

SGML/XML: Desarrollo en entornos documentales.

Autores:

José Luis Alonso Berrocal

Ángel F. Zazo Rodríguez

Carlos G. Figuerola

Universidad de Salamanca. Facultad de Documentación.

E-mail:

[berrocal | afzazo | figue]@gugu.usal.es

Dirección Postal:

Facultad de Documentación

C/ Francisco Vitoria, 6-16, 37008 – Salamanca

Tfno: + 34 923 294580

Fax: + 34 923 294582

Resumen:

Desde hace varios años, los denominados lenguajes de marca están adquiriendo un papel muy importante en el procesamiento y recuperación de la información, sobre todo desde que los documentos electrónicos están adquiriendo una gran importancia.

Muchos son los lenguajes de marca, pero destacan SGML y XML, que tienen diferentes desarrollos que los aplican, y que pretenden resolver problemas concretos.

1. Introducción

Cada vez con mayor frecuencia, estamos asistiendo a la necesidad de disponer de mecanismos que nos permitan estructurar adecuadamente la información en el entorno de los documentos electrónicos.

Los *lenguajes de marca* son hoy día una realidad en el ámbito documental, y pueden ofrecer respuestas interesantes a temas tales como la representación del conocimiento, las bibliotecas virtuales; pero además permiten establecer mecanismos útiles en el ámbito de la recuperación de información y en el intercambio de la misma.

2. SGML

Entre todos los lenguajes de marca destaca especialmente el SGML (Standard Generalized Markup Language), que con el paso del tiempo ha tenido diferentes implementaciones.

SGML se basa en el concepto de *marca*, que originalmente describía una anotación u otros símbolos para indicar a un maquetador de imprenta como diseñar una determinada página: tamaño de letra, negritas, centrados, etc. En el manual de estilo *The Chicago Manual of Style* define *markup* de la siguiente manera:

"Es el proceso de marcar un documento manuscrito con instrucciones acerca de cómo se deben utilizar los tipos de letra, los tamaños, los espacios, el sangrado, etc."

Con la producción automatizada de textos el término pasó a denominarse *Códigos de marca*, que se incluían en el texto electrónico para dar formato, imprimir y procesar un documento.

Las primeras ideas en la utilización de las marcas se dieron en la GCA (Graphic Communications Association), donde se reconoció la importancia del empleo de las marcas y promoviendo proyectos que analizaran su utilización. En 1967, William Tunnicliffe (pertenecía al comité de composición de la GCA) indicaba la necesidad de separar, dentro de un documento, la información de su formato.

En 1969 Charles Goldfarb (jefe de proyectos de investigación de IBM) junto a Edward Mosher y Raimond Lorie inventaron el Generalized Markup Language (GML) para permitir editar y formatear textos, así como para crear subsistemas de recuperación de la información que permitieran compartir documentos.

En 1978, gracias al comité de procesamiento de la información de la American National Standards Institute (ANSI) se dieron los primeros pasos para crear el SGML y convertirlo en un estándar. En 1985 se publicó un borrador del estándar que fue adoptado por la Oficina Oficial de Publicaciones de la Comunidad Europea y se publicó finalmente como norma ISO 8879:1986).

El SGML fue diseñado para permitir representar la información de manera sencilla, y que esta pudiera transferirse entre diferentes sistemas manteniendo la compatibilidad entre programas y plataformas. Se enfrentaba así a otros lenguajes de marca propietarios, dependientes de un software concreto o de una plataforma hardware determinada.

De esta forma SGML permite el intercambio de información entre distintas plataformas, soportes físicos, lógicos y entre diferentes sistemas de almacenamiento y presentación (bases de datos, edición electrónica, etc.), con independencia de su grado de complejidad.

Cuando se define un documento electrónico en SGML hay que tener en cuenta:

- El material que constituye un documento se puede distribuir en diferentes archivos, tantos como sean necesarios. Estos archivos además pueden estar almacenados en distintos ordenadores.
- Un archivo puede contener la portada, otro la introducción, otro una parte de una hoja de cálculo, otro un gráfico, otro un organigrama, otro la bibliografía, etc. En SGML, cada uno de estos objetos recibe el nombre de **entidad**. Las entidades se conciben como objetos independientes.
- Las entidades pueden tener cualquier tamaño, haber sido creadas por cualquier programa de software o estar guardadas en cualquier ordenador.
- Las entidades pueden estar compartidas por distintos documentos.
- Un documento estará definido en función de la **estructura de las entidades** que lo conforman.
- En el índice de materias de un documento no se encontrará ninguna referencia a los archivos que contienen las entidades.
- Las entidades se organizan en una *estructura lógica* de manera jerarquizada, en la que se definen conceptos como capítulos, tablas y párrafos y que configuran lo que se denomina **estructura de los elementos** del documento.

