

METADATOS : estado de la cuestión

Tesis de la Licenciatura en Bibliotecología y Documentación

María Carolina Rojas
Directora: Lic. Silvia Sleimen



FACULTAD DE HUMANIDADES
UNIVERSIDAD NACIONAL DE MAR DEL PLATA

2010

PÁGINA DE ACEPTACIÓN

METADATOS: Estado de la cuestión

Tesis presentada

por

BIBL. DOC. M. CAROLINA ROJAS

Aprobada en contenido y estilo por:

Lic. Silvia R. Sleimen, Directora

César O. Archuby, miembro del jurado

Gustavo Liberatore, miembro del jurado

Licenciatura en Bibliotecología y Documentación

Dpto. de Documentación / F.H. / UNMDP

Servicio de Información Documenta
Dra. Liliana B. De Boschi
Facultad de Humanidades
U.N.M.D.P.

Agradecimientos

A Eva Méndez Rodríguez, por su colaboración especializada, desinteresada y continua.

A Víctor Herrero, por sugerirme el tema de investigación.

A mis compañeros de trabajo Silvia Sleimen, Mela Bosch, Marcela Coringrato y Andrés Vuotto por sus continuos aportes.

A Rosalía Baltar, que revisó, muy generosamente, la escritura y organización de este documento.

A todos los que a lo largo de estos años no dejaron de apoyarme y me acompañaron en este proceso y me alentaron aplicar los conocimientos adquiridos, en especial a Noemí Conforti, César Archuby, Gustavo Liberatore, María Coira, Luis Porta y tantos otros de los que me estaré olvidando.

A mis amigos, a mi familia y a Claudio que nunca dejan de apoyarme y acompañarme en lo que necesito.

RESUMEN:

Los metadatos implican una nueva forma de estructurar los objetos de información digitales, como alternativa tecnológica para mejorar u optimizar su recuperación en contextos de red, fundamentalmente en *Internet* y en la *World Wide Web*, que han generado en los últimos años un debate profesional y un desarrollo vertiginoso, lo que nos ha llevado a elegirlos como objeto de estudio principal de esta investigación.

Se realiza una revisión bibliográfica acerca de los metadatos, sus tipos, funciones, esquemas y nivel de utilización, para mejorar la recuperación de recursos de información digital. Con ese fin, y para ejemplificar su aplicabilidad, se presenta un prototipo de uso de estas estructuras para la descripción y búsqueda de información en el sitio web del Departamento de Documentación de la Facultad de Humanidades de la Universidad Nacional de Mar del Plata (Argentina).

DESCRIPTORES:

Internet (World Wide Web) / Metadatos / Metainformación / Información digital / Documentación digital / Procesamiento de la información / Recuperación de información / Estructura de contenidos Web / Lenguajes de marcado / Sitios Web / LICAD / Motores de búsqueda / Buscadores locales

RESPONSABLES DEL PROYECTO

Grupo ejecutor: Bibl. Doc. M. Carolina Rojas

Directora: Lic. Silvia R. Sleimen

Entidad responsable: Dpto. de Documentación / F.H. / UNMDP

Tabla de Contenido

1. Introducción	3
2. Marco teórico, planteo, objetivo y método de la investigación.....	5
2.1 Ubicación del Proyecto	5
2.2. Probable impacto de la investigación	6
2.3. Objetivos de la investigación	7
Objetivo general:	7
Objetivos específicos:.....	7
2.4. Formulación del problema	8
2.5. Antecedentes, estado actual de la cuestión y marco teórico.....	9
2.6 Diseño metodológico	15
2.7 Administración del proyecto.....	18
Cronograma de actividades:.....	18
Recursos disponibles	19
Duración	19
Aporte e impacto de los resultados	19
3. Los metadatos: definición, tipos y funciones	20
Contexto para la gestión de metadatos (Greenberg)	25
4. Esquemas de metadatos	33
4.1. La elección de un esquema de metadatos	35
4.1.1. Consideraciones generales	35
4.1.2. Propuestas y modelos de metadatos	36
4.1.3. Software para creación, implementación y gestión de metadatos	38
4.1.4. Etapas del proceso de implementación de metadatos.....	40
4.1.5. La elección de un formato de metainformación para el sitio Web de la LICAD	43
5. El Modelo DC (ISO Standard 15836-2003 - ANSI/NISO Z39.85-2007 - IETF RFC 5013-2007).....	44
6. Aplicación de metadatos al sitio Web de la LICAD.....	61
7. Conclusiones y reflexiones finales.....	66
8. ANEXOS	68
A. Código en JavaScript del buscador	69
B. Arrays de búsqueda para el Módulo de Procesamiento de la información.....	81
9. Glosario.....	86
10. Bibliografía.....	91

Índice de ilustraciones

Ilustración 1 - Contexto para la gestión de metadata. Trad. por la autora, tomado del artículo de Jane Greenberg	25
Ilustración 2 - DataWareHouse – Asociación de metadatos embebidos.....	29
Ilustración 3- DataWareHouse – Asociación de metadatos along-with.....	29
Ilustración 4 - DataWareHouse – Asociación de metadatos service bureau.....	30
Ilustración 5 - DataWareHouse – Asociación de metadatos wrapped.....	30
Ilustración 6 - Relación de estándares de metadatos.....	37
Ilustración 7 - Modelo de etiqueta cualificada de Metadatos según formato DCMI	46
Ilustración 8 - Cabecera HTML en DC.....	47
Ilustración 9 - Cabecera en RDF un archivo HTML con metadatos DC.....	60
Ilustración 10 - Metabot Professional. Generación de un proyecto de edición de etiquetas – selección de esquema (en este caso “Basic Tags” y “Dublin Core”).....	62
Ilustración 11 - Metabot Professional. Generación de un proyecto de edición de etiquetas – selección de directorio.....	62
Ilustración 12 - Metabot Professional. Generación de un proyecto de edición de etiquetas - conjunto de documentos sin editar	63
Ilustración 13 - Metabot Professional. Generación de un proyecto de edición de etiquetas - conjunto de documentos editados	63
Ilustración 14 - Metabot Professional. “HEAD” de uno de los documentos sin procesar	63
Ilustración 15 - Metabot Professional. “HEAD” de uno de los documentos procesados	63
Ilustración 16 - Código buscador en JavaScript	64
Ilustración 17 - Arrays de búsqueda, con los metadatos extraídos de los archivos componentes del sistema	64
Ilustración 18 - Visualización del buscador en un navegador.....	65
Ilustración 19 - Visualización de los resultados de búsqueda	65

Índice de tablas

Tabla 1 - Tipos de metadatos	24
Tabla 2- Concepto de Generación Automática de Metadatos.....	27

1. Introducción

El desarrollo de las tecnologías implica en forma permanente y constante nuevos desafíos a la hora de relacionarse con las diferentes disciplinas. En el caso de las ciencias de la información, con el advenimiento de Internet, se plantean cuestiones en torno a la organización, gestión y recuperación efectiva de la información electrónica.

Entendemos que el caos de información que se ha suscitado con la fácil aplicabilidad de los lenguajes de edición de páginas web, sobre todo con la capacidad de edición *WYSIWYG* (*What You See Is What You Get*, lo que ves es lo que obtienes) -aplicación que permite ver el documento en su presentación final al momento de su procesamiento/creación; posibilita que cualquier usuario de la Red produzca y publique contenidos, sin que estos necesariamente cumplan con normas bien definidas-, impide una eficaz localización de recursos.

Para subsanar dicho problema se recurrió a la utilización formal de herramientas de descripción de propiedades previstas por quienes desarrollaron los formatos de archivos informáticos.

Nos ocupamos, desde el punto de vista documental, de la descripción de recursos electrónicos, mediante la utilización de esquemas de metadatos¹. Se

1 Metadatos: información acerca de la información; son datos sobre las características de un recurso de información, describen sus relaciones, y permiten su descubrimiento (recuperación) y uso efectivo. (Burnett, 1997)

tiene presente en todo momento que Internet funciona como una gran *base de datos*, ya que se sustenta en una serie de protocolos que permiten compartir datos almacenados en diferentes ordenadores y con diferentes plataformas sin que esto sea un inconveniente en el sistema de comunicación; esto obliga a que la información disponible tenga estructuras de datos definidas, en su nivel más simple de representación, y en especial si prestamos atención al código fuente de las páginas web.

Las funciones de descripción, tanto física (forma) cuanto lógica (representación de contenidos) que tienen y cumplen los metadatos facilitan la tarea de los indizadores automáticos (programas que representan el contenido conceptual de los documentos con vistas a su recuperación), además de tener un universo controlado de documentos electrónicos.

La exhaustividad que la aplicación de esta tecnología confiere a los sitios web permite el desarrollo de la *web semántica*, y a su vez, nos acerca a la aproximación de una *sociedad de la información por y para todos*.

A través de un proceso de investigación exploraremos diversas fuentes de información, impresas y electrónicas, para definir qué son los metadatos, sus características, funciones, aplicación y software. Y para completar este estudio se presenta un caso práctico de aplicación de metadatos en un sitio web educativo, el de la Licenciatura en Bibliotecología y Documentación A Distancia (LICAD) del Departamento de Documentación de la Facultad de Humanidades de la Universidad Nacional de Mar del Plata.

2. Marco teórico, planteo, objetivo y método de la investigación

En este capítulo introductorio, fundamental en un trabajo de esta naturaleza, se pretende, por una parte justificar la elección del tema, el marco teórico en el que se ubica, los antecedentes y presupuestos de los que partimos para abordarlo y los límites que establecemos en la investigación; y por otra, se definen los objetivos (generales y específicos) que se intenta alcanzar, a partir del análisis de variables concretas.

2.1 Ubicación del Proyecto

- a) Área de investigación: Procesamiento de la información.
- b) Materia de la investigación: Automatización, descripción y recuperación de la información
 - Temas generales: Información y Documentación digital; Internet (World Wide Web); Descripción de la información (catalogación, clasificación, indización); Recuperación de la información; Estructuración de la información Web; Marcado de documentos (HTML, XML, SGML).
 - Tema específico: Metadatos.
- c) Cobertura del proyecto: Revisión bibliográfica del concepto de metadatos tanto en fuentes impresas cuanto electrónicas, su construcción, desarrollo, aplicaciones y prospectiva de utilización en el contexto de la información electrónica en Internet, desde 1998 a 2005. Además, se presenta un ejemplo de aplicación que permita la utilización de un esquema básico de

metainformación en el sitio web de LICAD (Licenciatura en Bibliotecología y Documentación A Distancia, Departamento de Documentación, Facultad de Humanidades, Universidad Nacional de Mar del Plata), de tal forma que facilite una recuperación de información más precisa y pertinente.

- d) Campo de interés: Dirigida a los profesionales de la información y en especial a los que se dedican a la descripción y recuperación de información en Internet, así como a los encargados de la creación de sitios y recursos web. Adscribimos con este trabajo a la nueva tendencia de la organización de la información digital basada en la información semi-estructurada y en el mercado de documentos.
- e) Tipo de investigación: Estudio exploratorio-descriptivo, diseño transeccional descriptivo.

2.2. Probable impacto de la investigación

Esta investigación responde a un requisito académico de la Licenciatura en Documentación del Dpto. de Documentación / F.H. / UNMDP, y pretende acercar a los bibliotecarios y gestores de información una nueva tendencia del trabajo profesional en el ámbito de la información digital, abordando en ella un tema de especial interés en un contexto como es la metainformación.

Para ejemplificar este trabajo, el dominio informativo elegido es el Web de LICAD porque, por un lado, es el ámbito de información electrónica en cuyo desarrollo la autora está implicada por pertenecer al equipo técnico del Departamento de Documentación de la Universidad Nacional de Mar del Plata, y además, porque resulta un dominio informativo concreto y controlable para los

objetivos de esta investigación, en la que se trata de esclarecer qué son los metadatos y difundir su utilidad.

2.3. Objetivos de la investigación

Objetivo general:

Examinar de manera detallada el estado de la cuestión de los metadatos, a partir de la revisión de las principales fuentes en la materia, así como ejemplificar, a través de una experiencia específica de aplicación, su utilidad para la recuperación de información.

Objetivos específicos:

- ✓ Determinar las ventajas de describir y recuperar la información electrónica en la World Wide Web, mediante la aplicación de metainformación.
- ✓ Describir los metadatos, sus características, funciones, y modos de aplicación.
- ✓ Caracterizar los programas disponibles para su implementación en el ámbito de la documentación.
- ✓ Describir la presentación de un caso de aplicación de metadatos en un sitio web educativo.

2.4. Formulación del problema

Ante el aumento exponencial de recursos Web, es preciso tomar conciencia —no sólo desde una visión informática, sino también bibliotecológica—, de las nuevas tendencias, tecnologías y aplicaciones alternativas a la imposibilidad de procesamiento convencional de la información electrónica.

