

Definición y estilo de los objetos de información digitales y metadatos para la descripción

ANA M^a GARCÍA MARTÍNEZ
Universidad de Extremadura

Los sistemas de recuperación de información, que están en continua evolución debido al avance tecnológico, condicionan los modelos de recuperación de información, que provocan distintas líneas de investigación en este campo de especialización.

Una de estas líneas se sitúa en la evolución conceptual/procedimental, en la descripción de recursos para su localización o recuperación, nos referimos a los metadatos; aunque éstos se aplican a la descripción de recursos en red, no dejan de contribuir al mismo presupuesto que el de las técnicas de indización, puesto que en realidad se describe un objeto o recurso con la finalidad de recuperarlo, por lo que puede tener su campo de aplicación en documentos u objetos estructurados independientemente de su entorno de circulación.

PALABRAS CLAVE: Recuperación de Información, SGML, DTD, Metadatos, DSSSL, Formatos Metadatos, Descripción Recursos, Normalización

DEFINITION AND STYLE OF ELECTRONIC INFORMATION OBJECTS AND METADATA FOR DESCRIPTION

Information retrieval systems, ever evolving due to technological advances, condition information retrieval models and bring about different lines of research in this field of specialisation.

One such line concerns the conceptual/procedural changes in the description of resources for their retrieval. We refer here to metadata and although they apply to the description of network resources they constitute another indexing technique since their aim is to describe an object or resource for the purpose of its retrieval. As a result, they can also be applied to structured documents or objects regardless of their area of application.

KEYWORDS: Information Retrieval. SGML. DTD. METADATA. DSSSL. Resource Description. Standardisation.

INTRODUCCIÓN

Los sistemas de información tradicionales y los más novedosos o vanguardistas coinciden y siempre coincidirán en el objetivo último de los mismos, la satisfacción de las necesidades de información del usuario, de la forma más eficiente posible.

Para ello, en la corta historia de la ciencia de la documentación, se han ido aplicando una serie de técnicas, procesos, tecnología, metodologías, etc., que han sido descubiertas en el propio campo, o bien traídas de la mano de otras disciplinas, indirecta o directamente relacionadas, como es el caso de la informática como la más próxima hasta ahora, o la paleografía antes, pero tampoco hay que olvidarse de la lingüística o las matemáticas que cada día ocupan un papel más central en nuestro campo de estudio.

A pesar de esa diversidad de técnicas y procedimientos, siempre se ha constatado un fenómeno uniforme, la tendencia a incrementar la eficacia en la recuperación de información, y fundamentalmente de forma automática si fuera posible, o semiautomática, en mayor o menor medida.

Esta recuperación de la información necesita, previo al propio proceso como tal, ser representada y organizada, de forma tal que la configuración subyacente condicionará el resultado de la recuperación. En los sistemas automáticos y semiautomáticos, para llevar a cabo este proceso previo, generalmente se usan unos métodos de indexación. Los más usuales son los estadísticos y métodos lingüísticos, aunque también la combinación de ambos; tras los cuales y ya en el propio proceso de recuperación se aplica la lógica booleana u operadores de adyacencia o algún algoritmo de recuperación determinado, para realizar la búsqueda de la información deseada.

Pero también es necesario evidenciar la importancia que tiene el interfaz en la interacción hombre-máquina, tal y como propugnan White y McCain¹, el cuál es utilizado para llevar a cabo el proceso de diálogo entre el usuario y la máquina, es decir, la forma como se presenten los resultados y la disposición de los mecanismos para la estrategia de búsqueda condicionarán el mayor o menor éxito en la satisfacción del usuario.

Conforme transcurre el tiempo estos aspectos comunes que han venido manifestándose en los sistemas de recuperación de información, van a sufrir una alteración debido a la coincidencia de una serie de factores (como pueden ser entre otros, el incremento de la velocidad de proceso, de la capacidad de almacenamiento, aparición de nuevas técnicas informáticas, mejores protocolos, diferentes arquitecturas etc.), que vienen provocados por el avance tecnológico y que condicionan los modelos de recuperación de información, provocando distintas líneas de investigación.

1. White y McCain. The visualization of literatures. ARIST 1997

Una de estas líneas emergentes se está situando en el cambio o evolución conceptual/procedimental en la descripción de recursos para su localización o recuperación, fundamentalmente aplicados a recursos electrónicos en red (en particular la red Internet), nos referimos a los metadatos, que se pueden considerar los sustitutos de los datos catalográficos de los ficheros bibliotecarios. Aunque éstos se aplican a la descripción de recursos en red, no dejan de contribuir al mismo presupuesto que el de las técnicas de indización, puesto que en realidad se describe un objeto o recurso con la finalidad de recuperarlo en un futuro, por lo que se puede establecer como una línea paralela, sino confluyente, de desarrollo en la investigación documental para la mejora de la recuperación de información, aunque se ciña fundamentalmente al ámbito de la red, pero también tiene abierto su campo de aplicación a documentos u objetos estructurados independientemente de su entorno de circulación.

Otra vía de innovación o indagación que contribuye al avance de las investigaciones y la mejora en la recuperación de información, es la que afecta al modelo subyacente de los sistemas de recuperación de información; en este sentido una de las perspectivas o enfoques que nos encontramos, es la recuperación de información de documentos estructurados que, como demuestran algunos estudios², produce una recuperación mucho más eficaz que sobre documentos no estructurados.

A estas cuestiones también se suma el hecho de que existe una gran masa de información disponible en la red Internet, con una previsión de crecimiento acelerado en el futuro inmediato. En esta red Internet, que se está convirtiendo en el vehículo de telecomunicación por excelencia, el protocolo más útil para manipulación de documentos hipermedia es el WWW, en el que los documentos tienen estructura, que les es asignada a través del lenguaje de marcas HTML, o más recientemente XML, que es una DTD del metalenguaje SGML, creado para la definición de documentos estructurados y considerado como norma internacional ISO desde 1986.

Pero este metalenguaje, y por ende sus lenguajes derivados, suele ser engorroso de aplicar, por lo que para facilitar esta tarea de creación de documentos estructurados para WWW se elaboran unas hojas de estilo intermediarias entre el usuario y el documento, este proceso que suele ser arbitrario requería de una normalización para eliminar los problemas que conllevaba, para ello es por lo que se decide idear DSSSL, norma internacional ISO/IEC de 1996, con la que se puede crear el diseño

-
2. CALLAN, J. Passage-level evidence in document retrieval. *Conference on Research and Development in Information Retrieval* Dublin, 1994
 - Macleod, I. Storage and retrieval of structured documents. *Information Processing and Managements*, 26(2), 1990
 - SALTON, G., ALLAN, J. y BUCKLEY, C. Approach to passage retrieval in full text information systems. *Conference on Research and Development in Information Retrieval.*, Pittsburgh 1993

del documento para conseguir la estructuración interna a través de su lenguaje de estilo, y ofrece también la ventaja añadida de tener la capacidad de compatibilizar distintas DTDs de SGML a través de su lenguaje de transformación.