- Elementos y entidades pueden coincidir: un elemento lógico como *tabla* puede ser también una entidad en un archivo *hoja de cálculo*

El SGML está formado básicamente por las especificaciones de etiquetas que pueden utilizarse y por la definición de tipo de documento, y pasa a ser un lenguaje mediante el que podemos definir con precisión aquellos elementos que son necesarios en la elaboración de un documento, o de un grupo de documentos estructurados de forma similar. En la **Definición del Tipo de Documento** (DTD) se incluyen asimismo directrices o convenciones que conforman una aplicación SGML.

De esta forma, cuando enviamos un documento SGML podemos estar seguros de lo siguiente:

- que dicho documento comenzará con una **declaración SGML**, salvo que ésta se omita presuponiendo que tanto el sistema emisor como el receptor utilizan la sintaxis por defecto o la **sintaxis de referencia concreta**. La declaración SGML es un diagrama formal y normalizado que le indica al sistema receptor exactamente -entre otra mucha información- el conjunto de caracteres, los delimitadores y las características opcionales de SGML que se están utilizando (por ejemplo, la minimización es algo que unos sistemas soportan y otros no).
- que por consiguiente, el documento contendrá un **subconjunto de la declaración del tipo de documento**, un conjunto formal de declaraciones de elementos, atributos y entidades que le indican a un sistema exactamente el tipo de etiquetado que se utiliza en dicho documento. A menudo se sustituirá la DTD completa por una línea que indique que la DTD se edita como un texto público o se encuentra ya disponible en el sistema receptor.
- que en consecuencia, el documento contendrá una **muestra de documento**. Dicha muestra es lo que se entendía por el documento en sí: el contenido junto con el etiquetado.

Gracias a SGML se ha conseguido simplificar todos los pasos. En este sistema cada componente establece los valores y parámetros para el siguiente componente. El único etiquetado que aparece ha sido declarado en la DTD, y la sintaxis de la DTD se ha indicado mediante la declaración SGML definida por el estándar.

La auténtica ventaja de esta secuenciación de indicaciones es que los ordenadores pueden seguirla para comprobar si los documentos se adaptan a las reglas establecidas para ellos mismos. SGML, a pesar de ser un lenguaje incomprensible para los seres

humanos, es un lenguaje informático muy preciso. Esto significa que un programa informático **-un parser-** puede leer la declaración SGML y aprender sus reglas, a continuación leer la DTD y aprender las reglas del etiquetado y finalmente determinar si la muestra de documento cumple dichas reglas.

SGML tiene su mayor incidencia en el sector de la publicación electrónica. Andy van Dan, pionero en el desarrollo de hipertexto, declaraba en 1994:

"Si prefiero SGML como formato para intercambiar documentos frente a estándares industriales como PostScript o RTF es porque estos últimos únicamente aportan información del formato y no de la estructura de un documento, que podría llegar a visualizarse de distinta manera en medios diferentes.

El atractivo de SGML es que permite disponer de la información que se precisa para describir el documento no como una colección de páginas ya maquetadas, sino como una **estructura jerárquica de elementos multimedia.**"

Como metodología de edición de textos, SGML introduce conceptos de trabajo muy avanzados. Lo más novedoso es la idea de reemplazar los códigos y las macros procedimentales por **códigos declarativos** que separan el contenido (el valor funcional de los elementos de un documento) de su formato.

Así, frente a una etiqueta procedimental que señala un bloque de texto como cursiva o subrayado, sin indicar el motivo del cambio de tipografía, una etiqueta declarativa indica la condición del bloque: si se trata de una cita bibliográfica, de un extranjerismo, de un tecnicismo, o de un fragmento que se desea enfatizar, por ejemplo.

Esta posibilidad de **separar el contenido del formato** permite jugar de manera muy flexible con la información. Su estructuración, actualización, selección, combinación o presentación podrán ofrecerse según más convenga en cada oportunidad.