Este problema generalizado en el ámbito de la recuperación de información de Internet, se puede concretar más en el marco de las bibliotecas digitales y de cualquier servicio de información electrónica, y en nuestro caso, circunscribimos la cuestión a definir qué son los metadatos y cuál es el estado de la cuestión, realizando una revisión bibliográfica y elaborando un trabajo monográfico acerca de las diferentes facetas necesarias para su definición teórica. Agregaremos, a modo de ejemplo, una propuesta de aplicación para la mejora del sistema de descripción-recuperación de contenidos en el sitio Web de LICAD, en el que garantizar el acceso coherente a la información es una obligación por parte de nosotros y un derecho de los alumnos.

Problema: ¿Qué son los metadatos, sus características, funciones, modos de aplicación y de qué software disponemos para su implementación, en el ámbito de la bibliotecología y documentación?

2.5. Antecedentes, estado actual de la cuestión y marco teórico

La red conocida como Internet tiene más de 30 años de existencia, pero fue a principios de los años 90 cuando se establecieron las bases de lo que hoy conocemos como *World Wide Web*, y el momento a partir del cual se desarrolla exponencialmente como medio informativo. Se fundamentó en dos pilares: un protocolo de transferencia de hipertexto (HTTP, *HiperText Transfer Protocol*), y un sencillo lenguaje de marcado (*HTML, HyperText Markup Language*) interpretable por intuitivas aplicaciones de software o navegadores (*browsers*), que han hecho de la Web un entorno documental universal. El crecimiento que ha experimentado la Web en la última década constituye la base del problema que pretenden solucionar los metadatos y por ende, el fundamento del problema de esta investigación.

Cuando Tim Berners-Lee estableció el *HyperText Markup Language* (HTML) tenía como principal objetivo garantizar el funcionamiento de la red desde un punto de vista informático y con este lenguaje de marcado logró, además de un soporte *multi-plataforma*, un modo intuitivo de hacer accesible y manejable la información. El modelo hipertextual simple era más que suficiente para organizar la información interna del CERN (*Centre Europeen de la Recherche Nucleaire-European Council for Nuclear Research*), lugar en el que surge originalmente la

WWW, pero cuando éste fue llevado a escala mundial o global comenzaron a detectarse rápidamente sus limitaciones².

En un primer momento se consideró posible la recuperación de la información a través de *directorios* que, contruidos manualmente, organizaran la información en función de su naturaleza temática. Los casos paradigmáticos de este tipo de servicios de búsqueda de información son, por diversos motivos³:

-Yahoo! <<http://www.yahoo.com>> en el que puede encontrarse una estructura jerárquicamente ordenada de 2.500.000 enlaces⁴ aproximadamente.

-OPD (*Open Directory Project*) <<http://dmoz.org>>, que registra en torno a 4.000.000 de recursos enlazados⁵.

Ciertamente, estas cifras pueden parecer bastante grandes; sin embargo, sólo corresponden a una ínfima parte de la totalidad de la Red, por lo que este tipo de soluciones, si bien útiles, son al mismo tiempo insuficientes. Una forma alternativa de subsanar la limitación de la lentitud de la selección y procesamiento

2 Berners-Lee, Tim. Tejiendo la red: el inventor del World Wide Web nos descubre su origen. Madrid: Siglo XXI, 2000.

3 En el caso de Yahoo! el motivo fundamental para considerarlo como paradigma de este tipo de servicios es fundamentalmente su antigüedad (1994), que le otorga la capacidad de establecer un modelo. En el caso de OPD, su cobertura y la forma cooperativa y abierta de su realización, desde que Tim Bray propusiera el modelo a finales de 1998.

4 Según los últimos datos recogidos en *Search Engine Watch* (diciembre de 2004) <http://www.searchenginewatch.com/reports/> (2.600.000 OPD / 1.800.000 Yahoo!)

ODP and Yahoo Size Projection Charts (consultado diciembre de 2004 / creado 1diciembre 1999 / actualizado 10 enero 2004) <http://www.geniac.net/odp/> (4.000.000 OPD / 2.500.000 Yahoo!)

5 *Ibid.*

manual, en que se fundan los índices de búsqueda se encuentran en el procesamiento (indización) automático.

En esta línea se desarrollaron los motores o robots de búsqueda (*SearchEngine*), aplicaciones que construyen sus bases de datos a través de sistemas totalmente automatizados, conocidas con diversas denominaciones (crawlers, robots, agentes, rastreadores, engines, worms, ants, etc.), cuya principal finalidad consiste en recorrer la Web y extraer de ella la información. Esta tarea la realizan a través de cuatro elementos básicos:

-El robot, formado por un localizador y recolector. El localizador se encarga de revisar las direcciones de servidor almacenadas en su base de datos en función de recuperar y almacenar nuevos enlaces en la base de datos del recolector, cuya función será evaluar esas direcciones en términos de su preexistencia y en tal caso de posibles modificaciones, por medio del análisis del tamaño y fecha del recurso. Si el recurso es novedoso o ha sufrido cambios se los envía al analizador-indexador.

-El sistema de indexación automática o analizador-indexador es el programa encargado de extraer palabras del contenido textual de acuerdo con un conjunto de filtros. Los filtros son los que deciden si se debe indexar toda la página o sólo las primeras palabras, si indexará todas las etiquetas de un documento HTML o sólo el texto contenido en las etiquetas META, TITLE, Hn. Estos datos se

almacenan en un fichero inverso que será gestionado por una base de datos sobre la que buscarán los usuarios.

-El motor de recuperación: es importante considerar este ítem desde dos puntos de vista, el que tiene que ver con el hardware, mayormente en el aspecto de velocidad y capacidad de procesamiento, y el que se relaciona con el software, en cuanto a las capacidades de recuperación de información, vinculado con la estructura de las bases de datos, su forma de almacenamiento, lenguajes de recuperación admitidos (utilización de palabras simples, referencias compuestas, operadores booleanos, operadores relacionales, operadores de adyacencia, búsqueda cualificada o por campos, posibilidad de truncamiento, distinción entre mayúsculas y minúsculas, fichero de palabras vacías, combinación de búsquedas, sistemas de retroalimentación de relevancias, búsquedas por peso).

-La interfaz de usuario: es el subsistema que permite que un usuario pueda realizar una consulta y visualizar sus resultados. En general, los motores de búsqueda presentan un formulario simple, con un cuadro de texto en el que el usuario puede insertar su fórmula de búsqueda, con algunas características distintivas de delimitación de idioma, tipo de documento, fecha de publicación, según el perfil del buscador, y un botón que posibilita la ejecución de la consulta. Presentan, así, los resultados de acuerdo con un algoritmo de relevancia, aspecto que tiene especial importancia puesto que el

orden en que son presentados es determinante de la satisfacción de la demanda del usuario.⁶

Abundan ejemplos en esta línea: Altavista, Northen Light, Google, Excite, InfoSeek, HotBot, AllTheWeb, Lycos, y muchos más.

Aún así, las bases de datos que se conforman no abarcan la totalidad de la red -no es lo mismo usar uno u otro buscador- y jerarquizan los resultados de acuerdo con criterios que muchas veces responden más a los lugares de aparición de los términos buscados (como en títulos o encabezados del documento), o, incluso, a fines comerciales, que a criterios de relevancia de información.⁷

Ante esta situación, en diferentes ámbitos académicos se comenzó a estudiar la posibilidad de desarrollar y aplicar algún tipo de norma que permitiera la descripción, y posterior recuperación, de los contenidos informativos. En esta línea encontramos dos grandes temas: la estructuración del documento completo (XML) y, por otro lado, su descripción catalográfica, en un sentido amplio. La primera, se encuentra dominada por el estándar de descripción XML mientras que en la segunda hallamos los metadatos. Si bien en torno al XML se están realizando muchos estudios, nos centraremos en la segunda de las corrientes y profundizaremos en el análisis de los METADATOS.

6 Hernández Pérez, Antonio. La búsqueda y recuperación de información en internet. EN: La sociedad de la información: política, tecnología e industria de los contenidos. Mercedes Caridad Sebastián (Coord.). Madrid: Fundación Ramón Areces, 1999. p. 213-238.

7 Siri, Laura. Internet: búsquedas y buscadores. Buenos Aires: Norma, 2000.

Aunque los términos METADATA y METADATOS no significan exactamente lo mismo para todos los colectivos profesionales (informáticos, bibliotecarios, gestores de BD, etc.) en el entorno de las bibliotecas digitales se usan cada vez más para referirse a la disponibilidad de datos sobre los recursos de información. Los términos hacen referencia a cualquier elemento utilizado para identificar, describir y localizar los recursos electrónicos en un entorno de red. Los metadatos son, en otras palabras, datos acerca de datos. En el mundo bibliotecario este concepto es bastante familiar ya que los registros bibliográficos son esencialmente metadatos; proporcionan información descriptiva y analítica sobre un objeto de información. Los catalogadores los han empleado como método descriptivo desde hace décadas, especialmente como registros MARC⁸.

Las áreas que definen el mundo de la investigación en metadatos abarcan su definición teórica, la generación de estándares internacionales, la creación de metadatos para nuevos recursos electrónicos, el desarrollo de nuevos tipos de metadatos, la evaluación de su efectividad, la elaboración de lenguajes y herramientas de aplicación y los métodos de gestión e *interoperabilidad* entre diferentes formatos.

No se pretende aquí abordar toda la complejidad que alberga esta problemática sino presentar en esta investigación el resultado de la revisión bibliográfica, tanto en fuentes impresas cuanto electrónicas, acerca del concepto

8 Gilliland-Swetland, Anne J. Setting the Stage. EN: Introduction to Metadata : Pathways to Digital Information. Murtha Baca. Los Ángeles: Getty Research Institute, mayo 2000. [Versión HTML] <http://www.getty.edu/research/institute/standards/intrometadata/pdf/swetland.pdf>

de metadatos, su construcción, desarrollo, aplicaciones y prospectiva de utilización en el contexto de la información electrónica en Internet. Un ejemplo que permita la utilización de un esquema básico de metainformación es su aplicación al sitio web de LICAD, ya que facilita una recuperación de información más precisa y coherente. Se trata pues, de recortar como objeto de estudio —a través de una metodología teórico-práctica— la discusión científica que, desde hace varios años gira en torno a la recuperación de información digital, instando a la utilización de metadatos así como a la investigación en la materia.

2.6 Diseño metodológico

Tipo de estudio: exploratorio-descriptivo

Diseño de investigación: transeccional

Instrumentos de recolección: fichaje de documentos recuperados en bibliotecas y bases de datos on-line.

Universo de estudio: artículos de publicaciones periódicas, libros y documentos electrónicos, publicados entre los años 1998 y 2005.

Variable: Metadatos

Definición conceptual:

-Metadatos: información acerca de la información; son datos sobre las características de un recurso de información, describen sus relaciones, y permiten su descubrimiento (recuperación) y uso efectivo. (Burnett, 1997)

Definición operacional:

-Estructura: sintaxis, semántica.

-Características: extensibilidad, repetibilidad, opcionalidad, modificabilidad.

-Tipos y funciones: administrativos, descriptivos, de conservación, técnicos, legales, de uso o acceso; y para cada uno de ellos: procedencia (internos - externos), métodos de creación (automáticos generados por ordenador - manuales creados por individuos), carácter (creados por individuos que no son especialistas temáticos ni en información - creados por expertos), estatus (estáticos - dinámicos-de larga duración - de corta duración), estructura (estructurados - no estructurados), semántica (controlados - no controlados), nivel (de colección - individuales).

-Modos de aplicación:

- Embebido/Encajado (*Embedded*): aparecen contenidos en el objeto de información digital (EJ: TEI / sintaxis << meta >> HTML: DC)
- Junto con/A lo largo del (*Along-with*): son externos al recurso pero se proporcionarán con él mediante un mecanismo de transferencia de la respuesta a una recuperación de información.
- Oficina de servicios (*Service bureau*) [departamento de servicios: cuerpo que proporciona información computada al público y puede procesar datos]: se almacenan separados del recurso de información, por ejemplo, en bases de datos y permiten su acceso. (EJ. MARC/UNIMARC).
- Envuelto (*Wrapped*): recurso de información encapsulado en los metadatos que lo describen. EAD / URI / DOI⁹.
- Métodos de implementación: manual, automático
- Software de implementación: denominación; institución proveedora; dirección de descarga; versión; modo de distribución (freeware, shareware, de pago); generador de metaetiquetas-generador de metadatos; tipo (plantillas, software cliente, SIG-DLO [Sistema Integrado de Gestión de Objetos Documentales]); interfaz (amigabilidad, sencillez); disponibilidad de ayuda (manejo de la aplicación, descripción de esquemas utilizados, etiquetas que

9 Lupovici, Catherine. Identification des ressources sur Internet et métadonnées: diversité des standards. EN Documentaliste – Sciences de l'information, 1999, vol. 36, N°6, p. 221-325.

soporta); forma de visualización del DLO que se está describiendo (navegador propio, navegador integrado, sin visualización); permiso de visualización del código fuente a través de un editor de textos; flexibilidad (permiso de repetir etiquetas, permiso de agregación de etiquetas ad hoc, permiso para generar nuevos formatos); formatos de codificación de metadatos (los distintos esquemas); integración al sistema de almacenamiento, búsqueda y recuperación de metainformación; control de autoridades (autores y materias); validación de los elementos obligatorios del modelo elegido; inclusión de aspectos de sintaxis (elementos, subelementos, calificadores).

