un papel principal.

No entraremos aquí en la descripción del modelo propuesto por RDF. Tan sólo indicaremos que RDF es más que una simple aplicación de XML, pues al igual que este metalenguaje, contempla mecanismos para que cada comunidad pueda desarrollar de forma independiente vocabularios propios de metadatos, que en este contexto de la

⁹ La información oficial sobre este importante grupo de trabajo se encuentra disponible en <http://www.w3.org/2001/sw/>

¹⁰ RDF, una de las primeras aplicaciones surgidas del metalenguaje XML, trata de proporcionar un marco de trabajo válido para describir recursos electrónicos e intercambiar metadatos que han de ser servidos principalmente en la Web. La información oficial se encuentra disponible en <http://www.w3.org/RDF/>

Web semántica suelen denominarse *ontologías*¹¹. Las descripciones de estos vocabularios se realizan a través del modelo de esquema RDF (*RDF Schema*)¹², capaz de establecer y formalizar las relaciones semánticas así como las restricciones que se dan entre los elementos del lenguaje definido.

Se observa, pues, un aparente conflicto entre estos dos desarrollos del W3C, los modelos de esquema XML y RDF, pues ambos pueden ser utilizados para construir vocabularios específicos de marcado. Ello está generando en la actualidad no pocas polémicas pues no siempre es fácil discernir si es más conveniente hacer uso de las clases y subclases propias del esquema RDF o, por el contrario, emplear las capacidades de representación estructural que posibilita el esquema XML.

En las investigaciones llevadas a cabo por nuestro grupo se ha tratado de integrar las propiedades y beneficios aportados por ambos modelos, siguiendo las investigaciones iniciadas por J. Hunter y C. Lagoze en las que promueven la cohabitación y complementariedad de ambos modelos dentro de un desarrollo común de metadatos para la WWW [14]. De los modelos propuestos, se consideró más oportuno el que incrusta localmente las semánticas propias del esquema RDF dentro de las anotaciones del esquema XML, empleando el conjunto de elementos definidos por la iniciativa del *Dublin Core*¹³ para el establecimiento de dicha semántica universal al vocabulario de marcado desarrollado, dado que es el lenguaje de mayor aceptación en el entorno de los metadatos para la Web.

Señalaremos por último que igualmente se analizaron otras iniciativas en la asignación de metainformación a los documentos XML [15], alternativas al modelo propuesto por RDF, como han sido los desarrollos *XML Topics Maps (XTM)*¹⁴, el *Meaning Definition Language (MDL)*¹⁵ o, el más prometedor de todos ellos, el *DARPA Agent Markup Language (DAML)*.

Sin duda, la estrella de los lenguajes de marcado semántico para el desarrollo de la Web semántica lo está constituyendo últimamente DAML¹⁶, iniciativa procedente de la agencia norteamericana para la investigación en Defensa, DARPA, en la que participan investigadores del propio W3C, incluido el mismísimo T. Berners-Lee. Se trata de un potente mecanismo de descripción en el que se integran diferentes piezas: por un lado se encuentra el denominado DAML-ONT, lenguaje capaz de establecer expresiones más sofisticadas que las definiciones de clases que proporciona el modelo de esquema de RDF, y por otro, se combina con la iniciativa denominada *Ontology Infe-*

¹¹El concepto de “ontología” no es fácil de definir en este contexto y está sujeto a numerosas polémicas y controversias. El término, heredado de la filosofía (metafísica), donde hace referencia al ser en general y sus propiedades trascendentales, aplicado al campo de la gestión del conocimiento es una “especificación explícita de una conceptualización” o, dicho de otro modo, una descripción formal de conceptos y relaciones que pueden existir para un agente o comunidad de agentes. Para una más amplia información al respecto de este tema véase la información suministrada en <http://www.semanticweb.org/knowmarkup.html>

¹² La información oficial se encuentra en <http://www.w3.org/TR/rdf-schema/>

¹³ La información oficial de esta iniciativa se encuentra disponible en <http://dublincore.org/>

¹⁴ Información disponible en <http://www.topicmaps.org/>

¹⁵ El borrador de esta especificación se encuentra en <http://www.charteris.com/mdl/>

¹⁶ La información oficial puede ser consultada en <http://www.daml.org/>

rence Layer (OIL), capaz de proporcionar potentes y sofisticados sistemas de clasificación, facilitando la definición de tipos de datos basados en el modelo de esquema de XML. La combinación de todos ellos ha dado lugar a la iniciativa conocida como DAML+OIL, que integra, como apuntábamos antes, los principales beneficios aportados por los modelos de esquemas RDF y XML. La potencialidad de esta nueva iniciativa ha quedado puesta de manifiesto en un reciente estudio elaborado por Y. Gil y V. Ratnakar del *Information Sciences Institute and Computer Science Department* de la Universidad del Estado de California, en el que se analiza y compara este lenguaje de marcado semántico con los modelos de esquema de XML y RDF a través de una serie de parámetros o dimensiones (modularidad, inclusión de subclases propiedades, tipos de datos primitivos, restricciones de las propiedades, herencia, etc.) resultando demoledora esta comparativa en beneficio del DAML+OIL [16].