Pero SGML, aunque conocido por la utilización generalizada de etiquetas descriptivas ('tags'), no constituye un conjunto predefinido de etiquetas, ni una sintaxis de *etiquetado* ('markup'). SGML es en realidad un metalenguaje que permite el diseño y control de un conjunto de etiquetas y de una sintaxis a la medida del usuario o de la aplicación. Su adaptación a las necesidades del uso se concreta en la *declaración de tipo de documento* (DTD).

HTML es un ejemplo de DTD; es decir, dentro de las múltiples posibilidades de SGML, la DTD de HTML describe un conjunto concreto de etiquetas de hipertexto y una sintaxis para utilizarlas en Internet.

2.1. Otras aplicaciones de SGML

Pero además de HTML para WWW, existen muchas otras aplicaciones de SGML de gran envergadura y trascendencia. Varios organismos y asociaciones profesionales han establecido directrices propias que se materializan en el desarrollo de DTDs especializadas.

Así sucede con las directrices de la **American Association of Publishers (AAP)** -asociación de editores americanos-, en las que se contempla el etiquetado para una extensísima categoría de documentos (muchas de sus propuestas han sido reconocidas por las asociaciones de estandarización ANSI/NISO e ISO).

Un caso similar es el desarrollo del **Formato de Texto Universal (UTF)**, promovido por el Consejo de Telecomunicaciones de la Prensa Internacional (IPTC) y la **Newspaper American Association (NAA)** -asociación de editores de prensa diaria americana- para el intercambio de noticias entre agencias y clientes, reemplazando los anteriores formatos **IPTC 7901** y **ANPA 1312**.

En el ámbito de la descripción archivística nos encontramos con los trabajos que se están desarrollando para, partiendo de la aplicación de la ISAD(G) [International Standard Archival Description (General)], codificarla posteriormente utilizando un lenguaje de marca como SGML, empleando un DTD que se ha denominado EAD (Encoded Archival Description), que permitiría su tratamiento electrónico.

2.2. El proyecto TEI

El proyecto del consorcio internacional para el etiquetado de textos TEI (*Text Encoding Initiative*) es una iniciativa que ha partido de diversas asociaciones profesionales en el campo de las humanidades. Se trata de definir unas líneas generales de actuación para la representación de materiales textuales en formato electrónico.

El objetivo de TEI es fomentar el uso de etiquetas rigurosas y productivas para cualquier clase de texto, aunque su aportación más directa se produce en el campo de los textos con valor cultural y científico. Estas recomendaciones se recogen en un compendio

conocido como TEI P3 o [TEI Guidelines for Electronic Text Encoding and Interchange](#) (TEI, 2000)

El proyecto cuenta con el apoyo de tres asociaciones profesionales: ACL (Asociación de Lingüística Computacional), ALLC (Asociación de Literatura y Lingüística Computacional) y ACH (Asociación de Computación y Humanidades).

Charles F. Goldfarb, uno de los máximos responsables de la gestación de SGML, declaraba recientemente respecto a TEI:

"Si alguna vez alguien me pidiera una receta segura para lograr el caos, yo le sugeriría un proyecto en el que varios miles de especialistas pertenecientes a distintas disciplinas y a una docena o más de países, todos ellos muy entusiastas, recibieran el encargo de producir en un plazo menor a cinco años un documento de unas 1.200 páginas en el que se indicaran, de manera rigurosa y verificable por medios automáticos, las formas de expresar el conocimiento en cada una de sus especialidades.

Parece sociológica y tecnológicamente imposible que un equipo así llegara ni siquiera a acordar las materias de discusión del documento y mucho menos los detalles de la codificación. Sin embargo, con la misma sencillez con la que un abejorro desafía las leyes de la aerodinámica, TEI ha llevado a cabo con éxito su misión.

Las directrices de TEI son extraordinarias. Aunque nunca se pusieran en práctica, constituirían una contribución sin precedentes por el grado de detalle alcanzado en el análisis de los elementos informativos en una tipología tan amplia de textos. Y el caso es que tanto la comunidad científica como el mundo editorial ya han comenzado a utilizar estas directrices."

2.2.1. Funcionamiento de TEI

La operatividad de TEI se ha establecido entorno a cuatro comités que comparten la responsabilidad de elaborar las directrices.

1. El Comité de Documentación de Textos se encarga de definir las etiquetas para identificar los textos (procedencia, ubicación, clase, categoría, etc.).