2.7 Administración del proyecto

Cronograma de actividades:

1. Recopilación bibliográfica: publicaciones periódicas, monografías, manuales, páginas web. (3 meses)
2. Definición marco teórico. (3 meses)
3. Elaboración de un ejemplo que permita la aplicación de un esquema básico de metadatos al sitio web de la Licenciatura en Documentación a Distancia. Departamento de Documentación. Facultad de Humanidades. UNMDP. (2 meses)
4. Análisis de los resultados. (2 meses)
5. Elaboración de conclusiones y propuestas. (2 meses)

Recursos disponibles

PC con conexión a Internet, publicaciones periódicas especializadas, manuales de ciencias de la documentación e informática, software específico.

Duración

1 año

Aporte e impacto de los resultados

Como investigación básica: se realizará una revisión exhaustiva de los aspectos teóricos concernientes a los METADATOS dentro del ámbito de las ciencias de la información.

3. Los metadatos: definición, tipos y funciones

"Información acerca de la información" es la definición por excelencia que encontramos en todo documento acerca de metadatos, pero consideramos que con esto solo no basta para indicar el alcance del término.

Entendemos, y ampliando la anterior definición, que los metadatos están constituidos por aquellos elementos que permiten, de forma estructurada, describir las características de un objeto de información dado, para darle significado, ubicarlo dentro de un contexto y facilitar su organización, legibles por computadora.

A partir de este concepto podemos advertir una amplia correspondencia en el mundo de la información digital con las tareas de catalogación y clasificación, y su consiguiente normalización, desempeñadas tradicionalmente por los profesionales bibliotecarios en la organización de los documentos impresos, y es en este punto en el que se hace imprescindible la vinculación con los grupos que se ocupan del desarrollo de estándares internacionales de descripción de objetos digitales.

La estructura de los archivos digitales posibilitan que a cada objeto digital se le pueda asignar elementos de descripción según su formato: si se trata de un fichero generado a partir de una aplicación propietaria, el mismo software provee una herramienta de edición de las propiedades del documento: *Autor, Título,*

Palabras clave, Resumen, Fecha...; en el caso de los documentos generados a partir de lenguajes de marcado, se determinan etiquetas que permiten una descripción general del objeto digital (en el HEAD de un documento HTML mediante etiquetas META) o descripción de cada una de sus partes constitutivas.

La asignación de metadatos a un objeto o recurso de información digital se realiza a partir de la declaración de una propiedad, su valor y la referencia a un perfil normativo en el que se define cada propiedad y sus valores admitidos, y la sintaxis de una sentencia.

- Recursos: Todos los objetos descritos por expresiones RDF se denominan recursos. Un recurso puede ser una página Web completa, tal como el documento HTML <http://www.w3.org/Overview.html>, por ejemplo. Un recurso puede ser una parte de una página Web (un elemento HTML o XML específico dentro del documento fuente). Puede ser también una colección completa de páginas; es el caso de un sitio Web completo. También un objeto que no sea directamente accesible vía Web, p. ej. un libro impreso. Los recursos se designan siempre por URIs más identificadores de anclas opcionales (ver [URI]). Cualquier cosa puede tener un URI; la extensibilidad de URIs

permite la introducción de identificadores para cualquier entidad imaginable.

- **Propiedades:** Una propiedad es un aspecto específico, característica, atributo, o relación utilizado para describir un recurso. Cada propiedad tiene un significado específico, define sus valores permitidos, los tipos de recursos que puede describir, y sus relaciones con otras propiedades. Este documento no dirige cómo se expresan las características de las propiedades; para tal información, es posible dirigirse a la especificación del esquema RDF [RDF Schema specification].
- **Sentencias [declaraciones, enunciados]:** Un recurso específico junto con una propiedad denominada más el valor de tal propiedad para ese recurso es una sentencia RDF [RDF statement]. Estas tres partes individuales de una sentencia se denominan, respectivamente, sujeto, predicado y objeto. El objeto de una sentencia (es decir, el valor de la propiedad) puede ser otro recurso o puede ser un literal; es decir, un recurso

(especificado por un URI) o una cadena simple de caracteres [string] u otros tipos de datos primitivos definidos por XML. En términos RDF, un literal puede comprender en su contenido marcado XML pero ya no puede valorarse más por un procesador RDF. Existen varias restricciones sintácticas respecto al modo en que se puede expresar el marcado en literales.¹⁰

Desde una visión informática, Diché plantea una tipología de metadatos: a) *Comerciales o de usuario final*, que proveen definiciones comerciales de atributos específicos de los datos; en esta clase incluye códigos de designación de tipos de industria, descripciones de sistemas de origen de datos, sinonimia de datos, restricciones de acceso a los datos, valores válidos y reglas de la empresa para documentar datos derivados; b) *De base de datos*, que definen los términos usados en la base de datos: bases de datos, tablas y columnas. Y, por último, c) *De aplicación*, que explican ciertos términos y funciones contenidos en una herramienta de aplicación entregada al usuario final, como, por ejemplo, las ayudas para su uso.¹¹

10 Resource Description Framework(RDF) : Especificación del Modelo y la Sintaxis Recomendación el W3C 22 febrero 1999 (español). Editores: Lassila, Ora; Swick, Ralph R. Traducción de Eva Méndez Rodríguez. En: <http://www.sidar.org/recur/desdi/traduc/es/rdf/rdfesp.htm>

11 Diché, Hill. E-Data: transformando datos en información con Data warehousing – 1ª. Ed. – Buenos Aires : Prentice Hall, 2001. Pág. 182-184.

En los artículos acerca de la gestión documental de información podemos encontrar clasificados los metadatos en tres grandes grupos, a saber: descriptivos, estructurales y administrativos. Los primeros, se ocupan de la descripción e identificación de recursos de información. Los estructurales, facilitan la navegación y presentación de los recursos electrónicos. Y finalmente, los administrativos, que se encargan de la gestión y procesamiento de las colecciones digitales.¹²

Gilliland-Swetland¹³ amplía esta última clasificación:

<i>Tipo</i>	<i>Función</i>	<i>Elementos</i>
Administrativo	<i>Metadatos usados en la gestión y administración de recursos</i>	-Adquisición de información -Control de derechos y reproducciones -Documentación de requisitos legales -Información sobre localización -Criterios de selección para la digitalización -Control de las distintas versiones
Descriptivo	<i>Metadatos usados para describir o identificar información sobre recursos</i>	-Catalogación de registros -Ayudas para la búsqueda -Índices especializados -Relaciones hiperenlazadas entre recursos -Anotaciones de usuarios
Conservación	<i>Metadatos relacionados con la conservación de recursos de información</i>	-Documentación de recursos sobre la condición física -Documentación de acciones tomadas para conservar versiones físicas y digitales de los recursos, ej, regeneración y migración de datos
Técnico	<i>Metadatos relacionados con el funcionamiento de los sistemas o el comportamiento de los metadatos</i>	-Documentación de hardware o software -Digitalización de la información, p. Ej., formatos, ratios de compresión, ajustes -Control del tiempo de respuesta de los sistemas -Autenticación y seguridad de los datos, p.ej., claves cifradas, contraseñas
Uso	<i>Metadatos relacionados con el nivel y el tipo de uso de los recursos de información</i>	-Registro de exhibiciones -Seguimiento de usos y de usuarios -Uso repetido del contenido e información sobre versiones múltiples

Tabla 1 - Tipos de metadatos

12 Tutorial de Digitalización de imágenes – Metadatos. Departamento de preservación y conservación, Biblioteca de la Universidad de Cornell. En: <http://www.library.cornell.edu/preservation/tutorial-spanish/metadata/table5-1.html>

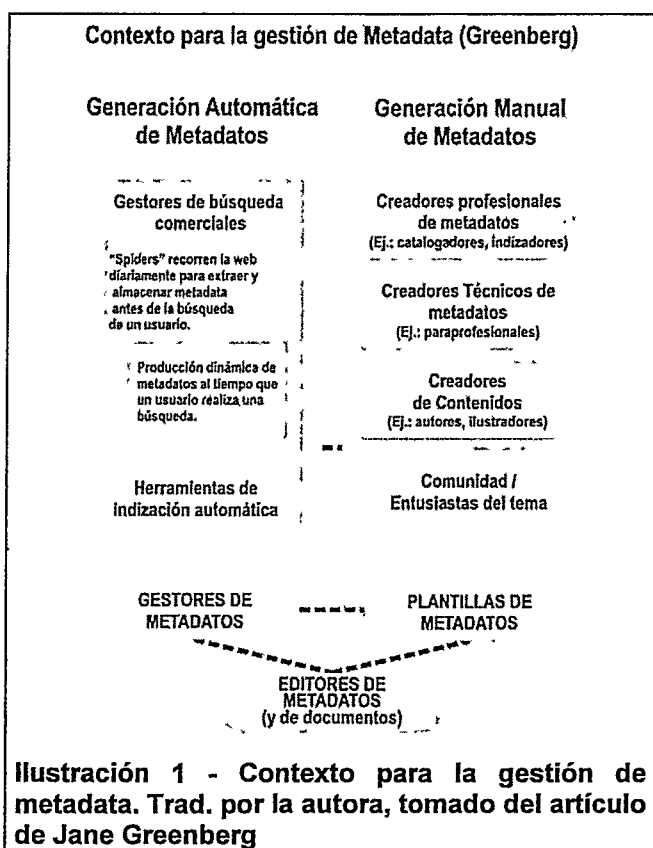
13 Gilliland-Swetland, Anne J. La definición de los metadatos. En: Introducción a los metadatos: vías a la información digital. Murtha Baca (ed.) Los Ángeles: Getty Institute Trust, 1999. pp. 1-9.

Según el tipo de relación entre el documento y el usuario encontramos que los metadatos permiten:

Desde el punto de vista del productor: Proporcionar una descripción de una entidad de información.

Desde el punto de vista del usuario: su localización, identificación, evaluación y/o selección.

Desde el punto de vista del gestor: Facilitar la gestión/codificación de los recursos descritos; mejorar el descubrimiento y localización de la información; controlar el acceso y uso de los recursos electrónicos.



Contexto para la gestión de metadatos (Greenberg)

Existen numerosas discusiones acerca de quién debe y a qué documentos se les deben aplicar esta herramienta. Coincidiendo con Jane Greenberg¹⁴, optamos por la siguiente descripción:

La generación de metadatos se realiza con el apoyo de los siguientes elementos:

14 Greenberg, Jane. Metadata Generation : Processes, People and Tools. En: Bulletin of the American Society for Information Science and Technology, December/January 2003. pp. 16-19.

- a) Los seres humanos: considerados como herramientas intelectuales con la capacidad de ejercer discreción y llevar a cabo la introducción de datos.
- b) Normas y documentación: las especificaciones de metadatos, las directrices de contenido, tesauros, listas de clasificación y otros tipos de normas y documentación que rigen la creación de metadatos.
- c) Dispositivos: compiladores técnicos de "captura" y "almacenamiento" de metadatos, ya sea en una base de datos o en el encabezado de los recursos (por ejemplo, la "HEAD" de un documento HTML o XML). Podemos incluir en esta categoría a las plantillas, editores y generadores.