Todo ello se ve particularmente contextualizado en los entornos social, político y económico, teniendo en cuenta la evolución de los sistemas de información a la nueva realidad tecnológica, y por ende, a las distintas exigencias que se plantean en las necesidades de información de los usuarios, aunque más que el cambio en las necesidades afecta al cómo satisfacerlas con una mayor garantía de éxito; que como bien establece Arms³ para la biblioteca digital, la clave reside en la arquitectura de ésta como marco de propósito general para una biblioteca donde objetos de todo tipo de material sean accesibles en las redes informáticas nacionales, y donde la transición a la biblioteca digital desde los servicios de red actuales viene condicionada por ocho principios básicos:

1. Coexistencia del marco técnico con un marco legal y social.
2. La terminología obstaculiza la comprensión de los conceptos de biblioteca.
3. La arquitectura subyacente debería estar separada del contenido almacenado en la biblioteca.
4. Los bloques básicos de construcción de la biblioteca digital son los nombres e identificadores.
5. Los objetos de la biblioteca digital son algo más que colecciones de bits.
6. El objeto de biblioteca digital que se usa es diferente del objeto almacenado.
7. Los depósitos deben cuidar la información que ellos alojan.
8. Los usuarios quieren trabajos intelectuales, no objetos digitales.

En esta evolución a la biblioteca digital también debemos tener en cuenta el entorno de meta-información que sugiere Smit⁴, que está basado en la organización de la biblioteca con la finalidad de facilitar el acceso a las colecciones de información controladas. Para ello se establece una distinción entre la biblioteca tradicional y la digital en función de las características organizacionales.

Las bibliotecas tradicionales poseen tres características organizacionales que, juntas, suministran una base para tal acceso, que son:

- La organización de la información en objetos de información (OI) físicos tales como libros.

3. ARMS, William Y. Conceptos clave en la arquitectura de una biblioteca digital

4. SMITH, Terence R. Meta-information entorn of digital libraries

- La organización física de las colecciones de OI de acuerdo a varios atributos, tales como materia temática y autor.
- Un entorno de información organizada que facilita acceso directo a los objetos de información basándose en atributos tales como autor, título, y materia temática, así como un grado limitado de acceso indirecto a la información contenida en los OI.

Esta última característica de una biblioteca tradicional generalmente implica múltiples fuentes de información para soportar el acceso, tales como bibliotecarios, catálogos, y la manera en la que las colecciones están organizadas físicamente. Ya que implica información sobre información, se denomina a esta característica el entorno de meta-información de una biblioteca.

Tal y como se concibe actualmente, las bibliotecas digitales son bibliotecas en las que las colecciones están en forma digital y el acceso a ellas se basa casi completamente en la tecnología digital. Desde un punto de vista del usuario, la tecnología digital cambia las tres características organizacionales de la biblioteca tradicional. Primero la organización de información en OI físicos es reemplazable por una organización más flexible en OI lógicos. Segundo, la organización física única de una colección de OI es reemplazable con múltiples organizaciones lógicas de IO. El tercer cambio, y más significativo ocurre en el entorno de meta-información de una biblioteca, tiene la ventaja de que al tener el OI en forma digital permite el uso de la tecnología digital en la extracción de información desde el OI. La información extraída puede satisfacer las necesidades últimas de información o puede ser empleada por los 'bibliotecarios digitales' en la caracterización de los OI en la colección. En el último caso, esta meta-información se puede emplear para suministrar acceso a la información codificada en el OI. Como desventaja está la pérdida de las importantes interacciones entre bibliotecarios y usuarios que ocurren en los entornos de meta-información de la biblioteca tradicional con la automatización del acceso a la información en la biblioteca digital.

METADATOS

En este ambiente entra en juego los metadatos, asumiendo un papel relevante en lo que respecta a la descripción, la localización y la recuperación de objetos de información en un entorno de información digital.

La contribución de los metadatos al campo biblioteconómico, hasta el momento, viene centrándose más en la parte descriptiva de los objetos que en la de su recuperación, quizás influenciados por el modelo tradicional imperante hasta el momento, con lo que se obtiene un producto resultante de la aplicación de nueva tecnología mediante la aplicación de un modelo bibliográfico consolidado.

Esta tendencia puede venir provocada por la introducción vaga del concepto en nuestra disciplina, ya que se evidencia un estado de confusión actual sobre el término metadatos, que se desprende del abuso excesivo de este concepto para usarlo en una gran variedad de los contextos entre los que se aprecian, por ejemplo los diccionarios de datos y, el control de las operaciones de sistemas de gestión de bases de datos, también para describir conjuntos de datos compartidos por los científicos, y en las bibliotecas digitales como apoyo al usuario en el acceso a la información, el correo electrónico, los protocolos de recuperación de información, los registros de archivos, en Internet para datos descriptivos estructurados de los documentos digitales etiquetados.

Por tanto sería necesario una aclaración del término a través de una cualificación del concepto en esta disciplina. En este sentido una de las versiones más populares es la que considera los metadatos como datos para la descripción de recursos, que están difundidos por toda la red para caracterizar atributos de gente, servicios, componentes de software, etc; y como no, implícitamente esta versión demanda una aclaración del término recurso, que puede ser considerado, en la avanzadilla, como un conjunto complejo de datos y servicios que pueden ser opacos al usuario o robot, y que además se pueden caracterizar por ser oportunistas, es decir, atendiendo sólo en respuesta a una conjunción particular de eventos o una consulta particular, además de que también puedan ser mutables y dinámicos, como puede apreciar cualquier persona que utilice la red con cierta frecuencia. Además, nos encontramos con una variada tipología de recursos que demandan prácticas descriptivas diferentes, lo que nos lleva en primer lugar a hacer un planteamiento de reconocimiento de esa tipología de recursos y por tanto las cuestiones de nominación e identificación de los mismos, y en segundo lugar a constatar la tensión que existe entre esa necesidad de establecer la simplicidad en la descripción y la complejidad que arrojan las exigencias de los distintos tipos de recursos, a lo que se añade la descripción de las relaciones de unos recursos a otros de forma múltiple.

Desde la perspectiva del usuario, la finalidad en el uso de los metadatos se puede circunscribir al descubrimiento de recursos, ampliación de información del recurso o información específica sobre el mismo, lo que provocaría una tipología determinada de metadatos, que soportarían procesos como el descubrimiento, la conservación, el uso, la selección, etc.

Por lo que, la observación del funcionamiento y uso de los metadatos en los distintos contextos, arrojan ciertos criterios que pueden soportar la base para ofrecer una somera definición de los metadatos como datos asociados a objetos que ayudan a los usuarios potenciales a tener un conocimiento anticipado de la existencia o características de los objetos; todo ello teniendo en cuenta que soportan una gran variedad de operaciones y que un usuario puede ser un programa o una persona.

Circunscribiéndonos al ámbito bibliotecario, los metadatos incluyen los datos descriptivos o temáticos tradicionalmente incluidos en los catálogos de bibliotecas. por lo que se ha tomado como referente el modelo bibliográfico, para aplicarlo a la descripción de recursos en red.

En este modelo se identifican tres niveles de descripción que se corresponden con la copia, la publicación, y la obra. A nivel de obra estarían los datos del contenido intelectual, que puede incluirse en una variedad de publicaciones, y que suele referirse a información sobre responsabilidad de autoría, materia, género, título; el nivel de publicación se centra en la manifestación de la obra, es decir es el título de donde tradicionalmente se han sacado los datos para la descripción, por lo que incluirá datos sobre el editor, distribuidor, derechos de propiedad intelectual, formato, etc.; mientras que el nivel de copia se considera una manifestación particular de la publicación, que contemplará información sobre propiedad, localización, procedencia, condiciones de uso, características físicas.