2. El Comité de Representación Textual se ocupa de describir física y lógicamente los textos. La descripción lógica abarca cuestiones como su estructura (capítulos, secciones, etc.), la tipografía, la maquetación, las notas, apéndices y referencias diversas, etc.
3. El Comité de Análisis e Interpretación de Textos trata el desarrollo de etiquetas que permiten la descripción lingüística y literaria del texto, así como cuestiones de intertextualidad, indización, etc.
4. El Comité de Cuestiones Metalingüísticas se ocupa de los problemas técnicos de la sintaxis utilizada en la etiquetación.

Las cuestiones dirimidas por estos comités de TEI dan idea del grado de complejidad y precisión al que se quiere llegar. La aplicación de estas directrices señala directamente a la idea que apuntábamos al comienzo, la de una transformación radical en la difusión y acceso al conocimiento.

3. XML

La popularización de Internet, y sobre todo el avance y desarrollo del World Wide Web, ha planteado nuevos retos para tratar de manejar adecuadamente la ingente cantidad de información. Se plantea además el problema de su falta de estructuración y la complejidad que va adquiriendo, que dificulta los procesos de gestión, mantenimiento y recuperación.

El XML (eXtensible Markup Language) nace en el año 1996 (ratificada por el W3C en 1998 la especificación XML 1.0), con el fin de permitir intercambiar documentos muy estructurados a través del WWW, permitiendo al mismo tiempo la utilización de la característica hipertexto.

XML es una versión abreviada de SGML, lo que permite una mayor agilidad en su implementación, pero hay que tener en cuenta que carece de muchas de las características del propio SGML y precisamente por ello XML no sustituye a SGML.

Los objetivos de XML (W3C-REC-XML) pueden ser los siguientes:

- Se ha creado para el Web.
- Sobre XML se puede desarrollar una gran cantidad de aplicaciones.
- XML debe ser compatible con SGML.

- Debe ser fácil la escritura de los programas que procesen los documentos XML.
- Por compatibilidad, el número de características opcionales debe tender a cero.
- Los documentos XML deben ser legibles por los usuarios y no desvirtuar el contenido.
- El diseño de XML debe ser muy rápido.
- Los documentos XML deben ser fáciles de crear.
- XML precisa de otros estándares asociados como, Unicode e ISO/IEC 10646 para el tratamiento de caracteres, Internet RFC 1766 para las marcas de identificación de lenguaje, ISO 639 para los códigos de nombre de lenguaje, ISO 3166 para los códigos de nombre de país.

XML describe una clase de objetos de datos llamados documentos XML y parcialmente describe el comportamiento de programas de ordenador que puedan procesarlos.

Los documentos XML están formados por unidades de almacenamiento llamadas entidades, que contienen datos procesados o sin procesar. Los datos procesados están formados por caracteres, algunos de los cuales forman datos de carácter, y otros por marcas. Las marcas codifican la descripción del esquema de almacenamiento y la estructura lógica del documento, pudiendo establecer mecanismos de restricción, tanto al esquema de almacenamiento como a la estructura lógica.

La potencia de XML proviene de la separación que ofrece entre el interface de usuario y la estructura de los datos., centrándose en la definición de los contenidos. Se separan los datos de la representación y del procesamiento, permitiendo mostrar y procesar los datos como se desee, dependiendo de las diferentes aplicaciones u hojas de estilo empleadas.

Esta separación de los datos de su representación permite una integración absoluta entre datos que provienen de diferentes fuentes, facilitando el intercambio de datos online fácilmente. Algunos beneficios (Microsoft, 2000) que aporta el empleo de XML pueden ser:

- Las búsquedas de información se ajustan mejor a las necesidades del usuario.
- Se pueden desarrollar aplicaciones Web muy flexibles que permitan:
 - o Integración de datos de diferentes fuentes.
 - o Manipulación y tratamiento de los datos por parte del cliente.

- Múltiples posibilidades de visualización de los datos.
- Actualización granular de los datos, que permite eliminar la necesidad de reenviar una estructura de datos entera, cuando se modifican unos pocos datos, y solamente los elementos modificados son los que es preciso enviar del servidor al cliente.

Un aspecto interesante del XML es que es un conjunto de tecnologías, basadas en estándares abiertos, que forman módulos opcionales, que amplían las posibilidades del mismo. Algunos de estos módulos son:

- Xlink, que describe como añadir hiperenlaces a un fichero XML.
- XPointer y Xfragments, que son sintaxis que indican la forma de apuntar a una parte, dentro de un documento XML.
- CSS, el lenguaje de hojas de estilo.
- XSL, lenguaje avanzado para expresar hojas de estilo, que se basa en XSLT.
- DOM, conjunto estándar de funciones para realizar llamadas, desde un lenguaje de programación, para manipular ficheros XML.
- XML Namespaces, especificación que describe como asociar un URL con cada etiqueta y atributo en un documento XML.