La función de asignación de metadatos a un documento electrónico puede darse a través de dos formas:

-Automática-semiautomática, para describir estos conceptos traducimos las definiciones presentadas por Greenberg¹⁵:

Concepto de Generación Automática de Metadatos	Ejemplo/s
Extracción de metadatos Es el proceso automático por el cual se extraen los metadatos del contenido de un recurso. El contenido del recurso es recorrido para producir metadatos estructurados ("etiquetados") que representan al objeto.	La extracción de metadatos para una página web incluye la extracción de metadatos del contenido de un recurso que es visualizado a través de un navegador web.
Recolección de metadatos Proceso de recolección automática de metadatos que el recurso tiene embebidos o	Un software que recogerá los metadatos que se encuentran en la sección del código fuente de "HEAD" de un documento HTML (o XHTML),

15 Greenberg J., Spurgin, K., Crystal, A., Cronquist, M., Wilson, A.. Final Report for the AMeGA (Automatic Metadata Generation Applications) Project, UNC School of information and library science, 2005. http://www.loc.gov/catdir/bibcontrol/lc_amega_final_report.pdf

asociados. La recolección de metadatos, originalmente es realizada por personas o puede realizarse a través de un software de modo total o semiautomáticamente.	(por ejemplo, "Palabras clave" meta-tags). O los metadatos de un documento de Microsoft Word que se encuentran en las propiedades del archivo (por ejemplo, "Tipo de archivo", que se genera automáticamente, y "Palabras clave", que puede ser añadido por el autor del recurso).
Generación completamente automatizada de metadatos Completa (o total) dependencia de procesos automáticos para crear metadatos.	El software de edición web (por ejemplo, Macromedia Dreamweaver y FrontPage de Microsoft) y el software de gestión del documento (por ejemplo, Microsoft Word y Acrobat) que producen automáticamente los metadatos al momento en que un recurso es creado o actualizado (por ejemplo, "Fecha de creación" o "Fecha de modificación"), sin intervención humana
Generación semi-automática de metadatos dependencia parcial del software para crear metadatos, una combinación de procesos totalmente automáticos y humanos para crear metadatos	<p>1. Técnicas completamente automáticas que se utilizan para generar los metadatos (por ejemplo, "Palabras clave") como un primer paso, y el programa presenta los metadatos a una persona que de forma manual puede editar los metadatos.</p> <p>2. El software puede presentar a una persona (por ejemplo, autor de los recursos o arquitecto web) una "plantilla" que guía la introducción manual de los metadatos, a continuación, convierte automáticamente los metadatos a una codificación adecuada (por ejemplo, las etiquetas XML). El software puede incluso integrar automáticamente los metadatos a un recurso.</p>

Tabla 2- Concepto de Generación Automática de Metadatos

(Traducción de la autora)

, y

-Manual: Cabe señalar que las personas que realizan esta tarea se pueden clasificar del siguiente modo:

- o Creadores profesionales de metadatos, en este grupo encontramos a los catalogadores, indizadores y otras personas con formación específica en Ciencias de la Información o Bibliotecología.
- o Creadores técnicos de metadatos: paraprofesionales o auxiliares técnicos con entrenamiento básico en la aplicación de estas herramientas.
- o Creadores de contenidos: los autores intelectuales de los contenidos del documento digital a describir, entre otros encontramos a los autores del texto, ilustradores, fotógrafos.
- o Comunidad/Entusiastas del tema: Cualquier usuario de la información que mediante la utilización de herramientas de anotación puede describir la información conforme a sus opiniones/necesidades.

Las etapas que tienen que ver con el proceso de producción, disposición y provisión de acceso al objeto digital en cuestión se podrían definir de la siguiente manera:

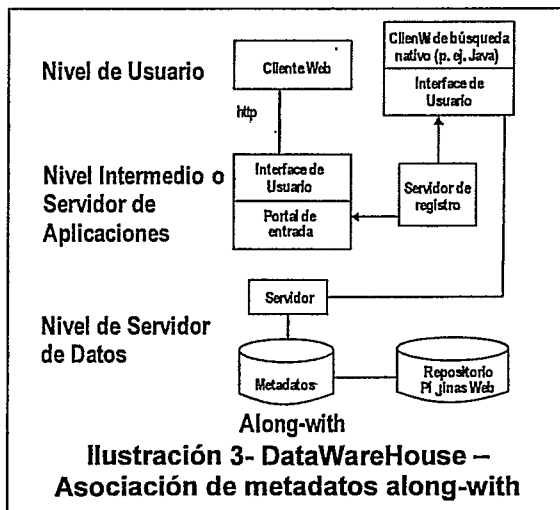
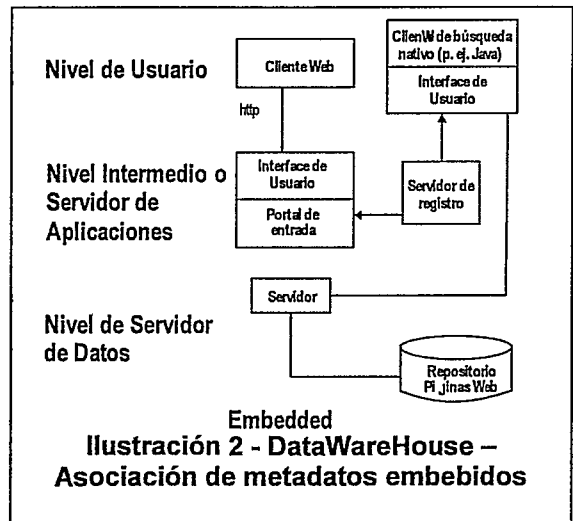
- Etapa de producción: autor del documento.
- Etapa de publicación: Servidor: Permite una declaración efectiva y una mejor gestión.
- Asignación de forma externa: por parte de un buscador o un sistema bibliotecario, cuando el recurso se haya disponible.

Con respecto al sujeto que debe realizar este tipo de operaciones, adherimos a los postulados de DC, que sugieren y facilitan que la inclusión de esquemas de metadatos en los e-docs debe ser realizada por los propios autores de los objetos en cuestión.

En cuanto a qué describir, se sostiene mayoritariamente, y por razones tanto de costo temporal cuanto económico y de utilidad, que los documentos a los que se les debe aplicar esta tecnología son los de tipo científico (comunicación científica).

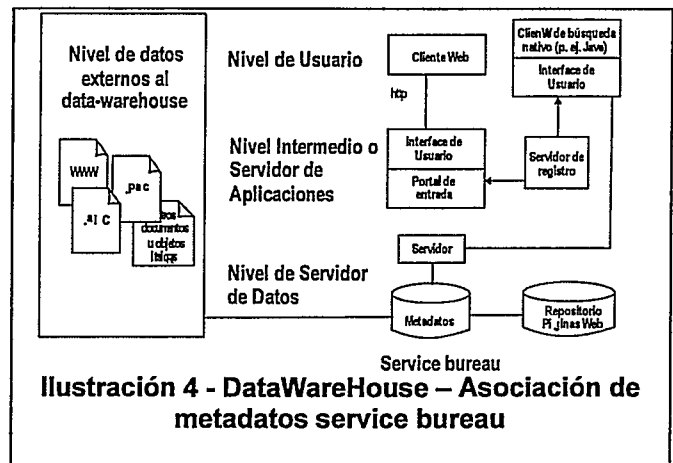
Los metadatos se pueden asociar a los documentos según los siguientes modos:

- Embebidos / incrustados: es decir encapsulados dentro de los recursos (EJ TEI / sintaxis <<meta>> HTML: DC)



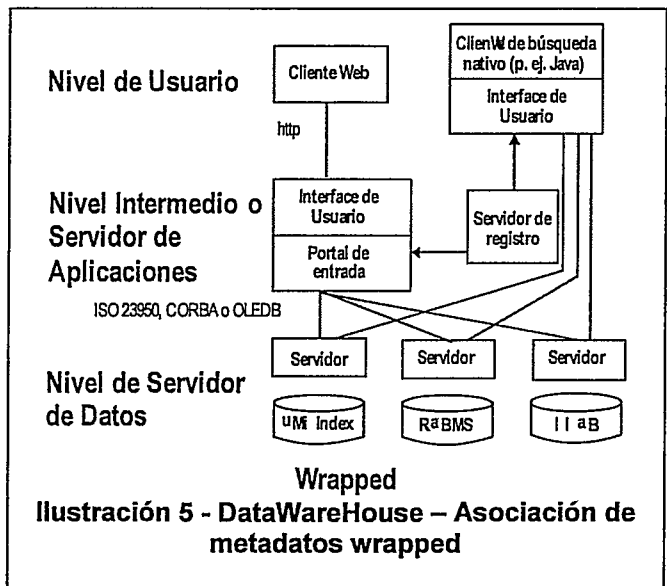
- *Along-with* (junto con): los metadatos son externos al recurso son equiparados mediante un mecanismo de transferencia como respuesta a una recuperación de información.

- *Service bureau* (departamento de servicios: cuerpo que proporciona información computada al público y puede procesar datos), utilizados



separadamente de los recursos por una recuperación de información, eventualmente de diferentes bases de datos. (EJ. MARC/UNIMARC).

- *Wrapped* (envuelto): el recurso está encapsulado en los metadatos que lo describen. EAD / URI / DOI.¹⁶



Para realizar la tarea de asignación de metadatos es fundamental

ajustarse a normativas internacionales, consistentes, adaptables a diferentes escalas y bien aceptadas, dado que permitirán asegurar la interoperabilidad entre

¹⁶ Lupovici, Catherine. Identification des ressources sur Internet et métadonnées: diversité des standards. EN Documentaliste – Sciences de l'information, 1999, vol. 36, N°6, p. 221-325.

sistemas actuales y futuros.¹⁷ En palabras de Ángela Sorlí Rojo y José Antonio Merlo Vega: “La normalización es un proceso necesario que afecta a un gran número de ámbitos del pensamiento, la gestión, la ciencia y la técnica. La aprobación de estándares comunes posibilita que los procedimientos se puedan realizar de igual forma en distintos países y zonas, con las posibilidades de entendimiento y cooperación que ello supone. Por lo general, las organizaciones responsables de la normalización suelen ser agencias nacionales, ya sean estatales o entidades en las que se ha delegado esa potestad; no obstante, existe un gran número de asociaciones especializadas en un sector y de federaciones que agrupan a organismos nacionales o regionales.”¹⁸

Cada comunidad científica ha dedicado esfuerzos en desarrollar un estándar adecuado a las necesidades de la disciplina representada por tanto se pueden encontrar diferentes iniciativas con características particulares. De todas formas, independientemente de su origen, cada una de ellas debe cumplir con una serie de condiciones:

- Ser fácil de crear y mantener.
- Utilizar una semántica que pueda entenderse de forma común.
- Poder crearse de forma automática.
- Poder describir la forma, el contenido y la localización de la información.

17 Estableciendo normas [OCLC - Why OCLC?] EN
<http://www.oclc.org/americalatina/es/digitalpreservation/why/standards/default.htm>

18 Sorlí Rojo, Ángela y Merlo Vega, José Antonio. Organizaciones de normalización en Internet. Revista Española de Documentación Científica, jul.-sept. 2000, vol. 23, n. 2, p. 327-340.

- Su estructura debe permitir ser contenida en otros objetos.
- Se debe poder usar para construir múltiples índices.
- Poder interoperar en los sistemas de indización.
- Poder ampliarse según las necesidades.

Estos formatos constan de un conjunto de elementos y la definición de una estructura tanto sintáctica como semántica.

- La sintaxis responde a la definición de la forma en que debe incluirse cada elemento, por ejemplo, cómo indicar que se trata de un elemento de metadatos, su denominación, el modo de describirlo y si se pueden aplicar atributos o no.
- La semántica está dada por el significado que se le confiere a cada elemento-componente de un esquema y su descripción general y de atributos específicos.
- Mediante los atributos se indican esquemas utilizados provenientes de normas establecidas.

Para poder responder de forma coherente a las necesidades de descripción de información se pueden utilizar estándares de metadatos.

4. Esquemas de metadatos

Se han desarrollado diferentes formatos según fueron surgiendo iniciativas por grupos pertenecientes a determinadas especialidades, quienes se ocuparon de desarrollar esquemas para cubrir las necesidades específicas requeridas para el tratamiento de la información en su ámbito de desarrollo.

Se intenta generar formatos con un mayor potencial de uso por parte de diferentes comunidades y que permita la interoperabilidad entre los sistemas y compatibilidad entre formatos.

Según Stuart Weibel¹⁹ un esquema de metadatos responde a una especificación formal de semántica y estructura de una colección/conjunto coherente de atributos que pueden ser asignados en la descripción de un recurso, además puede indicar las restricciones en la manera en que pueden aplicarse tales descripciones. Analizando separadamente los términos del concepto:

- *Formal*: en esta definición, significa que es mantenido por una agencia autorizada (authoritative). Esto implica adicionalmente que la especificación en sí misma puede ser bien definida como estructura de campos, etiquetas de campos, y tipos de datos permitidos.

19 Weibel, Stuart. RE: metadata schemes. EN: OCLC Office of Research Listserv Archives List_Serv: Metamarda-L METAMARDA-L@cornell.edu, <http://orc.dev.oclc.org:5103/metamarda-l/msg00104.html>. [99/04/01] Copyright © 1998 OCLC Online Computer Library Center.

- *Semántico*: se refiere a que los atributos sean comprensibles (plenos de significado) para los humanos.
- *Estructural*: las características de codificación de un valor (¿es un valor discreto?, ¿Un rango?, ¿un valor compuesto ordenadamente?...)
- *Coherente*: es usado en el sentido de una relación ordenada de elementos como constitutivos de un conjunto lógico.
- *Atributos*: son campos o elementos que pueden tener, a su vez, una subestructura.
- *Restricciones*: pueden incluir número y orden de los valores posibles, opcionalidad, tipos de datos permitidos, especificación de secuencia (collating sequence), etc.

Dempsey y Heery²⁰ definen que los formatos se pueden clasificar en tres bandas de acuerdo con su complejidad o por su riqueza semántica o estructural más el entorno de uso, método de creación, función del registro, protocolos de búsqueda y recuperación, y su estado como norma internacional:

- a) Banda 1. Formatos simples: elaborados por quienes los utilizan (sistemas propietarios) a través de mecanismos de indización full-text (*Lycos, Altavista, Yahoo,...*)

20 1998

- b) Banda 2. Formatos estructurados: estándares en desarrollo, estructura de campos (*Dublin Core, IAFA, RFC 1807, SOIF, LDIF*)
- c) Banda 3. Formatos complejos o Arquitecturas: formatos ricos, estándares internacionales, etiquetas elaboradas. Convenciones comunes sobre la semántica, sintaxis, estructura y diseño de infraestructuras que posibiliten la gestión de información y proporcionen la capacidad de transformar el web en un recurso de información más útil y poderoso (*ICPSR, CIMI, EAD, TEI, MARC*).