Este modelo, por naturaleza mantiene datos comunes en cada nivel, de forma que los niveles más específicos heredan las propiedades de los niveles más generales, así la copia asume las propiedades de la publicación y de la obra

En el contexto biblioteconómico, además de la función descriptiva, los metadatos también pueden cumplir una función de normalización en la práctica bibliotecaria, como es la normalización tanto del contenido como de la estructura, que no se han reconocido en otros dominios, y que tan importante es para la consecución de los buenos resultados en la recuperación de la información.

En este sentido la función de normalización la cumplen los metadatos a través de la univocidad en la asignación de funciones a los distintos campos etiquetados para representar un concepto y, evidentemente, al hecho de normalizar el contenido de cada campo habría que sumarle también la normalización en la terminología a usar; ya que todos ellos son factores que influyen indiscutiblemente en la eficacia de la recuperación de información.

Y por lo que se refiere a la parte descriptiva, los datos de la catalogación tradicional no son suficientes para tener una buena catalogación en un entorno digital y en red, por lo que para tener una visión global de la información descriptiva en esa infraestructura de información sería necesario contemplar tipos de metadatos como los siguientes:

Términos y condiciones. Estos metadatos describen las condiciones de uso de un objeto. Pueden incluir una lista de acceso de quién puede visualizar el objeto, una “condición de uso” establecida que puede ser visualizada antes de permitir el acceso al objeto, un programa de precios y cuotas para usar el objeto, o una definición de usos permitidos de un objeto.

Datos administrativos. Metadatos relacionados con la gestión de un objeto en un servidor particular o depósito. Algunos ejemplos de información almacenada en datos administrativos son la fecha de la última modificación, fecha de creación, y la identidad del administrador.

Clasificación de contenidos. Esto es una descripción de los atributos de un objeto en un esquema de clasificación de escala multidimensional, asignado por alguna autoridad de clasificación; un ejemplo puede ser la conveniencia del contenido para varias audiencias. El subcomité técnico de PICS en el IETF es un esfuerzo para crear un marco para definir tales clasificaciones de contenido.

Procedencia. Este dato define la fuente de origen del contenido de algún objeto, por ejemplo la localización de algún artefacto físico desde el que se pueda escanear el contenido. También puede incluir un resumen de todas las transformaciones algorítmicas que se han aplicado al objeto (filtering, decimation, etc.)

Datos de enlace o relación. Estos son datos sobre la relación del contenido de un objeto a otros; ejemplos son las relaciones entre conjuntos de artículos y una revista que los contenga, entre una traducción y el trabajo en lengua original, entre una edición posterior y el trabajo original, y entre los componentes de un trabajo multimedia.

Datos estructurales. Estos son datos que definen los componentes lógicos de objetos compuestos o complejos y cómo acceder a estos componentes. Un ejemplo simple es una tabla de contenidos. Un ejemplo más complejo es la lista de componentes de un juego de software.

También hay que contemplar los metadatos desde otra perspectiva en nuestro campo, además de la catalográfica o descriptiva, que se materializa en el contexto tradicional y de Internet, con la finalidad de facilitar y mejorar la recuperación de información; la perspectiva científica que se desarrolla en un contexto académico con el objetivo de medir la recuperación en términos de exhaustividad y precisión, a través de metadatos que identifican conceptos clave de los recursos de información que mejoran esta ratio de efectividad de la recuperación de información.

Formatos de metadatos

Los metadatos que se crean para la gestión de recursos en red se conciben en sus inicios en el ámbito de comunidades especializadas y sectores del mercado que dan lugar a diferentes formatos de metadatos con propósitos de cubrir las necesidades específicas que requería el tratamiento de la información en su ámbito; esto da lugar a una variedad de formatos propietarios, que se caracterizan por el esfuerzo en desarrollar formatos especializados para adecuarlo al propósito perseguido, por la inversión en formación de personal y documentación para el conocimiento del formato, y por la inversión en el desarrollo de sistemas para manipular y suministrar servicios basados en estos formatos. Estas particularidades provocan un fuerte apego a los formatos propios, y por tanto una dificultad para renunciar a ellos y elegir

otros formatos con un mayor potencial uso por parte de diferentes comunidades, que además permitiera la interoperabilidad entre los sistemas y compatibilidad entre formatos.

Los formatos vigentes actualmente son variados dependiendo de las diferentes características que incorporen, pero para dar una idea de la tipología de formatos que actualmente están en uso, nos podemos remitir a la clasificación de Dempsey y Heery⁵, en la que clasifica los formatos de metadatos en tres bandas, en función de la complejidad de los mismos, aunque el criterio que da lugar a estas tres bandas clasificatorias sea el nivel de complejidad del formato por su riqueza semántica o estructural, también condicionan dicha clasificación aspectos tales como el entorno de uso, el método de creación, la función del registro, los protocolos de búsqueda y recuperación y, su estado como norma internacional.

Formatos generales

Los formatos más conocidos a nivel general, y también por ser los primeros que se concibieron son TEI y DOCBOOK, que han servido de modelo para el desarrollo de otros formatos, y a la vez han sido base para su aplicación a otros campos mediante la adaptación de su estructura al contenido del tipo de documento.

DocBook es una DTD de SGML diseñada para documentación de informática, particularmente libros y artículos sobre hardware y software. A pesar de que es una DTD robusta, ha sido adoptado por una creciente comunidad de autores que escriben libros de todas clases, debido a que su estructura principal se corresponde con la noción general de lo que constituye un libro.

TEI, es un formato orientado a la codificación de textos electrónicos de investigación científica, intentando abarcar una amplia variedad de tipos de texto científicos, que se consideran para permitir el establecimiento de un número mínimo de características que compartan todos los textos y permitan usarlos con conjuntos de características más especializadas definibles por el usuario. Lo que dio lugar al origen de este formato fue la constatación de que la mayoría de los esquemas de codificación reflejaban los intereses científicos de sus creadores y eran aplicables a un área temática únicamente, por lo que ninguno era lo suficientemente flexible o generalizable para aplicarlo a la codificación de materiales de un amplio espectro de aplicaciones científicas. El objetivo que se perseguía con la creación de este formato era especificar un formato de intercambio común para textos legibles por máquina, suministrando un conjunto de recomendaciones para codificar materiales textuales nuevos que especificarían las características a codificar y cómo representarlas, al tiempo que se documentaba el esquema de codificación.

5. DEMPSEY, Lorcan y HEERY, Rachel. Metadata: a current view of practice and use. En: *Journal of Documentation* v.55, n°2 (marzo 1998)

Formatos específicos

Para cubrir las necesidades específicas de metadatos, que vienen impuestas por el tipo de información que se utiliza en una organización o en un campo o dominio específico, hay que recurrir a la definición de formatos que se adecuen a esos requerimientos. En este caso nos encontramos con el desarrollo de formatos como por ejemplo FGDC para información geográfica, lo mismo ocurre para otras especialidades, y concretamente en el ámbito documental hallamos tres formatos significativos, que son DC, CIMI, EAD

CIMI es un formato que está ideado para su aplicación al ámbito museístico.

EAD (Encoding Archival Description), es un formato que cae en el campo archivístico, en concreto destinado a la descripción de instrumentos de descripción archivística.

DC (Dublin Core), es un formato simple, con una intención de aplicación genérica, pero que se ha gestado en el ámbito bibliotecario y por lo tanto su principal aplicación la tiene en este campo.