Progresivamente han ido apareciendo soluciones a diferentes problemas documentales en el Web, basadas en XML (Rosa, 1999a) como RDF (Resource Description Framework) para intercambio de metadatos (Méndez, 1999), XML-QL como lenguaje de interrogación de bases de datos, WIDL para desarrollo de interfaces web, EDI (Electronic Document Interchange) para intercambio electrónico de datos, etc.

XML ha permitido un mayor desarrollo al desprenderse de las complicaciones de SGML para poder estructurar la información y al no ser obligatorio disponer de DTD, permitiendo que muchas soluciones ahorren tiempo y simplifiquen el desarrollo del software de aplicación.

4. Algunas aplicaciones

- Hay un proyecto de trabajo para convertir el formato MARC a formato SGML y en un primer paso se está estableciendo el MARC DTD, que debe permitir convertir todos los registros en formato ISO 2709 (MARC) a ISO 8879 (SGML) y viceversa. En este

momento existe una versión MARC DTD de la Biblioteca del Congreso y algunas utilidades de conversión bidireccional MARC-SGML escritas en Perl (Cover, 1998).

- Muy interesante es la propuesta de norma EAD (Encoded Archival Description), para codificación de instrumentos de descripción archivísticos, para crear un formato de intercambio de la información archivística, que abre unas muy buenas perspectivas, sobre todo debido a la escasa aceptación del formato MARC AMC (Santamaría, 1999).
- Formato Dublin Core, creado por la iniciativa de OCLC (On Line Computer Library Center), que trata de “ubicar, en el entramado de Internet, los datos necesarios para describir, identificar, procesar, encontrar y recuperar un documento introducido en la red” (San Segundo, 1998).
- Proyecto BNC (British National Corpus) de creación de un corpus de Inglés moderno con 100 millones de palabras, para el que se ha creado el formato CDIF (Hábeas Document Interchange Format) basado en SGML.
- Proyecto MULTEXT (Multilingual Text Tools and Corpora), basado en TEI, para el tratamiento de bases de datos multilingües.
- Proyecto CES (Hábeas Encoding Standard), para realización de aplicaciones que procesen en lenguaje natural.
- Proyecto de biblioteca digital de la Universidad de Berkeley.
- Proyecto CIMI (Consortium for Interchange of Museum Information).
- Tesoros y proyectos terminológicos (Rosa, 1999).
- Proyecto sobre concordancia en paralelo de la Universidad de Nancy II.

Una lista completa de aplicaciones reales se encuentran en (Cover, 1999a) y (Cover, 1999b).

5. Conclusiones

La utilización del SGML y del XML abre una nueva perspectiva en el tratamiento y recuperación de la información y se puede convertir en los próximos años en una herramienta de trabajo fundamental en los diferentes entornos documentales.

Es deseable que todas las experiencias que surjan, traten siempre de ofrecer compatibilidad con los estándares en los que se basan, evitando propuestas fuera de un estándar establecido, con el fin de conseguir sistemas fiables y que puedan ser utilizados

por todos, posibilitando, en definitiva, el buen aprovechamiento de la información en todas sus facetas.