4.1. La elección de un esquema de metadatos

4.1.1. Consideraciones generales

La elección de un esquema de metadatos podrá basarse en los siguientes pilares:

1) Cumplimiento de las características fundamentales de un esquema de metadatos:

- Ser fácil de crear y mantener.
- Utilizar una semántica que pueda entenderse de forma común.
- Poder crearse de forma automática.
- Poder describir la forma, el contenido y la localización de la información.
- Su estructura debe permitir ser contenida en otros objetos.
- Se debe poder usar para construir múltiples índices.

- Poder interoperar en los sistemas de indización.

- Poder ampliarse según las necesidades.

2) Finalidad del sitio a describir; hay diferentes esquemas según el tipo de información y objeto documental y los usuarios a quienes está dirigida.

3) Disponibilidad de hardware y software. Se tomarán en cuenta decisiones en torno al sistema de asociación de metadatos a utilizar y generación de herramientas de búsqueda y recuperación de información.

4.1.2. Propuestas y modelos de metadatos

La gran diversidad de objetos digitales y de comunidades que los producen, comparten y difunden, obliga a que se adopten normas para la descripción mediante metadatos. Estos modelos, producidos por diversas comunidades, deben definir sus reglas, indicar su forma de uso y ser compatibles entre sí.

Todas estas propuestas y modelos se pueden clasificar -relacionándolos con otras normas, vocabularios y herramientas de codificación- desde varios puntos de vista, como pueden ser:

- △ el proceso en el que están involucrados -creación, adquisición, organización, diseminación y/o preservación-,
- △ los tipos de DLO que describen -texto, imagen fija, imágenes en movimiento, sonido-
- △ la institución o comunidades profesionales que los desarrollan - instituciones expertas en la gestión de información, medicina, leyes, comercio, industria de la construcción, geografía- y
- △ la función que cumplen -modelos conceptuales, estándares de contenido, vocabularios controlados, contexto/tecnología, lenguajes de marcado, formatos de grabación, normas de estructura-²¹.

21 Riley, Jenn. Seeing Standards: A Visualization of the Metadata Universe.
<http://www.dlib.indiana.edu/~jenrile/metadatamap/>

Como podemos visualizar en la siguiente figura:



Para atenernos solo a los formatos de metadatos, y para simplificar su comprensión, podemos adoptar la clasificación propuesta por Méndez Rodríguez²², señalando que existen:

- △ “Metadatos de propósito general: aquellos formatos dirigidos a cualquier tipo de información, es decir, destinados a describir cualquier DLO. El ejemplo por excelencia es el Dublin Core, pero también otros formatos como METS (*Metadata Encoding and Transmission Standard*, dirigido a describir cualquier DLO en el seno de una biblioteca digital).
- △ Metadatos de propósito específico: vinculados a un tipo de información digital o a un dominio temático o comunidad informativa, en particular, como por ejemplo:
 - Formatos de metainformación para información geográfica: U.S. Federal Geographic Data Committee (FGDC). Content Standards for Digital Geospatial Metadata, o la norma ISO 19115-2003.

22 Méndez Rodríguez, Eva. Metadatos y recuperación de información : estándares, problemas y aplicabilidad en bibliotecas digitales. España : Ediciones Trea, 2002.

- Formatos para describir información de archivo, como EAD, *Encoded Archival Description*.
 - Modelos de metadatos para la información textual de contenido literario y humanístico, como el TEI-Header en la Text Encoding Initiative.
 - Esquemas de metadatos para imágenes, como VRA-Core de la Virtual Resources Association.
- Además de todos estos esquemas de metadatos el W3C ha desarrollado RDF, *Resource Description Framework*, una infraestructura para la descripción de recursos en la Web que más que un mero formato de metainformación es un "metamodelo de metadatos" que permite codificar distintos esquemas de metadatos y además crear otros vocabularios específicos a través del *Lenguaje para la Descripción de Vocabularios RDF* o RDF schema (RDFS)."

4.1.3. Software para creación, implementación y gestión de metadatos

“Cada vez existe más software para la creación, implementación y gestión de metadatos. Estas herramientas, de índole, objetivos y procedencias diversa pueden ser tipificadas genéricamente como: plantillas que funcionan en modo servidor y aplicaciones cliente, además de distinguir, por supuesto, software libre,

de fuente abierta o comercial²³. A continuación, señalamos algunos ejemplos de estas herramientas:

Retomando el planteo realizado por Greenberg²⁴, podemos indicar tres grandes categorías de software para la creación, implementación y gestión de metadatos:

- Las plantillas: son listados del conjunto de elementos de un esquema de metadatos en particular que permiten la introducción manual de los valores correspondientes al DLO en cuestión y que generan como salida un formato con los atributos de los elementos y sus valores para vincular o integrar al DLO. No contienen información acerca de las normativas o modos en que deben cargarse los valores del esquema utilizado.
- Los gestores o generadores automáticos de metadatos: son los que permiten la extracción automática de metadatos. Mediante el acceso al DLO utilizan un algoritmo de rastreo del contenido del objeto, incluyendo su código fuente, asignan automáticamente los metadatos. Producen automáticamente un registro de acuerdo a un esquema específico que luego es posible asignar al documento o integrar a una base de datos. Los resultados del uso de esta herramienta aún son bastante erráticos para elementos intelectualmente exigentes, como los de descripción temática. ;y

²³ Méndez Rodríguez, Eva y Senso, José A. Introducción a los metadatos ; estándares y aplicación. Madrid : Sedic, 2004. Méndez Rodríguez, Eva y Senso, José A. Introducción a los metadatos ; estándares y aplicación. Madrid : Sedic, 2004. EN <http://www.sedic.es/autoformacion/metadatos/index.htm>.

²⁴ Greenberg, Jane. Metadata Generation : Processes, People and Tools. En: Bulletin of the American Society for Information Science and Technology, December/January 2003. pp. 16-19.

- Los editores: Son similares a las plantillas, aunque más sofisticados, ya que proporcionan acceso directo a las normas y a la documentación de la creación de metadatos subyacente. También suelen brindar asistencia automática sobre los aspectos sintácticos de creación de los metadatos.

A las descritas les podemos sumar herramientas híbridas, que son aquellas que combinan aspectos de los editores y los generadores automáticos de metadatos, que son las que permiten la edición de los valores de los metadatos luego de que estos son generados automáticamente.

Cabe destacar también la existencia de herramientas de conversión de formatos, que son las que permiten traducir un formato en otro y que se utilizan para trabajar en ámbitos de interoperabilidad.

Las herramientas de metadatos son generalmente desarrollados para apoyar esquemas de metadatos o conjuntos de elementos específicos. Los sitios web de un esquema particular, con frecuencia, tienen vínculos a conjuntos de herramientas pertinentes.

4.1.4. Etapas del proceso de implementación de metadatos

- **Etapas 1:** Planificación o producción de requerimientos
 1. Delimitación del ámbito del proyecto
 2. Estudio de viabilidad
 3. Análisis de riesgos

4. Estimación de costos
5. Planificación temporal
6. Asignación de recursos

- **Etapas 2:** Análisis – elicitación de requerimientos

Descripción clara y completa de qué es lo que se pretende, incluyendo la presentación de los resultados que se desean obtener (salidas) y la forma en que se va a utilizar la aplicación (interfaz de usuario):

1. Relevar los estándares internacionales de metadatos.
2. Seleccionar el esquema o esquemas más apropiados a los efectos de descripción del corpus documental a describir.
3. Determinar las herramientas de software y hardware necesarias para la implementación del esquema de metadatos seleccionado.

- **Etapas 3:** Implementación

1. Etiquetación de contenidos.
 - Crear una plantilla de documento con el formulario de entrada de metadatos.
 - Crear un proceso simple de revisión de metadatos.
 - Crear un registro de metadatos XML que contenga sólo los campos seleccionados utilizando la DTD IMS o XML-Schema.

- Crear un modelo completo con valores de metadatos en cada campo seleccionado.
- Revisar los metadatos. Ajustar lo que sea necesario.
- Mostrar cómo trasladar los campos de datos desde el formulario al un registro XML vacío usando un editor XML.
- Crear un directorio enlazando los nombres de archivos a los títulos de los módulos.
- Guardar el registro XML resultante.
- Escribir instrucciones para ingresar metadatos para los usuarios.
- Hacer una pequeña biblioteca de registros XML.

2. La base de datos.

- Definir un archivo de búsqueda. Qué campos desea para esta operación?
- Crear la estructura de registros de la Base de Datos.
- Crear un método para mapear o agregar los campos a los campos XML a los registros de la base de datos.
- Cargar la base de datos desde la biblioteca XML.

3. Búsqueda en la base de datos.

- Desarrollar la interfaz de búsqueda.
 - Realizar un examen de la interfaz de búsqueda.
 - Usar la interfaz de búsqueda.
- **Etapa 4: Evaluación y retroalimentación del sistema**

4.1.5. La elección de un formato de metainformación para el sitio Web de la LICAD

La constitución de un sitio web educativo de calidad requiere, entre otras cosas, de un sistema eficiente de identificación y recuperación de información, y con más razón tratándose de un sitio que se ocupa de temáticas del área de la Bibliotecología y la Documentación. Por lo tanto, se hace imperioso que los documentos que lo conforman se encuentren descritos con el mejor nivel de detalle que sea posible:

- Condicionamientos.
- Posibilidades
- Keywords y description
- Modelo Dublin Core Metadata Initiative [DCMI].
- Sintaxis de codificación: HTML
- Justificación: motor de búsqueda

En el capítulo siguiente dedicaremos especial atención a Dublin Core Metadata Initiative [DCMI] y Resource Description Framework [RDF].

5. El Modelo DC (ISO Standard 15836-2003 - ANSI/NISO Z39.85-2007 - IETF RFC 5013-2007)

El esquema de metadatos Dublin Core se almacena en el área de la cabecera (<HEAD>) de un documento HTML y no es visible por los web-browsers. Al estar embebido en el documento, permanece con éste al ser creado, movido o actualizado.

La habilidad para buscar o encontrar información se ve mejorada por los vocabularios controlados asociados a los elementos de metadatos. Adicionalmente, como metadatos combinados con índices temáticos controlados, permite una mayor precisión en la búsqueda y gestión de documentos.

Concebido originalmente para la descripción de recursos Web por autor, se ha convertido en foco de atención de comunidades involucradas en la descripción formal de recursos, tales como museos, bibliotecas, agencias gubernamentales y organizaciones comerciales.

La adopción del esquema de metadatos DC permite una mayor facilidad y exactitud en la recuperación de información, y son razones importantes para su adopción:

- Facilidad de creación y sencillez de descripción
- Simplicidad para indizar y describir localización de recursos, formatos, etc.
- Ofrece vocabulario controlado que posibilita una precisión mayor en las recuperación que en las búsquedas a texto completo

- Es un estándar aprobado (y fomentado) por el W3C
- Ofrece extensibilidad e interoperabilidad con otros estándares
- Acrecienta la calidad en la gestión de recursos

Entre sus principales características distinguimos:

- Extensibilidad: se podrá ampliar el grupo con elementos más complejos para la descripción intrínseca, de particular importancia en una comunidad determinada.
- Opcionalidad: todos los elementos son opcionales.
- Repetibilidad: todos los elementos son repetibles.
- Modificabilidad: cualquier elemento puede ser modificado por uno o más cualificadores.

El esquema de metadatos DC para descripción de materiales accesibles vía red, definido por el Dublin Core Working Group (W3C / NISO / ANSI), consta de una lista descriptiva de 15 elementos, entre los que reconocemos tres grupos:

- Los elementos relacionados principalmente con los contenidos del recurso (título, materias y palabras clave, descripción, fuente, idioma, relación, cobertura)
- Los relacionados con la propiedad intelectual (autor o creador, editor, otros colaboradores, derechos)
- Estatus (Instantiation) (Fecha, tipo, formato, identificador)

La sintaxis para la aplicación del esquema DC está constituida por la codificación de cada elemento con la etiqueta META, incluyendo dos atributos básicos NAME (elemento) y CONTENT (contenido), pudiendo adicionar, en los casos que sea necesario, los calificadores SCHEME (esquema) o TYPE (tipo).

“Los calificadores del DCMI permiten aumentar la especificidad y precisión de los metadatos. [...] En la determinación de los calificadores, se dio preferencia a vocabularios, notaciones y términos

<pre><META NAME="DC.Date" SCHEME="ISO8601" CONTENT="2002-11-25"></pre>
--

Ilustración 7 - Modelo de etiqueta cualificada de Metadatos según formato DCMI

actualmente mantenidos por agencias establecidas. Inevitablemente, algunos usuarios encontrarán que los calificadores no les son familiares y son para dominios muy específicos.