Este formato es particularmente atractivo en el campo bibliotecario, ya que recientemente la biblioteca en general y de investigación científica en particular ha experimentado una evolución en cuanto a la clase de material que proporciona, debido a que ahora son accesibles desde Internet grandes colecciones de textos, imágenes y sonidos de muchas comunidades científicas sólo existen de forma electrónica. A menudo el paradero y situación de este material solo es conocido de manera informal por los miembros de una comunidad determinada; y para los ajenos a la misma este material es difícil de e inasequible.

Con esto se constata la dificultad a la hora de encontrar muchos de estos objetos en Internet o la WWW, y a pesar de que los servicios de localización indexan los recursos disponibles en el web y actualizan las localizaciones de las BD, los índices resultantes son útiles en pequeñas colecciones pero conforme se expande su amplitud los índices sucumben ante problemas de recuperación de grandes conjuntos y de generalidades semánticas de disciplinas cruzadas.

Para mejorar la búsqueda y recuperación se necesitarán registros más ricos, creados por expertos de contenido. Esta riqueza la pueden suministrar normas formales tales como TEI y MARC, pero debido a la inversión de tiempo en su creación y mantenimiento sólo se aplicarán a los recursos más importantes.

Una solución alternativa a estos extremos sería la creación de un registro que sea más informativo que un índice pero menos completo que un registro de catalogación formal. Se describirían más objetos si sólo se requiriera un pequeño esfuerzo humano para crear los registros, especialmente si se animara al autor a crear la descripción. Y si la descripción del registro siguiera una norma establecida, sólo la

creación del registro requeriría intervención humana; las herramientas automatizadas descubrirían estas descripciones y las recogerían. Podría definirse un registro simple de metadatos que describa suficientemente un amplio rango de objetos electrónicos.

Estas son las premisas sobre las que se asienta DC.

El DC es un formato simple de descripción de recursos, que ha atraído una considerable atención, porque se ha situado como una potencial solución para tres requerimientos acuciantes. El primero es tener un formato simple de descripción de recursos aceptable en general que aloje la descripción de un amplio rango de recursos. El segundo uso al que se dirige es suministrar una base semántica para metadatos embebidos o adjuntados a documentos HTML. El tercer uso al que se dirige es suministrar una base para la interoperabilidad semántica de metadatos más ricos entre dominios. Los formatos de registros más ricos pueden representar un conjunto central de datos sobre DC que proporcionen un conjunto común de elementos para propósitos de descubrimiento

DC enfoca un aspecto de los metadatos, la descripción simple, pero hay una tensión evidente para ampliar el conjunto de elementos que posibiliten la descripción más compleja para dominios especializados particulares, así como extender los tipos de recursos que se describen, tales como material impreso.

Arquitecturas

La WWW da acceso sin precedentes a la información distribuida globalmente. Los metadatos o datos estructurados sobre datos, mejora el descubrimiento de y el acceso a tal información. Sin embargo el uso efectivo de metadatos entre aplicaciones requiere convenciones comunes sobre la semántica, la sintaxis, la estructura, y el diseño de infraestructuras que posibiliten la gestión de información y proporcione la capacidad de transformar el Web, en un recurso de información más útil y poderoso.

Las comunidades individuales de descripción de recursos definen la semántica, o significado, de los metadatos que tratan sus necesidades particulares. La sintaxis, la organización sistemática de los elementos de datos para el procesamiento por máquina, facilita el intercambio y uso de los metadatos entre aplicaciones múltiples. La estructura se puede considerar como una restricción formal sobre la sintaxis para la representación consistente de la semántica.

Para esto se necesita definir una arquitectura necesaria para soportar los metadatos del web, que imponga restricciones estructurales necesarias para suministrar métodos no ambiguos para expresar la semántica de la codificación consistente, el intercambio y el procesamiento informático de los metadatos. Además de suministrar medios para editar vocabulario, tanto legibles por el hombre como procesables por la máquina, diseñados para fomentar el intercambio, uso y extensión de la semántica de los metadatos entre comunidades de información dispares.

Las implicaciones potenciales de esta aplicación general redundará en que la búsqueda sobre el web será más fácil en cuanto que los motores de búsqueda tengan más información disponible, y de este modo la búsqueda se puede focalizar más. También se abrirán puertas para agentes software automatizados que recorran el web, buscando la información para nosotros o negociando por nosotros. El Web de hoy, la vasta masa de información no estructurada, en el futuro puede transformarse en algo más manejable, y de este modo algo mucho más útil.

Con el rápido incremento en el número y variedad de recursos en red, hay una necesidad creciente de una arquitectura que asocie diversos tipos de metadatos con esos recursos. Este requerimiento incrementa obviamente en el actual WWW donde las principales herramientas para encontrar recursos en red son “web-crawlers” o “spiders” que indexan el texto completo de páginas html. Aunque el valor de estas herramientas no debe desestimarse, sus deficiencias son obvias como muestran los resultados de efectividad en la recuperación de información

En este contexto los marcos más conocidos actualmente son WF y RDF.

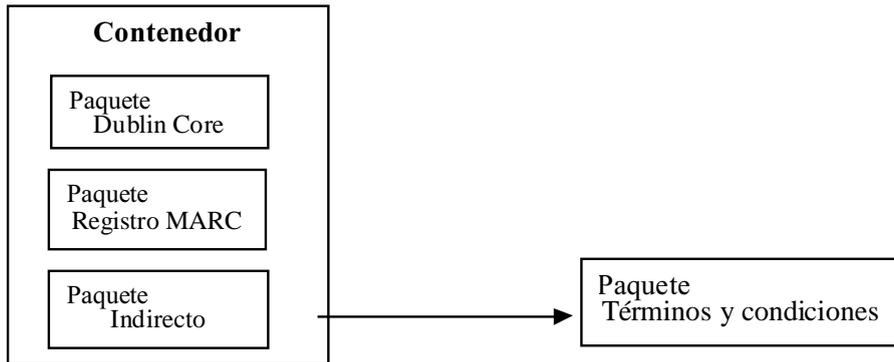
WF es una arquitectura contenedor, conocida como Warwick Framework, que se requería para el intercambio de paquetes de metadatos, en la que un paquete es concebido como un objeto de metadatos especializado para un propósito particular. Un paquete puede ser un registro DC, otro podría ser un registro MARC, también un registro de términos y condiciones. Los usuarios o agentes software deben tener la capacidad de agregar estos paquetes de metadatos discretos, de aquí la noción de arquitectura paquete-contenedor. Esta arquitectura deberá ser *modular*, es decir que tenga en cuenta objetos metadatos de diferente tipo, *extensible*, que tenga en cuenta los nuevos tipos de metadatos, *distribuida* permitiendo así referenciar los objetos metadatos externos y, *recursiva* que permita a los objetos de metadatos ser tratados como ‘contenido de información’, y a su vez, tener objetos de metadatos asociados a ellos.

WF, como arquitectura para agregar múltiples conjuntos de metadatos, tiene dos componentes fundamentales. Un *contenedor* es la unidad para agregar los conjuntos de metadatos tipificados, que se conocen como *paquetes*.

Mientras que el contenedor puede ser de dos tipos, transitorio (objeto de transporte entre depósitos, clientes y agentes.) y permanente, (se almacena en uno o más servidores y es accesible desde estos servidores usando un identificador accesible globalmente (URI)). Los paquetes son de tres tipos:

1. *Metadatos (conjunto)*: son paquetes que contienen los metadatos actuales. Algunos ejemplos de esto son los paquetes que son registros MARC, registros DC, y condiciones y términos codificados
2. *Indirecto*: es un paquete que es una referencia indirecta a otro objeto en la infraestructura de información.