6. Bibliografía

- [ABAITUA, 2000] Abaitua, Joseba. *La telaraña* [documento www], 2000. Disponible en: <http://orion.deusto.es/~abaitua/konzeptu/byte.htm> [Consulta: Marzo de 2000]
- [ARBORTEXT, 1995] A guide to SGML and its role in information management [documento www], 1995. Disponible en: http://www.arbortext.com/Think_Tank/SGML_Resources/Getting_Started_with_SGML/getting_started_with_sgml.html [Consulta: Marzo de 2000]
- [BURNARD, 1995] Burnard, Lou. *Text encoding for Information Interchange* [documento www], 1995. Disponible en: <http://www.uic.edu/orgs/tei/info/teij31/> [Consulta: Marzo de 2000]
- [COVER, 1998] Cover, Robin. *MARC (Machine Readable Cataloging) and SGML/XML* [documento www], 1998. Disponible en: <http://www.oasis-open.org/cover/marc.html> [Consulta: Marzo de 2000]
- [COVER, 1999] Cover, Robin. *SGML: general Introductions and Overviews* [documento www], 1999. Disponible en: <http://www.oasis-open.org/cover/general.html> [Consulta: Marzo de 2000]
- [COVER, 1999a] Cover, Robin. *Academic Applications* [documento www], 1999. Disponible en: <http://www.oasis-open.org/cover/acadapps.html> [Consulta: Marzo de 2000]
- [COVER, 1999b] Cover, Robin. *General SGML/XML Applications* [documento www], 1999. Disponible en: <http://www.oasis-open.org/cover/gen-apps.html> [Consulta: Marzo de 2000]
- [GAYNOR, 1996] Gaynor, Edward. *From MARC to Markup: SGML and Online Library Systems* [documento www], 1996. Disponible en: http://www.lib.virginia.edu/speccol/scdc/articles/alcts_brief.html [Consulta: marzo de 2000]
- [GIMENO, 1998] Gimeno Montoro, María José; Barrueco Cruz, José Manuel; García Testal, Cristina. Catalogación de recursos electrónicos accesibles en internet: revisión de propuestas para una normativa. *Jornadas Españolas de Documentación* (6ª. 1998. Valencia). Disponible en: http://www.florida-uni.es/~fesabid98/Comunicaciones/mj_gimeno.htm

- [ISO 8879, 1986] International Organization for Standardization. (1986). *ISO 8879-1986 (E). Information Processing --- Text and Office Systems --- Standard Generalized Markup Language (SGML), first edition --- 1986-10-15*. Geneva: International Organization for Standardization.
- [MÉNDEZ, 1999] Méndez Rodríguez, Eva M^a. RDF: un modelo de metadatos flexible para las bibliotecas digitales del próximo milenio. *Jornades Catalanes de Documentació*. 1999, 7, 487-498. Disponible en: <http://www.cobdc.org/7es/1.pdf>
- [ROSA, 1999] Rosa, Antonio de la. Instrumentos terminológicos en el www: xml. *El profesional de la información*, 1999, octubre, v. 8, n. 10, pp. 14-36
- [MICROSOFT, 2000] Microsoft. *Benefiting from XML* [documento www], 2000. Disponible en: <http://msdn.microsoft.com/xml/general/benefits.asp> [Consulta: Marzo de 2000]
- [ROSA, 1999a] Rosa Piñero, Antonio de la. Entornos documentales WWW: entorno XML. 2^a Jornadas Andaluzas de Documentación (2^a. 1999. Granada), pp. 357-372
- [SAN SEGUNDO, 1998] San Segundo Manuel, Rosa. Organización del conocimiento en Internet: metadatos bibliotecarios Dublín Core. *Jornadas Españolas de Documentación* (6^a. 1998. Valencia). Disponible en: http://www.florida-uni.es/~fesabid98/Comunicaciones/r_sansegundo.htm
- [SANTAMARÍA, 1998] Santamaría González, Fernando. XML (eXtensible Markup Language): nuevo estándar para la descripción de documentos en la World Wide Web. *Jornadas Españolas de Documentación* (6^a. 1998. Valencia). Disponible en: http://www.florida-uni.es/~fesabid98/Comunicaciones/f_santamaria/f_santamaria.htm
- [SANTAMARÍA, 1999] Santamaría Gallo, Abelardo. Codificación de instrumentos de descripción archivísticos con SGML/XML: La norma EAD v.1.0. *Tabula*, 1999, n. 4, pp.69-125
- [TEI, 2000]TEI Guidelines for Electronic Text Encoding and Interchange [documento www] Disponible en: <http://etext.virginia.edu/TEI.html> [Consulta: Marzo de 2000]
- [W3C-CSX-NOTE] WORLD WIDE WEB CONSORTIUM. *Comparison of SGML and XML* [documento www], Nota de 15 de Diciembre de 1997, Disponible en: <http://www.w3.org/TR/NOTE-sgml-xml-971215> [Consulta: Marzo de 2000]

[W3C-REC-XML] WORLD WIDE WEB CONSORTIUM. *Extensible Markup Language (XML) 1.0* [documento www], Febrero 1998, Disponible en: <http://www.w3.org/TR/1998/REC-xml-19980210>

6.1. Direcciones Web relacionadas

<http://orion.deusto.es/~abaitua/konzeptu/>

<http://www.w3c.org/XML>

<http://msdn.microsoft.com/xml/general/intro.asp>

<http://www.w3.org/MarkUp/SGML/>

<http://lcweb.loc.gov/ead/>

<http://purl.org/DC>

<http://www.w3.org/RDF>