El DCMI (Dublin Core Metadata Initiative) reconoce dos grandes clases de calificadores:

- **Refinación de elementos (TYPE):** Estos calificadores hacen que el significado de un elemento sea más estrecho o más específico. Un elemento refinado comparte el significado del elemento no calificado, pero con un alcance más restrictivo. Si un agente no entiende un término de refinamiento específico para un elemento, debe ser capaz de ignorarlo y tratar el valor del metadato como si estuviese sin calificar. Las definiciones de términos para refinamiento de elementos deben estar públicamente disponibles.
- **Esquema de codificación (SCHEME):** Estos calificadores identifican esquemas que ayudan en la interpretación del valor de un elemento. Dichos esquemas incluyen vocabularios controlados y notaciones formales o reglas de parseo (reglas de equiparación automatizada de variables). Un valor expresado usando un esquema de codificación será entonces un símbolo (token) escogido de un vocabulario controlado (por ejemplo, un término de un sistema de clasificación) o un “string” con formato de acuerdo con una notación formal (por ejemplo, “2002-01-01” como la expresión estándar de una fecha). Si un esquema de codificación no es entendido por un agente, el valor será de todas maneras, útil para un lector humano. La descripción definitiva de un esquema de codificación para

calificadores debe estar claramente identificada y disponible para uso público.²⁵

```
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/">
<link rel="schema.DCTERMS" href="http://purl.org/dc/terms/">
<meta name="DC.title" lang="es" content="Módulo Procesamiento - Eje 2: Los lenguajes documentales">
<meta name="DC.creator" content="Gustavo Liberatore, Marcela Coringrato">
<meta name="DC.subject" lang="es" content="lenguajes documentales, tesauros, ciclo documental,
indización, lenguaje natural, sistema, función, conceptos, núcleo, información">
<meta name="DC.description" lang="es" content="Eje 2 de la asignatura Procesamiento de la Información
de la carrera Licenciatura en Bibliotecología y Documentación a Distancia de la UNMDP">
<meta name="DC.publisher" content="Universidad Nacional de Mar del Plata. Departamento de
Documentación">
<meta name="DC.contributor" content="Andrés Vuotto">
<meta name="DC.date" scheme="DCTERMS.W3CDTF" content="2005-04-30">
<meta name="DC.type" scheme="DCTERMS.DCMIType" content="Text">
<meta name="DC.format" content="text/html">
<meta name="DC.format" content="5844 bytes">
<meta name="DC.identifier" content="proseje2.htm">
<meta name="DC.language" content="spanish">
<meta name="DC.relation" content="Módulo Procesamiento - Eje 2: Los lenguajes documentales">
<link rel="DC.rights" href="http://www.mdp.edu.ar/humanidades/documentacion/licad/rights.htm">
```

Ilustración 8 - Cabecera HTML en DC

Los 15 elementos del DC y sus características son:

Elemento 1: Título

Etiqueta: DC.Title

Descripción: El nombre dado al recurso por el creador/editor

Esquema: No necesario.

Tipo:

- DC.Title (incompetente)

25 Calificadores del estandar de metadatos Dublin Core (Dublin Core Qualifiers). Traducción al español de Dublin Core Qualifiers <http://www.dublincore.org/documents/dcmes-qualifiers/>. Marcela Calderón Corail (Traductora) EN <http://www.ciw.cl/proyectos/calificadoresDC.html>

- DC.Title.Alternative (usado para cualquier otro título aparte del principal incluyendo subtítulos, etc.)

Elemento 2: Autor

Etiqueta: DC.Creator

Descripción: La persona(s) u organización(es) principalmente responsable del contenido intelectual del recurso. Por ejemplo, autores en el caso de documentos de texto, artistas, fotógrafos, o ilustradores en el caso de recursos visuales.

Esquema: LCNAF (Library of Congress Name Authority File).

Tipo:

- DC.Creator (incompetente)
- DC.Creator.PersonalName
- DC.Creator.CorporateName (incluye nombre de conferencia)
- DC.Creator.PersonalName.Address (incluye cualquier tipo: dirección postal, e-mail u otro)
- DC.Creator.CorporateName.Address

Nótese que los calificadores enumerados en el documento de Knight/Hamilton extienden el elemento más que refinarlo (ej., código postal, teléfono, fax, afiliación, etc.). Solamente la dirección se incluye aquí porque se ha utilizado con frecuencia en proyectos actuales de metadata. El e-mail se puede considerar una manera alternativa de especificar al autor. Si otros calificadores son necesarios (e.j. afiliación), deben ser utilizados como extensiones locales.

Elemento 3: Palabras clave

Etiqueta: DC.Subject

Descripción: El t3pico del recurso. T3picamente, materia expresada como palabras clave o frases que describen el tema o contenido del recurso. El uso de vocabularios controlados y esquemas formales de clasificaci3n es aconsejable.

Esquema: (incompetente: Palabra clave por defecto)

- LCSH (Library of Congress Subject Headings)
- MeSH (Medical Subject Headings)
- AAT (Art and Architecture Thesaurus)
- LCNAF (Library of Congress Name Authority File): para nombres usados como materia
- DDC (Dewey Decimal Classification)
- LCC (Library of Congress Classification)
- NLM (National Library of Medicine Classification)
- UDC (Universal Decimal Classification)

Estos cualificadotes son los usados m3s frecuentemente. El esquema podr3a se3alar la lista controlada mantenida por la Biblioteca del Congreso en la lista de c3digo de USMARC para "Relators", fuentes, convenciones de descripci3n, que incluye muchos otros (la parte III: Fuentes y parte IV de la clasificaci3n: Fuentes Del T3rmino De Subject/Index; estos documentos est3n actualmente bajo revisi3n para una nueva edici3n). Es posible utilizar tesauros o esquemas de clasificaci3n localmente definidos.

Tipo:

- (incompetente)

Elemento 4: Descripción

Etiqueta: DC.Description

Descripción: Una explicación textual del contenido del recurso, puede incluir resúmenes, tabla de contenidos, descripción textual del contenido en el caso de objetos gráficos o texto libre.

Esquema: (Incompetente: Keywords por defecto)

- Abstract (default)
- URL

Tipo:

- (incompetente)

Elemento 5: Editor

Etiqueta: DC.Publisher

Descripción: La entidad responsable de que el recurso esté disponible en la forma actual, como una casa editora, un departamento universitario, o una entidad corporativa.

Esquema: (incompetente)

- No obligatorio

Tipo:

- DC.Publisher (sin calificar)
- DC.Publisher. PersonalName
- DC.Publisher. CorporateName
- DC.Publisher. PersonalName.Address
- DC.Publisher. CorporateName.Address

Elemento 6: Colaborador

Etiqueta: DC.Contributor

Descripción: Persona u organización no especificada en el elemento "creador", que ha realizado contribuciones intelectuales significativas al recurso pero que ese aporte es secundario respecto del realizado por la persona u organización especificadas en dicho elemento (por ejemplo, editor, transcriptor e ilustrador).

Esquema:

- LCNAF (Library of Congress Name Authority File)

Tipo:

- DC.Contributor (sin calificar)
- DC.Contributor.PersonalName
- DC.Contributor.CorporateName
- DC.Contributor.PersonalName.Address
- DC.Contributor.CorporateName.Address

En el DC4 se decidió que el rol no es necesario. De manera que no está incluido aquí. Si fuera necesario, se podría incluir como extensión local. En la Biblioteca del Congreso, por ejemplo, hemos encontrado que todos los nombres se buscan en el mismo índice y allí no tenemos ninguna necesidad de especificar rol más amplio en el trabajo. Se ha dejado de indicar esto en expedientes bibliográficos hace largo tiempo.

Elemento 7: Fecha

Etiqueta: DC.Date

Descripción: Fecha asociada a la creación o disponibilidad/accesibilidad del recurso. Nótese que esta definición ha cambiado en el DC5. La fecha puede ser una fecha simple o un rango de fechas.

Esquema:

- (Default is ISO 8601)
- ANSI X3.30
- IETF RFC 822
- (Otros?)

Tipo:

Se observa que los nombres de los subelementos no se han convenido. El Date Working Group continúa trabajando en éstos. Existió una cierta controversia en el final de DC5 si el primero de los dos tipos de subelementos debe ser combinado puesto que era también incierto incluir los últimos tres subelementos.

- DC.Date.Creation_of_intellectual_content
- DC.Date.Creation/Modification_of_present_form
- DC.Date.Formal_publication
- DC.Date.Available
- DC.Date.Valid (includes verification)
- DC.Date.Acquisition/Accession
- DC.Date.Accepted
- DC.Date.DataGathering

Elemento 8: Tipo de Recurso

Etiqueta: DC.Type

Descripción: La categoría del recurso –naturaleza o género–, como home-page, novela, poema, certificado de trabajo, informe técnico, ensayo, diccionario. En favor de la interoperabilidad, el tipo de recurso puede ser

seleccionado de una lista enumerada que está bajo desarrollo en el workshop series.

Esquema:

- No necesario, pero sí se requiere de una lista enumerativa de tipos de recursos.

Elemento 9: Formato

Etiqueta: DC.Format

Descripción: El formato del recurso, usado para identificar el software y el posible hardware necesario para visualizar/presentar u operar el recurso. Normalmente, se utiliza para incluir tipos de medios o dimensión de un recurso. En función de la interoperabilidad, puede ser seleccionado de una lista predefinida bajo desarrollo en el workshop series.

Esquema:

- IMT (i.e. MIME)
- DCPMT (Dublin Core Physical Medium Type) Este es un nuevo concepto discutido en el Subelement Working Group del DC5. Podría ser una lista enumerada de tipos de medios físicos cuando la definida por IMT resulte inadecuada (ej. un recurso no electrónico).

Tipo:

- No necesario

Elemento 10: Identificador de recurso

Etiqueta: DC.Identifier

Descripción: Cadena de caracteres, signos, letras o números usados para identificar de modo unívoco el recurso.

Ejemplos para recursos de red incluyen URLs y URNs (cuando son implementados). Otros identificadores globales-unívocos, como el International Standard Book Numbers (ISBN) u otras denominaciones formales que también puedan ser aplicables para este elemento en el curso de recursos off-line.

Esquema:

- (URL is default)
- URN (Uniform Resource Name)
- ISBN (International Standard Book Number)
- ISSN (International Standard Serial Number)
- SICI (Serial Item and Contribution Identifier)
- FPI (Formal Public Identifier)

Tipo:

No necesario

Elemento 11: Fuente

Etiqueta: DC.Source

Descripción: Una cadena de caracteres o número usado para identificar unívocamente el trabajo desde el que el recurso fue derivado, si es aplicable. Por ejemplo, una versión PDF de una novela derivada de un libro físico podría contener en el elemento fuente el ISBN correspondiente a esa edición.

Esquema:

- (Free text is default)
- URL
- URN
- ISBN
- ISSN

Tipo:

No necesario

Elemento 12: Idioma

Etiqueta: DC.Language

Descripción: Idioma(s) del contenido intelectual del recurso. En la práctica, el contenido de este campo coincide con RFC 1766. Ver: <http://ds.internic.net/rfc/rfc1766.txt>.

Esquema:

- IETF RFC 1766
- Z39.53
- ISO 639-1
- ISO 639-2/B (after final publication)

Tipo:

No necesario.

Obsérvese que la ISO 639 actual (hacer ISO 639-1 cuando el código 3-character de la ISO 639-2 es aprobado), que se utiliza en RFC 1766, cubre solamente cerca de 140 idiomas con respecto a la ISO estándar nuevamente desarrollada 639-2, que incluye cerca de 400 idiomas. Por ejemplo, la ISO 639-1 no distingue lenguas antiguas de idiomas modernos. Muchas agencias necesitarán la lista más extensa que la ISO 639-2 proporciona. La ISO 639-2 fue aprobada en 1997; la aprobación

del proyecto de norma internacional final es próxima, seguida por la publicación de la ISO. Se debe tener en cuenta, además, que la ISO 639-2 incluye un sistema bibliográfico (ISO 639-2/B) y un sistema terminológico (ISO 639-2/T). El primero, se prefiere para el metadata debido a su uso extendido en agencias bibliográficas. El nuevo estándar está más cercano a la norma Z39.53 que a la ISO 639-1.

Elemento 13: Relación/Vinculación

Etiqueta: DC.Relation

Descripción: Relación con otros recursos. La intención de especificación de este elemento es proveer una forma de expresar relaciones entre recursos que tienen relaciones formales con otros, pero existen como recursos discretos (diferenciados / independientes) entre sí. Por ejemplo, imágenes de un documento, capítulos de un libro, o ítems de una colección. Una especificación formal de este elemento está actualmente bajo desarrollo. Usuarios y desarrolladores entienden que el uso de este elemento puede ser considerado actualmente como experimental.

Esquema:

- (Default is free text)
- URL
- URN
- ISBN

Tipo:

Se considera que este elemento se va a desarrollar rápidamente en el contexto de las descripciones de colección-nivel. Los subelementos podrán ser definidos después. Las siguientes categorías de Relation fueron definidas en el DC5; éstas no deben considerarse como nombres formales para los subelementos.