3. *Contenedor*: es un paquete que él mismo es un contenedor. No hay límite definido para esta recursión.



El contenedor en este ejemplo contiene tres paquetes lógicos de metadatos. Los dos primeros, un registro DC y uno MARC, están contenidos en el contenedor como un par de paquetes. El tercer conjunto de metadatos, que define los términos y condiciones para acceder al objeto de contenido, se referencia indirectamente vía URI en el contenedor.

RDF, Resource Description Framework, desarrollado bajo los auspicios de W3C, es una infraestructura que posibilita la codificación, el intercambio, la reutilización de los metadatos estructurados. Esta infraestructura permite la interoperatividad de los metadatos a través del diseño de mecanismos que soportan convenciones comunes de la semántica, la sintaxis y la estructura. RDF no estipula la semántica para cada comunidad de descripción de recursos, sino que suministra la capacidad de definir tantos elementos de metadatos como necesiten estas comunidades.

RDF suministra un modelo para describir recursos, en el que los recursos tienen propiedades (atributos o características), y se define el recurso como cualquier objeto que es unívocamente identificable por un URI (Identificador Uniforme de Recursos).

Las propiedades asociadas a los recursos se identifican por tipos-propiedades, y estos tienen sus correspondientes valores, los tipos-propiedades expresan la relación que hay entre los valores asociados a los recursos, que por naturaleza pueden ser atómicos (cadena de texto, número, etc.) u otros recursos, que a su vez pueden tener sus propias propiedades, y una colección de estas propiedades que se refieren al mismo recurso se llama descripción.

Se requiere una sintaxis que represente este modelo para almacenar instancias en ficheros legibles por máquina y comunicar estas instancias entre aplicaciones.

RDF usa XML como sintaxis común para el intercambio y procesamiento de metadatos, e impone una estructura formal sobre XML para soportar la representación consistente de la semántica.

RDF tiene en cuenta la capacidad de las comunidades de descripción de recursos para definir la semántica; sin embargo, es importante desambiguar esta semántica entre comunidades. Por ejemplo la propiedad tipo “autor” puede tener un significado más general o más específico dependiendo de las necesidades de las diferentes comunidades. Lo que sería problemático es que múltiples comunidades usaran la misma propiedad tipo para significar diferentes cosas. Para prevenir esto, RDF únicamente identifica propiedades-tipo mediante el mecanismo namespace. Namespace XML suministra un método para identificar sin ambigüedad la semántica y las convenciones que gobiernan el uso particular de las propiedades tipo, únicamente identificando la autoridad gobernante del vocabulario.

Los esquemas RDF se usan para declarar vocabularios, los conjuntos de propiedades-tipo de la semántica definidos por una comunidad particular. Los esquemas RDF definen las propiedades válidas en una descripción RDF dada, además de cualquier característica o restricción de los valores mismos de la propiedad-tipo. El mecanismo namespace XML sirve para identificar Esquemas RDF.

Una descripción de un esquema RDF, humana y procesable por máquina, puede accederse mediante la referencia del esquema URI. Si el esquema es procesable por máquina, una aplicación tiene la posibilidad de aprender algo de la semántica de las propiedades-tipo nombrados en el esquema. Comprender un esquema RDF particular es comprender la semántica de cada una de las propiedades en esa descripción.

Los esquemas RDF están estructurados basándose en el modelo de datos RDF. Por lo tanto, aunque una aplicación no entienda un esquema particular será posible analizar la descripción en la propiedad tipo y los correspondientes valores y podrá transportar la descripción intacta

LENGUAJES DE MARCADO

El procedimiento para marcar documentos que produjeran una presentación visual distinta de la estructura y atributos del documento, es decir un formato determinado, ha ido variando, desde los procesos puramente manuales en la época de la composición tipográfica donde el editor marcaba un manuscrito con instrucciones específicas para que lo ejecutara un compositor, y en el que no había ninguna conexión entre las instrucciones y la estructura de los documentos; en una fase posterior los sistemas automatizados continuaban este modelo añadiendo el proceso específico de marcado al fichero de documento legible por máquina; este marcado constaba de instrucciones de procesamiento específico, aunque ahora estaban en lenguaje de un programa de formateo, más que en un compositor humano, pero

tenía el inconveniente de que el fichero no se podría usar para un propósito diferente, o sobre un sistema informático diferente, sin cambiar todo el marcado. Con la evolución de los procesadores se desarrollan modelos que palian este problema, introduciendo llamadas a macros o a formato para identificar localizaciones en el documento donde ocurría el procesamiento, por lo que las instrucciones de procesamiento estaban fuera del documento almacenadas en procedimientos donde se podrían cambiar más fácilmente.

El siguiente paso da lugar a la aparición de la codificación genérica como práctica de formateo, al establecerse los ‘identificadores genéricos’ de los tipos de elementos en lugar de las llamadas a macros como nombres que sugerían un procesamiento particular, que vino provocado por la constatación de que las mismas estaban al comienzo o final de los elementos de documento.

La codificación genérica fue un gran paso hacia la construcción de los sistemas de procesamiento de texto automatizado, que refleja la relación natural entre los atributos del documento y el procesamiento.

El advenimiento de los “lenguajes de marcado generalizado” en los comienzos de los 70 llevó esta tendencia más allá suministrando una base de lenguaje formal para la codificación genérica.

Un lenguaje de marcado generalizado contempla dos principios fundamentales:

- a) Predomina el marcado descriptivo y se distingue de las instrucciones de procesamiento.

El marcado descriptivo incluye identificadores genéricos y otros atributos de los elementos del documento que dan lugar las instrucciones de procesamiento. Las instrucciones de procesamiento pueden estar en cualquier lengua, normalmente están recogidas fuera del documento en procedimientos.

- b) El marcado es definido formalmente por cada tipo de documento.

Un lenguaje de marcado generalizado formaliza el marcado de un documento incorporando “definiciones de tipo de documento”. Las definiciones de tipo incluyen una especificación (como una gramática formal) de los elementos y atributos que pueden ocurrir en un documento y en qué orden. Con esta información es posible determinar si el marcado para un documento individual es correcto (esto es, cumple con la definición de tipo) y, también suministrar el marcado que está ausente porque puede deducirse sin ambigüedad de otro marcado que está presente.

SGML normaliza la aplicación de la codificación genérica y los conceptos de marcado generalizado. Suministra una sintaxis coherente y clara para describir cualquier cosa que un usuario elija para identificar dentro de un documento. El lenguaje incluye:

- Una *sintaxis abstracta* para el marcado descriptivo de los elementos del documento.
- Una *sintaxis concreta de referencia* que ciñe la sintaxis abstracta a cantidades y caracteres delimitadores particulares. Los usuarios pueden definir sintaxis de referencia alternativas para satisfacer sus requerimientos.
- Declaraciones de marcado que permiten al usuario definir un vocabulario específico de identificadores genéricos y atributos para diferentes tipos de documentos.
- Disposición para datos de contenido artificiales. En el marcado generalizado los datos son cualquier cosa que no está definida por el lenguaje de marcado. Esto puede incluir 'notaciones de datos de contenido' especiales que requieren interpretaciones diferentes según el texto general: fórmulas, imágenes, alfabetos no latinos, texto formateado previamente, o gráficos.
- Referencias de entidad: una técnica no específica del sistema para referirse al contenido localizado fuera de lo principal del documento, tal como capítulos escritos separadamente, fotografías, etc.
- Delimitadores especiales para las instrucciones de procesamiento, distinguiéndolo así del marcado descriptivo. Las instrucciones de procesamiento se pueden introducir cuando se necesiten para situaciones que no puedan manejarse con los procedimientos, pero que puedan ser fácilmente halladas y modificadas posteriormente cuando un documento sea enviado a un sistema de procesamiento diferente.