Conclusiones

Como conclusiones a esta ponencia se puede señalar que el conjunto de tecnologías del XML para la Web resultan adecuadas para el tratamiento y difusión en Internet de información jurídica, como se ha demostrado en las diversas investigaciones realizadas y en curso en el seno del grupo de *Tecnologías de la Información* del Departamento de Biblioteconomía y Documentación de la Universidad Carlos III de Madrid. El modelo inicial de la DTD de XML para la construcción de lenguajes específicos de marcado descriptivo, aunque válido para un primer acercamiento a la definición estructural y semántica de los documentos jurídicos (en el caso de la investigación relacionada aquí, documentos jurisprudenciales emanados del Tribunal Constitucional español), resulta insuficiente para establecer un marcado riguroso debido a sus serias limitaciones. Se hace, por tanto, recomendable la migración hacia el modelo de esquema XML del W3C. Sin embargo, este modelo no es capaz de proporcionar una respuesta satisfactoria para el establecimiento de metadatos o, dicho de otro modo, para la asignación de una semántica universal al vocabulario elaborado. Para la construcción de una verdadera Web semántica se hace imprescindible obtener un modelo de esquema capaz de generar verdaderos lenguajes de marcado semántico. Si en un principio el modelo propuesto por el esquema RDF parecía válido por sí solo para llevar a cabo este cometido, algunas investigaciones recientes han apostado por la integración y combinación de ambos modelos de esquema del W3C para, de este modo, obtener los beneficios estructurales y de contenido que proporcionan los esquemas XML junto con los derivados del esquema RDF para establecer definiciones semánticas. Finalmente, se deben seguir muy estrechamente las investigaciones llevadas a cabo dentro de la iniciativa del DAML+OIL pues además de ofrecer dicha integración proporciona otros mecanismos propios auxiliares para la creación de potentes lenguajes de marcado tanto descriptivo como semántico.

Referencias

1. Martín Galán, Bonifacio. *Tratamiento y Difusión en Internet de información jurisprudencial mediante tecnologías XML: Aplicación al caso del Tribunal Constitucional* [Tesis doctoral]. J. Tomas Nogales Flores (dir.). Getafe: Dpto. de Biblioteconomía y Documentación, Universidad Carlos III de Madrid, febrero 2002.
2. Marchal, Benoît. *XML by Example*. Indianapolis: Que, 2000.
3. Martín Galán, Bonifacio; Rodríguez Mateos, David. "Estructuración de la información mediante XML: un nuevo reto para la gestión documental". En: *Jornadas Españolas de Documentación* (7ª. 2000. Bilbao). Bilbao: Universidad del País Vasco, 2000, pp. 113-123.
4. Nogales Flores, J. Tomás; Martín Galán, Bonifacio; Arellano Pardo, Mª del Carmen. "Informática, Derecho y Documentación. Experiencias y posibilidades de aplicación de los lenguajes de marcado de texto (SGML, HTML y XML) a los documentos jurídicos". En: Miguel Ángel Davara Rodríguez (coord.). *Encuentro sobre Informática y Derecho* (16º. 2002. Madrid). Madrid: Instituto de Informática Jurídica, Facultad de Derecho, UPCO, 2003. (en prensa).
5. J. André, J.; Furuta, R.; Quint, V. (eds.). *Structured Documents*. Cambridge: Cambridge University Press, 1989.
6. Eve Maler, Eve; El Andaloussi, Jeanne. *Developing SGML DTDs: From Text to Model to Markup*. Upper Saddle River, NJ: Prentice Hall, 1996.
7. Salminen, A. ; Lyytikäinen, V. ; Tiitinen, P. "Putting documents into their context in document analysis". *Information Processing and Management*, v. 36, nº 4, 2000, Pp. 623-641.
8. Sjöberg, Cecilia Magnusson. *Critical Factors in Legal Document Management: A study of standardised markup languages*. Stockholm: Jure AB, 1998.
9. Marcoux, Yves, Sévigny, Martin. "Why SGML? Why Now?". *Journal of the American Society for Information Science*, 1997, v. 48, nº 7, pp. 584-592.
10. Ioannides, Demetrios. "XML schema languages: beyond DTD". *Library Hit-Tech*, 2000, v. 18, nº 1, pp. 9-14.
11. Laurent, Simon St. *Describing Your Data: DTDs and XML Schemas* [documento HTML]. XML.com, December 1, 1999. Disponible en <http://www.xml.com/pub/1999/12/dtd/index.html> (consultado el 10 de septiembre de 2002).
12. Cover, Robin. *XML and Semantic Transparency* [documento HTML]. OASIS, rev. November 24, 1998. Disponible en <http://www.oasis-open.org/cover/xmlAndSemantics.html> (consultado el 12 de septiembre de 2002).
13. Bray, Tim. *What is RDF?* [documento HTML]. XML.com, January 24, 2001. Disponible en <http://www.xml.com/pub/a/2001/01/24/rdf.html> (consultado el 10 de septiembre de 2002).
14. Hunter, Jane; Lagoze, Carl. *Combining RDF and XML Schemas to Enhance Interoperability Between Metadata Application Profiles* [documento HTML] Queensland: DSTC, University of Queensland, November 2000. Disponible en <http://archive.dstc.edu.au/RDU/staff/jane-hunter/www10/paper.html> (consultado el 14 de septiembre de 2002).
15. Ahmed, Kal... [et al.]. *Professional XML Meta Data*. Birmingham: Wrox Press, 2001.
16. Gil, Yolanda; Ratnakar, Varun. "A Comparison of (Semantic) Markup Languages" [documento PDF]. En: *International FLAIRS Conference* (15th. 2002. Pensacola Beach). Disponible en http://trellis.semanticweb.org/expect/web/semanticweb/flairs02_comparison.pdf (consultado el 10 de septiembre de 2002).