- Creative (e.g. translation, annotation)
- Mechanical (copy, format change, mirror copy)
- Version (edition, draft)
- Inclusion (collection, part)
- Reference (citation)

Elemento 14: Cobertura

Etiqueta: DC.Coverage

Descripción: Se trata de las características espaciales, temporales y/o jurisdiccionales del recurso. La especificación formal de cobertura está actualmente bajo desarrollo. Usuarios y desarrolladores entienden que el uso de este elemento puede ser considerado actualmente como experimental.

Esquema:

- A ser definida por el Coverage Working Group

Tipo:

Los siguientes fueron determinados por el Coverage Working Group. Información más detallada en:

<http://www.sdc.ucsb.edu/~mary/coverage.htm> .

- DC.Coverage.PeriodName
- DC.Coverage.PlaceName
- DC.Coverage.t
- DC.Coverage.x
- DC.Coverage.y

- DC.Coverage.z
- DC.Coverage.Polygon
- DC.Coverage.Line
- DC.Coverage.3d

Elemento 15: Derechos legales

Etiqueta: DC.Rights

Descripción: Es un enlace a una nota de copyright o a derechos de uso o al servicio que podría proveer información acerca de los términos de acceso al recurso. La especificación formal de derechos está actualmente bajo desarrollo. Usuarios y desarrolladores entienden que el uso de este elemento es plausible de ser considerado actualmente como experimental.

Esquema:

- (Por defecto es texto libre)
- URL
- URN

Tipo:

No obligatorio. Los subelementos se aplican en implementaciones específicas.

RDF (Resource Description Framework): es una infraestructura que posibilita la codificación, intercambio y reutilización de metadatos estructurados; permite su interoperatividad a través del diseño de mecanismos que soportan convenciones

comunes de la semántica (standard de descripción de contenidos. Ej.: DC, TEI, AACR2, etc.), estructura (RDF: modelo de datos para esquemas específicos) y sintaxis (XML: lenguaje de marcado para estructurar información web). RDF suministra la capacidad de definir tantos elementos de metadatos cuantos necesite una comunidad de descripción de recursos, sin estipular la semántica para esa comunidad.

```
<?xml version="1.0"?>
<!DOCTYPE rdf:RDF SYSTEM
"http://dublincore.org/documents/2002/07/31/dcmes-xml/dcmes-xml-
dtd.dtd">

<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description rdf:about="proseje2.htm">
    <dc:title>
      0Ydulo Procesamiento - Eje 2: Los lenguajes documentales
    </dc:title>
    <dc:creator>
      Gustavo Liberatore, Marcela Coringrato
    </dc:creator>
    <dc:subject>
      lenguajes documentales,tesauros,ciclo
      GRFupentD, LnGLzDFLYn, OengudMe
      npturD, sLstED, funFLYn, FRnFestrs, núFOer, LnfrRPDFLYn
    </dc:subject>
    <dc:description>
      EMe 2 Ge Od dsLgnDturD 3rRFesDPLeNR Ge Od 'nfrRPDFLYn Ge
      Od FRrreRD /LfenFLDturD en BLEOLRteFRORgtD y
      DRFupentDFLYn D DLstDnFLD Ge Od U10D3
    </dc:description>
    <dc:publisher>
      rniversidad Nacional de Mar del Plata. Departamento de
      DRFupentDFLYn
    </dc:publisher>
    <dc:contributor>
      AnGrPs 9urttr
    </dc:contributor>
    <dc:date>
      2005-04-30
    </dc:date>
    <dc:type>
      Text
    </dc:type>
    <dc:format>
      text/html
    </dc:format>
```

```
<dc:format>
  5844 bytes
</dc:format>
<dc:language>
  spanish
</dc:language>
<dc:relation>
  Módulo Procesamiento
</dc:relation>
<dc:rights>
  http://www.mdp.edu.ar/humanidades/documentacion/licad/right
  s.htm
</dc:rights>
</rdf:description>
</rdf:oac>
```

Ilustración 9 - Cabecera en RDF un archivo HTML con metadatos DC

6. Aplicación de metadatos al sitio Web de la LICAD

Para agregarle utilidad y dinamismo a un sistema de información digital, tal como el que presenta actualmente la carrera de Licenciatura a Distancia en Documentación (LICAD) que dicta el Departamento de Documentación de la Facultad de Humanidades de la UNMDP, se requería una herramienta que permitiera de forma ágil la recuperación de cualquiera de sus componentes desde cualquier página que el usuario (alumno/docente) estuviera visitando. También debía tenerse en cuenta, como requisito fundamental, que dicho sistema de recuperación de información funcionara sin que mediara una conexión on-line para su lectura.

Se evaluaron diferentes sistemas de diseño/creación de buscadores para el Web, hasta encontrar la solución en la aplicación de comandos en lenguaje JavaScript, y a partir de la recolección de diferentes ejemplos y programas automatizadores de la tarea, se decidió aplicar en la primera etapa la herramienta automática de generación de un motor de búsqueda sencillo llamada SearchMakerPro, que a través de la extracción de los metadatos básicos (4 elementos: Título, Palabras Clave, Descripción, Autor) en las páginas contenidas a partir de un directorio dado, y en todas sus subcarpetas, genera un archivo de arrays (arreglos) y un formulario de búsqueda para ese archivo, que permite efectuar una recuperación por relevancia de las páginas indizadas. Desestimado luego, por su básica funcionalidad.

En función de lo expuesto, se decidió diseñar un buscador en JavaScript basado en el esquema DC de metadatos, para que cuando los sistemas funcionaran en línea cumplieran con los estándares internacionales.

Para describir esas páginas, se solicitó a los docentes de las asignaturas que señalaran palabras clave para cada una de ellas –el resto de la descripción fue realizado por la autora- y, mediante el programa MetaBot Pro (v3.51 de Watchfire), fueron integrados a las páginas siguiendo este procedimiento:

A) Generación de un proyecto de metadatos, mediante la opción de visualización en forma de planilla de todos los documentos contenidos en un directorio dado, en modo de acceso local y planteando la implementación de los esquemas “basic tags” y “Dublin Core” de metadatos (el programa permite utilizar esquemas predefinidos, u optar entre más de 50 etiquetas diferentes).

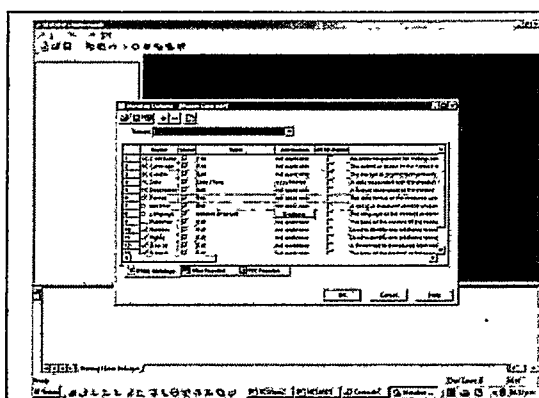


Ilustración 10 - Metabot Professional.
Generación de un proyecto de edición de etiquetas – selección de esquema (en este caso “Basic Tags” y “Dublin Core”)

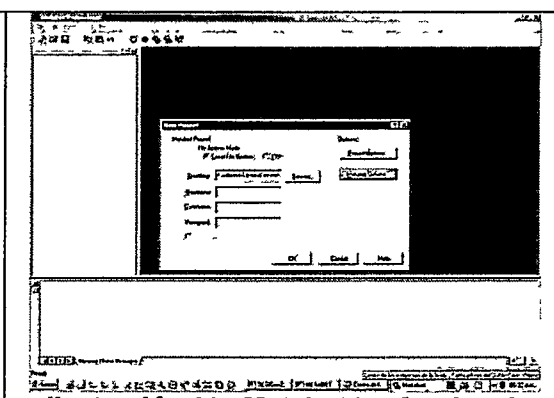


Ilustración 11 - Metabot Professional.
Generación de un proyecto de edición de etiquetas – selección de directorio

B) Se procedió a la descripción de cada uno de los documentos, y una vez finalizada ésta y revisados todos los campos se realizó su integración en los documentos del proyecto.

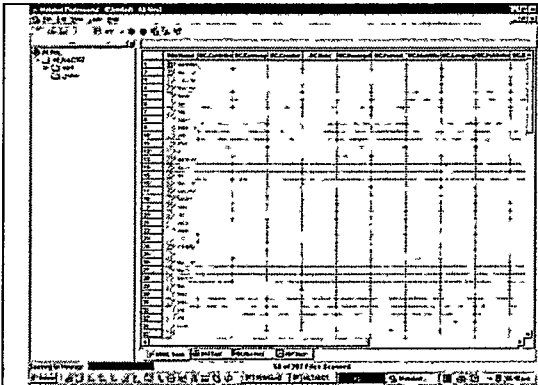


Ilustración 12 - Metabot Professional.
Generación de un proyecto de edición de etiquetas - conjunto de documentos sin editar

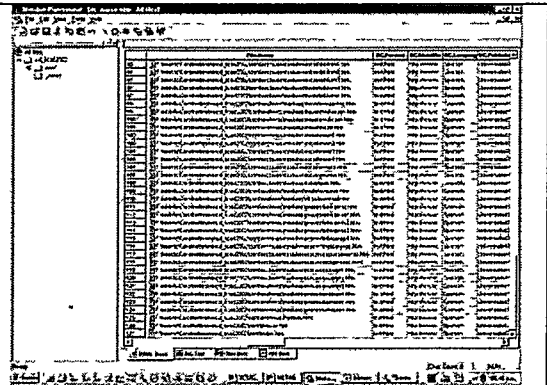


Ilustración 13 - Metabot Professional.
Generación de un proyecto de edición de etiquetas - conjunto de documentos editados

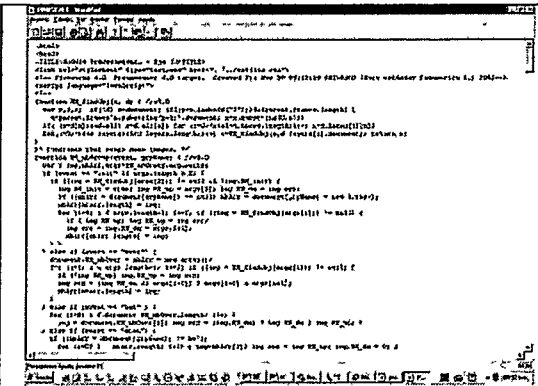


Ilustración 14 - Metabot Professional.
“HEAD” de uno de los documentos sin procesar

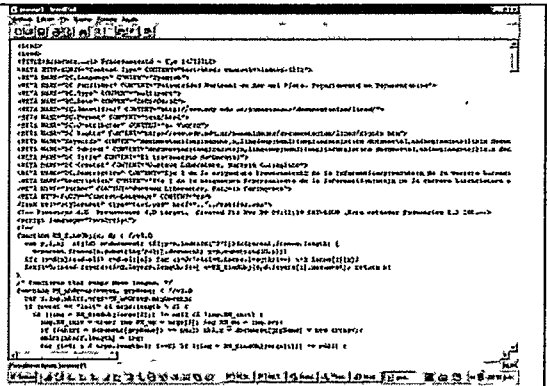


Ilustración 15 - Metabot Professional.
“HEAD” de uno de los documentos procesados

C) Al guardar el proyecto se integran las etiquetas de metadatos a cada documento que lo compone.

D) Una vez integrados los metadatos a las cabeceras de los documentos se procedió al diseño de un buscador en base a los elementos de DC sobre los archivos contenidos en el directorio de LICAD. Como resultado se generó el código fuente en JavaScript correspondiente al buscador y se almacenó en el archivo INDEX.HTML.

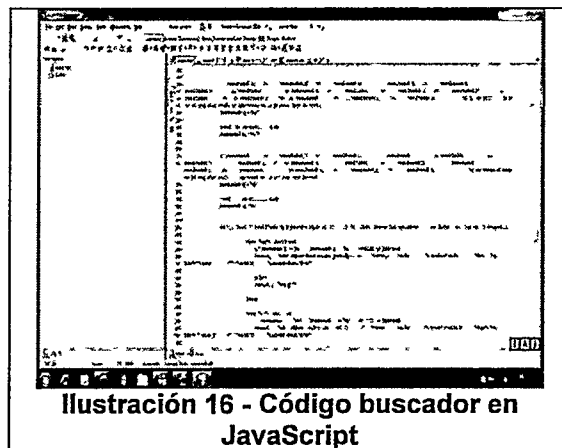


Ilustración 16 - Código buscador en JavaScript

E) El siguiente archivo BD.JS contiene la descripción de los archivos integrantes del módulo tomado como ejemplo para este trabajo.