Para que un lenguaje de marcado generalizado sea una norma aceptable, requiere algo más que suministrar las capacidades funcionales requeridas. El lenguaje debe tener propiedades metalingüísticas, en orden a satisfacer las restricciones impuestas por la necesidad de usarlo en una multiplicidad de entornos. Estas propiedades que cumple SGML, se centran en:

- a) Los documentos 'marcados' con el lenguaje deben ser procesables por un amplio rango de sistemas de procesamiento de texto y de palabras.
- b) Se deben soportar los millones de mecanismos de entrada de texto existentes.
- c) No deberá haber dependencia del conjunto de caracteres, ya que los documentos deben ser tecleados sobre una variedad de mecanismos.
- d) No deberá haber dependencia de procesamiento, sistema, o mecanismo.
- e) No debe haber predisposición de lenguaje nacional.
- f) El lenguaje debe complacer las convenciones del procesador de palabra y la máquina de escribir familiares

- g) El lenguaje no debe depender de una cadena de datos u organización física del fichero.
- h) El texto marcado debe coexistir con otros datos.
- i) El lenguaje debe ser utilizable por humanos y programas.

Se establece la organización física de un documento SGML como estructura de entidad, y la organización lógica de un documento SGML como estructura de elemento, partiendo de la base de la consideración del documento como una estructura de varios tipos de elementos.

SGML se puede usar para documentos que son procesados por cualquier sistema de procesador de texto o palabra. Es particularmente aplicable a:

- a) Documentos que se intercambian entre sistemas con diferentes lenguajes de procesamiento de texto.
- b) Documentos que se procesan en más de una forma, incluso cuando los procedimientos usan el mismo lenguaje de procesamiento de texto.

Los documentos que solamente están en forma final formateada no entran en el campo de aplicación de esta norma internacional.

NORMALIZACIÓN DE ESTILO Y LENGUAJE DE TRANSFORMACIÓN

DSSSL es una norma internacional nominada como ISO/IEC 10179:1996 Tecnología de la Información –lenguajes de procesamiento– Lenguaje de especificación y Semántica de Estilo de Documento (DSSSL)

Esta norma internacional define el Lenguaje de Especificación y la Semántica de Estilo de Documento (DSSSL) usado para especificar el formateo y la transformación de documentos SGML. En origen DSSSL se centra en el formateo de medios electrónicos y de papel y sobre la transformación de documentos SGML marcados de acuerdo a diferentes DTDs. DSSSL se puede usar con cualquier documento SGML sin requerir modificaciones o restricciones de las definiciones de tipo de documento.

El principal objetivo de esta Norma Internacional es suministrar un lenguaje para expresar el formateo y otras especificaciones de procesamiento de documentos de una manera formal y rigurosa de forma que estas especificaciones puedan ser procesadas por un amplio rango de formateadores, de forma natural o usando mecanismos de traslación.

El lenguaje de estilo DSSSL permite a los usuarios especificar los tipos de formateo a aplicar a los objetos durante la composición.

El lenguaje de transformación DSSSL permite al usuario especificar la transformación de documentos desde una aplicación de marcado SGML en otra.

DSSSL está diseñado para especificaciones que se aplican a una clase de documentos. Estas especificaciones son aplicables a todos los documentos SGML posibles para aplicaciones SGML así como a documentos SGML particulares.

La normalización de la semántica de formateo se suministra en DSSSL a través de un conjunto de estructuras básicas conocidas como objetos de flujo y un conjunto asociado de características que se aplican a estos objetos. DSSSL suministra mecanismos para definir y extender las construcciones semánticas de forma que los diseñadores de aplicaciones DSSSL puedan construir las aplicaciones DSSSL que mejor se adapten a sus entornos de aplicación.

Los conceptos que hay tras DSSSL están asociados con el desarrollo de la codificación genérica y específicamente con SGML, el lenguaje de marcado generalizado estándar (ISO 8879).

Tradicionalmente se aplicaba la codificación específica a los textos electrónicos, que contenían códigos de control o llamadas a macros dando como resultado que el documento fuera formateado de una manera particular. En contraste la codificación genérica usa etiquetas descriptivas (por ejemplo 'heading' más que 'espacio 2 líneas; arial 12'), en la que lo importante del concepto de este tipo de codificación es la separación del contenido de información de los documentos del formato o apariencia del contenido. El concepto de codificación genérica ganó importancia en la década de los 70, consiguiéndose posteriormente con el desarrollo de SGML.

SGML suministra un lenguaje para modelar clases de documentos, pero no describe ningún modelo particular o conjunto de etiquetas predefinido. Una aplicación SGML se refiere a un conjunto de reglas (que constan principalmente de una DTD y su documentación de apoyo) que aplica SGML a una clase de documentos.

SGML normaliza la representación de la estructura del documento, dejando a los usuarios el desarrollo de sus propias técnicas para interactuar con los formateadores y otros procesadores, precisamente DSSSL está ideado para apoyar la segunda clase de aplicaciones suministrando una arquitectura normalizada para formateo y otras especificaciones de procesamiento, permitiendo a los usuarios intercambiar tales especificaciones en un marco normalizado.

Una especificación DSSSL es normalmente externa al documento SGML al cual se aplica, y así las múltiples especificaciones pueden ser aplicadas a los documentos SGML dados para producir varias presentaciones de los mismos datos.

SGML proporciona la capacidad de distinguir entre el contenido intrínseco y la estructura de un documento, de una parte, y de otra, las especificaciones para procesarlo. Con DSSSL, el formateo y otras especificaciones de procesamiento

pueden ser intercambiadas en conjunción con documentos SGML para suministrar la especificación normalizada de la visualización del documento mientras se preserve la distinción esencial entre contenido y formato.

Esta norma internacional está diseñada para especificar el procesamiento de documentos SGML válidos. DSSSL define la semántica, la sintaxis, y el modelo de procesamiento de dos lenguajes para la especificación del procesamiento de documentos:

- a) El lenguaje de transformación para transformar documentos SGML marcados de acuerdo a una o más DTDs en otros documentos SGML marcados de acuerdo a otras DTDs. La especificación de este proceso de transformación está completamente definido por esta norma internacional.
- b) El lenguaje de estilo, en el que el resultado se consigue aplicando un conjunto de características de formateo a partes de los datos, y la especificación es, a veces, tan precisa como requiera la aplicación, dejando algunas decisiones de formateo, tales como decisiones de final de línea o final de columna, a los procesos de composición.

El lenguaje de estilo DSSSL está ideado para ser usado en una amplia variedad de entornos con requerimientos tipográficos que varían desde las disposiciones simples de única columna a disposiciones complejas de múltiples columnas. Esta norma internacional proporciona los medios por los que una implementación puede exteriorizar características de estilo y otras técnicas para asociar información de estilo con un documento SGML.

DSSSL suministra un mecanismo para especificar el uso de ‘procesos externos’ para manipular datos. La naturaleza de estos procesos está fuera del alcance de DSSSL, pero puede incluir funciones de gestión de datos típicos, tales como ordenación e indexación; las funciones de composición típicas; y procesos multimedia y gráficos para datos no SGML.

Los documentos que ya han sido formateados o no contienen información estructural jerárquica o marcado genérico no están en el campo de aplicación de esta norma internacional.

DSSSL expresa especificaciones a realizar por un procesador que acepte un documento de entrada y produzca uno de salida. DSSSL es independiente del tipo de formateador, sistema de formateo, u otro procesador de transformación.

DSSSL incluye:

- a) Construcciones que suministran acceso a, y control de, toda la posible información marcada en un documento SGML, así como mecanismos para procesamiento de cadenas que permitan la manipulación de datos no marcados. Esto es suministrado por el componente SDQL (Standard Document Query Language) de DSSSL.

- b) Disposiciones para especificar la relación entre uno o más documentos SGML como entrada a un proceso de transformación y cero o más documentos SGML resultantes como salida del proceso.
- c) Disposiciones para especificar las relaciones entre los documentos SGML, expresados en la fuente DTDs, y el resultado del proceso de formateo. Disposiciones para describir el estilo tipográfico y la composición de un documento.
- d) Definiciones de una sintaxis procesable por máquina para la representación de una especificación DSSSL y sus varios componentes.
- e) Disposiciones para crear nuevas características DSSSL y sus valores asociados, así como nuevas clases de objetos de flujo. Estas son declaradas en las declaraciones en la parte del lenguaje de estilo de la especificación DSSSL.

Esta norma internacional está ideada para usar en una amplia variedad de entornos de aplicación SGML, incluyendo tanto edición electrónica como impresión convencional.

Una característica clave del marcado generalizado es que el formateo y otra información de procesamiento asociada con el documento está separada de las etiquetas genéricas embebidas en él.

En cualquier esquema de marcado generalizado, hay un método para asociar especificaciones de procesos con el marcado SGML. Este método de asociación permite que la información se una a las instancias específicas de los elementos así como a las clases generales de los tipos de elementos. El primer objetivo de DSSSL es suministrar un marco normalizado y métodos para asociar la información de procesamiento con el marcado de documentos SGML o parte de ellos.

DSSSL está ideado para usar con documentos estructurados como una jerarquía de elementos.

DSSSL mejora el formateo y otras especificaciones de proceso para asociarlas con estos elementos para producir un documento formateado para la presentación. Además durante el proceso de transformación DSSSL, la información de formateo puede ser añadida al resultado de la transformación. Esta información puede ser representada como atributos SGML; éstos, a su vez, pueden ser usados por el lenguaje de estilo para crear características de formateo con valores específicos.

DSSSL suministra cuatro áreas distintas de normalización:

- a) Un lenguaje y modelo de procesamiento para transformar uno o más documentos SGML en cero más documentos SGML distintos.

Este es llamado el lenguaje de transformación. Esta transformación está controlada por la especificación de transformación. Una especificación de transformación contiene una lista de asociaciones. Una asociación consta de tres

partes: las expresiones de consulta, las expresiones de transformación, las opcionales expresiones de prioridad. Funcionalmente, esta especificación permite al usuario especificar la creación de nuevas estructuras, la duplicación de estructuras existentes, y la reordenación y reagrupación de estructuras existentes.

- b) Un lenguaje para especificar la aplicación de características de formateo sobre un documento SGML.

El proceso que aplica el formateo y otras características de proceso relacionadas con el formateo a un documento SGML es denominado proceso de formateo. Este proceso es controlado por la especificación de estilo. Una especificación de estilo contiene una secuencia de reglas de construcción, de las que hay varios tipos.

- c) Un lenguaje de consulta, Standard Document Query Language, usado para identificar partes de un documento SGML.

SDQL es parte del lenguaje de transformación y de estilo de DSSSL. Se usa para navegar a través de la estructura jerárquica del documento SGML, identificando las piezas relevantes del marcado SGML y el contenido sobre el que se va a realizar el procesamiento.

- d) Un lenguaje de expresión

El lenguaje de expresión DSSSL es usado en SDQL, en el lenguaje de transformación y en el lenguaje de estilo. Se usa para crear y manipular objetos.

Modelo conceptual

El modelo conceptual tiene dos procesos distintos: 1) un proceso de transformación y 2) un proceso de formateo. Los dos procesos pueden usarse conjuntamente o por separado.

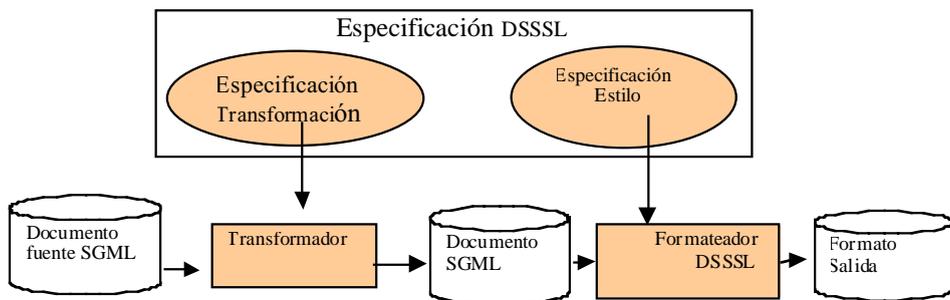
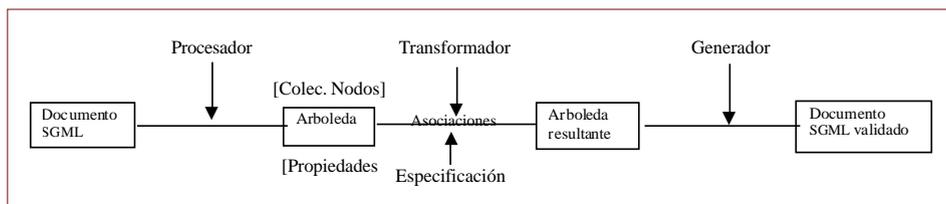


Ilustración del modelo conceptual

Cada uno de los procesos DSSSL está controlado por el apropiado lenguaje DSSSL. El lenguaje de transformación controla el proceso de transformación. De la misma manera el lenguaje de estilo controla los aspectos del proceso de formateo.

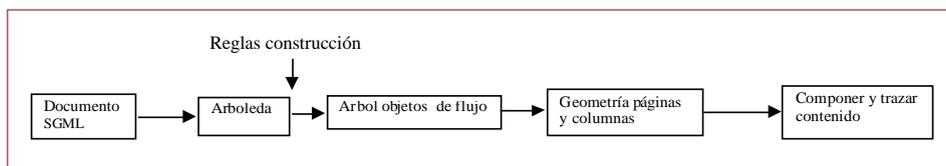


Esquema: Proceso de Transformación

El primer componente del proceso de transformación es el procesador de construcción de la arboleda, que actúa sobre el documento de entrada SGML al que analiza y representa por una colección de nodos llamada arboleda, en la que las relaciones de la misma son expresadas en términos de propiedades.

El segundo elemento que interviene es el transformador, que opera sobre la arboleda mediante la especificación de transformación, que consta de un conjunto de colecciones, en la que cada asociación especifica la transformación de objetos en el documento fuente en objetos en la arboleda resultante, que es el resultado de esta intervención.

El tercer componente del proceso es el generador que trabaja con la arboleda resultante para convertirla en documento SGML validado, para el intercambio, validación y entrada al proceso de formateo, en su caso.



Esquema: Proceso de Formateo

El proceso de formateo usa el mismo paso de construcción de la arboleda que el proceso de transformación para convertir el documento SGML en una arboleda de objetos estructurados jerárquicamente.

La arboleda es después procesada, usando las reglas de construcción, para crear un árbol de objetos de flujo que conste de objetos de flujo con el apropiado formateo y características de composición de página.

Posteriormente se define la geometría de páginas y columnas por características sobre el objeto de flujo de secuencia de página y objetos de flujo de secuencia de conjunto de columnas referentes a los modelos de página y de conjunto de columnas respectivamente.

Por último hay que componer y trazar el contenido basado en las reglas especificadas por la semántica de las clases de objeto de flujo y los valores de las características asociadas con estos objetos.

Estos procesos se llevan a cabo teniendo en cuenta que una arboleda es un conjunto de nodos construidos de acuerdo a un plan arboleda; cada nodo en la arboleda pertenece a una clase nominada en el plan arboleda, un nodo es un conjunto de asignaciones de propiedad, cada una de ellas consta de un nombre de propiedad y un valor de propiedad.

Un plan arboleda define un conjunto de clases y, para cada clase, un conjunto ordenado de propiedades. Para cada asignación de propiedad, hay una única propiedad que se corresponde con la clase del nodo cuyo nombre es el mismo que la parte nombre de la asignación de propiedad. Un plan arboleda especifica una selección de clases y propiedades de un conjunto de propiedades, en el que un conjunto de propiedades es definido por una definición de conjunto de propiedades expresado en SGML.

Para cualquier fuente de la arboleda, el juego de propiedades determina la *arboleda completa* que se construirá usando un plan arboleda que seleccione todas las clases y propiedades desde el juego de propiedades.

CONCLUSIONES

Ante el nuevo panorama de espacio de información que se nos presenta, fundamentalmente radicado en los entornos e infraestructuras de información digital, los profesionales de la información tenemos que aprovechar las nuevas oportunidades que se presentan de la mano de la tecnología, y fundamentalmente las destinadas a los objetos de información electrónicos, contexto en el que nos encontramos con herramientas relativamente recientes para la gestión de la información.

Por lo que frente a este nuevo espacio de información, el profesional se enfrenta con un nuevo reto en esa gestión de información, con unos objetos de información que son generados por terceros, pero que en esta nueva concepción de la gestión los intermediarios pueden también actuar como creadores de información aplicándole un valor añadido al objeto de información mientras que atraviesa la cadena de circulación; es decir, las instituciones de información pueden aplicar al objeto de información una serie de características que van a facilitar la identificación, la recuperación o la localización del objeto ante una determinada demanda de información para la que ese objeto se revela como pertinente a la necesidad expresada por el

usuario. Para llevar a cabo este cometido se pueden utilizar los metadatos, procedimiento que en el campo bibliotecario se ha utilizado más para la actividad descriptiva que cualquier otra de la gestión de información, pero que puede tener distintos usos potenciales, todos ellos encaminados al descubrimiento y recuperación de objetos de información de la forma más sencilla y eficaz posible. En este sentido se han desarrollado algunos formatos de metadatos específicos en el campo bibliotecario, que parece ser que están dando buenos resultados, aunque no están totalmente consolidados, debido a la fase de experimentación en que se encuentran, sí que están respaldados por la normalización internacional, por el hecho de ser definiciones de tipo de documento del metalenguaje SGML.

En esta normalización internacional se está desplegando un campo de actividad bastante dinámico, y es que además de las DTDs específicas de cada comunidad, se ofrece un marco para poder conjugar o trabajar con distintas DTDs de un mismo campo o de disciplinas diferentes, esto, puede hacerse mediante la construcción de arquitecturas de metadatos, permitiendo la interpretación de las distintos formatos de metadatos en una misma plataforma; pero la comunidad internacional ha ido más allá queriendo normalizar procesos tanto de estilo de objetos estructurados como de la transformación de los objetos codificados según un DTD en otra DTD, posibilitándolo a través de DSSSL.

En este espacio de información, parece obvio que en el campo biblioteconómico entrarán en juego toda una serie de herramientas para la gestión de información, algunas conocidas como es la parcela de la edición electrónica, pasando por otras más recientes como el marcado de objetos de información y arquitecturas de metadatos vigentes, así como aquéllas que son perfectamente desconocidas, como son los procesos de transformación y de estilo.

Por último no debemos olvidar que este marco técnico es necesario integrarlo con un marco social y legal, los cuales también están en continua evolución y adaptación a las nuevas realidades, observándolo de forma global bajo el prisma de la sociedad de la información.

BIBLIOGRAFÍA

- ADLER, Sharon C. The "ABCs" of DSSSL. *Journal of The American Society for Information Science*, 1997, vol. 48 (7)
- ARMS, Williams Y. Key concepts in the architecture of the digital library. *D-Lib Magazine*, 1995
- ALSCHULER, Liora. *ABCD... SGML: a user's guide to structured information*. International Thomson Computer Press, 1995
- CHILVERS, Alison y FEATHER, John. The management of digital datos: a metadata approach. *The Electronic Library*, 1998, vol. 16, n. 6

CATHRO, Warwick. Metadata: an overview.

<http://WWW.nla.gov.au/nla/staffpaper/cathro3.html>

DEMPSEY, Lorcan y HEERY, Rachel. Metadata: a current view of practice and use. *Journal of Documentation*, 1998, vol.55, n.2

DUCHARME, Bob. *SGML CD*. Prentice Hall, 1998

FAUSEY, Jon y SHAFER, Keith. All my data is in SGML. Now what?. *Journal of The American Society for Information Science*, 1997, vol. 48, n. 7

GERMAN, Daniel M. An introduction to DSSSL

– <http://csg.uwaterloo.ca/~dmg/DSSSL/tutorial.html>

– International Standard ISO/IEC 10179:1996. Information technology – processing languages – Document Style Semantics and Specification Language (DSSSL)

– International Standard ISO/IEC 8879:1986. Information processing – Text and office systems – Standard Generalized Markup Language (SGML)

LAGOZE, Carl. The Warwick Framework: a container architecture for diverse sets of metadata. *D-Lib Magazine*, july/august 1996

MILLER, Eric. A introducción to the Resource Description Framework. *D-Lib Magazine*, may 1998

MILSTEAD, Jessica y FELDMAN, Susan. Metadata: cataloging by any other name. *Online*, january/february 1999, vo. 25

PRESCOD, Paul. Introduction to DSSSL

– <http://itrc.uwaterloo.ca/~papresco/DSSSL/tutorial.html>

SMITH, Terence R. The Meta-Information Environment of Digital Libraries. *D-lib Magazine*, july/august 1996

Text Encoding Initiative

– <http://ota.ahds.ac.uk/teilight/teiu5f3.html>

– The ACH/ACL/ALLC Text Encoding Initiative: an overview

– <http://WWW-tei.uic.edu/orgs/tei>

WEIBEL, Stuart. Metadata: the foundations of resource description. *D-Lib Magazine*, july 1995

WHITE, Howard D. y McCAIN, Katherine W. The visualization of literatures. *Annual Review of Information Science and Technology (ARIST)*, 1997, vol. 32