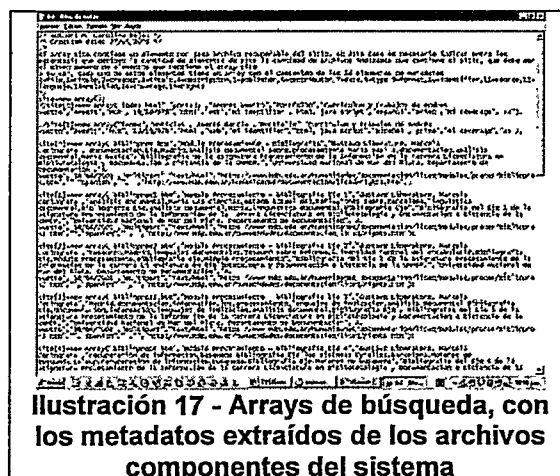
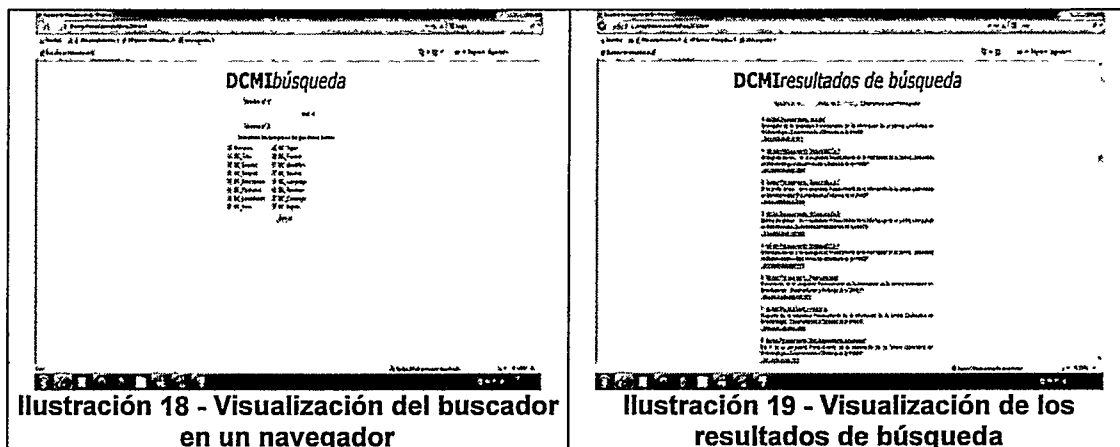


Ilustración 17 - Arrays de búsqueda, con los metadatos extraídos de los archivos componentes del sistema

F) El resultado final se puede observar en las siguientes figuras:



Como próxima fase en el desarrollo del sistema de búsqueda dentro de un sitio web navegable off-line nos resta generar los arreglos de acuerdo a cada módulo existente en el sitio web de LICAD, es decir, para que permita la búsqueda en todo el sitio, en una o en varias secciones. Y, finalmente, integrar esta herramienta a cada página del sitio y así facilitar la navegación dentro de la web de LICAD en su conjunto.

7. Conclusiones y reflexiones finales

Es indudable que Internet va a seguir creciendo indefinidamente y que por tanto es necesario que se redoblen nuestros esfuerzos en el desarrollo de sistemas efectivos de gestión de contenidos de documentos digitales.

“Entendemos que los esquemas de metadatos son una de las piezas fundamentales en este camino. El desafío es trabajar en el desarrollo de sistemas que aumenten la interoperabilidad entre los diferentes formatos, formatos para nuevos recursos, nuevos tipos de metadatos, evaluación de la funcionalidad y estudios métricos a través de metadatos, teoría y fundamentos.

“Los bibliotecarios como especialistas en la recuperación y búsqueda de información, y en el comportamiento de los usuarios de información, y con la experiencia adquirida a lo largo de todos estos años, podemos y debemos participar en el desarrollo de recursos electrónicos.”

Relativas a la aplicación al sitio web de LICAD.

- a) Se provee de una herramienta de descripción normalizada de cada uno de los componentes del sitio web de LICAD que permite que el mismo sea relevado automáticamente en buscadores tanto especializados como generales, aumentando su visibilidad notoriamente.
- b) Se enriquece la funcionalidad del sitio web de LICAD con una herramienta de búsqueda que permite la recuperación por distintos puntos de acceso.

- Líneas futuras de investigación:

- Formatos de metadatos para la descripción de objetos digitales: RDA, DC, SCORM
- Lenguajes de marcado, en especial XML
- Desarrollo de sistemas normalizados de gestión de documentos digitales.
- Web semántica.

8. ANEXOS

- A. Código en JavaScript del buscador
- B. Arrays de búsqueda para el Módulo de Procesamiento de la información

A. Código en JavaScript del buscador

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN">
<html>
<head>
<title>Buscador de Metadatos en JS</title>
<meta name="author" content="Andr s Vuotto">
<meta name="author" content="M. Carolina Rojas">
<meta name="keywords" content="buscador, javascript, dublin core, metadatos">
<meta name="description" content="Buscador de Metadatos DC en JavaScript">
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1">
<link rel="stylesheet" type="text/css" href="archivos/estilos.css">
<script language="JavaScript" src="archivos/bd.js"></script>
<script>
<!--
function buscar(a,b){

//definiciyn de la variable que contiene el conector lgico
var bool;
bool=document.miform.bool.options[document.miform.bool.selectedIndex].value;

/*
Los array con los metadatos(o sea la base de datos) se encuentran en el archivo
bd.js guardado en la carpeta archivos
*/

resultado=new Array();

resultadoA=new Array();
resultadoB=new Array();
busquedaA=new Array();
busquedaB=new Array();
result=new Array();
var recup;
recup="";

    for (var i = 0; i<site.length;i++){//este for recorre cada registro y
forma variables con los resultados de bsqueda

        resultadoA[i]=new Array();
```

```

resultadoB[i]=new Array();
//recorrido de la caja de texto 1
if(document.miform.a.checked){
resultadoA[i][0]=site[i][0].indexOf(a,0);
}
else{
resultadoA[i][0]=-1;
}

if(document.miform.b.checked){
resultadoA[i][1]=site[i][1].indexOf(a,0);
}
else{
resultadoA[i][1]=-1;
}

if(document.miform.c.checked){
resultadoA[i][2]=site[i][2].indexOf(a,0);
}
else{
resultadoA[i][2]=-1;
}

if(document.miform.d.checked){
resultadoA[i][3]=site[i][3].indexOf(a,0);
}
else{
resultadoA[i][3]=-1;
}

if(document.miform.e.checked){
resultadoA[i][4]=site[i][4].indexOf(a,0);
}
else{
resultadoA[i][4]=-1;
}

if(document.miform.f.checked){
resultadoA[i][5]=site[i][5].indexOf(a,0);
}
else{
resultadoA[i][5]=-1;
}

```

```

    }

    if (document.miForm.g.checked) {
    resultadoA[i][6]=site[i][6].indexOf(a,0);
    }
    else{
    resultadoA[i][6]=-1;
    }

    if (document.miForm.h.checked) {
    resultadoA[i][7]=site[i][7].indexOf(a,0);
    }
    else{
    resultadoA[i][7]=-1;
    }

    if (document.miForm.i.checked) {
    resultadoA[i][8]=site[i][8].indexOf(a,0);
    }
    else{
    resultadoA[i][8]=-1;
    }

    if (document.miForm.j.checked) {
    resultadoA[i][9]=site[i][9].indexOf(a,0);
    }
    else{
    resultadoA[i][9]=-1;
    }

    if (document.miForm.k.checked) {
    resultadoA[i][10]=site[i][10].indexOf(a,0);
    }
    else{
    resultadoA[i][10]=-1;
    }

    if (document.miForm.l.checked) {
    resultadoA[i][11]=site[i][11].indexOf(a,0);
    }
    else{
    resultadoA[i][11]=-1;

```

```

    }

    if(document.miform.m.checked){
    resultadoA[i][12]=site[i][12].indexOf(a,0);
    }
    else{
    resultadoA[i][12]=-1;
    }

    if(document.miform.n.checked){
    resultadoA[i][13]=site[i][13].indexOf(a,0);
    }
    else{
    resultadoA[i][13]=-1;
    }

    if(document.miform.o.checked){
    resultadoA[i][14]=site[i][14].indexOf(a,0);
    }
    else{
    resultadoA[i][14]=-1;
    }

    if(document.miform.p.checked){
    resultadoA[i][15]=site[i][15].indexOf(a,0);
    }
    else{
    resultadoA[i][15]=-1;
    }

    //recorrido de la caja de texto 2
    if(document.miform.a.checked){
    resultadoB[i][0]=site[i][0].indexOf(b,0);
    }
    else{
    resultadoB[i][0]=-1;
    }

    if(document.miform.b.checked){
    resultadoB[i][1]=site[i][1].indexOf(b,0);
    }
    else{

```

```
resultadoB[i][1]=-1;
}

if(document.miform.c.checked){
resultadoB[i][2]=site[i][2].indexOf(b,0);
}
else{
resultadoB[i][2]=-1;
}

if(document.miform.d.checked){
resultadoB[i][3]=site[i][3].indexOf(b,0);
}
else{
resultadoB[i][3]=-1;
}

if(document.miform.e.checked){
resultadoB[i][4]=site[i][4].indexOf(b,0);
}
else{
resultadoB[i][4]=-1;
}

if(document.miform.f.checked){
resultadoB[i][5]=site[i][5].indexOf(b,0);
}
else{
resultadoB[i][5]=-1;
}

if(document.miform.g.checked){
resultadoB[i][6]=site[i][6].indexOf(b,0);
}
else{
resultadoB[i][6]=-1;
}

if(document.miform.h.checked){
resultadoB[i][7]=site[i][7].indexOf(b,0);
}
else{
```

```

resultadoB[i][7]=-1;
}

if (document.miForm.i.checked) {
resultadoB[i][8]=site[i][8].indexOf(b,0);
}
else{
resultadoB[i][8]=-1;
}

if (document.miForm.j.checked) {
resultadoB[i][9]=site[i][9].indexOf(b,0);
}
else{
resultadoB[i][9]=-1;
}

if (document.miForm.k.checked) {
resultadoB[i][10]=site[i][10].indexOf(b,0);
}
else{
resultadoB[i][10]=-1;
}

if (document.miForm.l.checked) {
resultadoB[i][11]=site[i][11].indexOf(b,0);
}
else{
resultadoB[i][11]=-1;
}

if (document.miForm.m.checked) {
resultadoB[i][12]=site[i][12].indexOf(b,0);
}
else{
resultadoB[i][12]=-1;
}

if (document.miForm.n.checked) {
resultadoB[i][13]=site[i][13].indexOf(b,0);
}
else{

```



```

        resultadoB[i][13]==-1;
    }

    if(document.miform.o.checked){
        resultadoB[i][14]=site[i][14].indexOf(b,0);
    }
    else{
        resultadoB[i][14]==-1;
    }

    if(document.miform.p.checked){
        resultadoB[i][15]=site[i][15].indexOf(b,0);
    }
    else{
        resultadoB[i][15]==-1;
    }

        if(resultadoA[i][0]==-1 && resultadoA[i][1]==-1 &&
resultadoA[i][2]==-1 && resultadoA[i][3]==-1 && resultadoA[i][4]==-1 &&
resultadoA[i][5]==-1 && resultadoA[i][6]==-1 && resultadoA[i][7]==-1 &&
resultadoA[i][8]==-1 && resultadoA[i][9]==-1 && resultadoA[i][10]==-1 &&
resultadoA[i][11]==-1 && resultadoA[i][12]==-1 && resultadoA[i][13]==-1 &&
resultadoA[i][14]==-1 && resultadoA[i][15]==-1){//si no se encontro nada en el
registro segun lo ingresado en la primera caja de texto
        busquedaA[i]="no";
    }
    else{//si se encontro algo
        busquedaA[i]="si";
    }

        if(resultadoB[i][0]==-1 && resultadoB[i][1]==-1 &&
resultadoB[i][2]==-1 && resultadoB[i][3]==-1 && resultadoB[i][4]==-1 &&
resultadoB[i][5]==-1 && resultadoB[i][6]==-1 && resultadoB[i][7]==-1 &&
resultadoB[i][8]==-1 && resultadoB[i][9]==-1 && resultadoB[i][10]==-1 &&
resultadoB[i][11]==-1 && resultadoB[i][12]==-1 && resultadoB[i][13]==-1 &&
resultadoB[i][14]==-1 && resultadoB[i][15]==-1){//si no se encontro nada en el
registro segun lo ingresado en la segunda caja de texto
        busquedaB[i]="no";
    }
    else{//si se encontro algo
        busquedaB[i]="si";
    }

```

```

switch (bool){//resultados de búsquedas según el
conector utilizado, genero las variables con los datos a mostrar en la respuesta

case "and"://busco and
    if(busquedaA[i]=="si" &&
busquedaB[i]=="si"){//evalúo la búsqueda
        result[i]="<div style='text-
align: justify'>"+i+" - "+site[i][0]+site[i][1]+"</a><br>"+site[i][4]+"<br>"+<a
href='"+site[i][10]+'>"+site[i][10]+"</a><br><br></div>";
        }
        else{
            result[i]="&nbsp;";
        }
        break;

case "or"://busco or
    if(busquedaA[i]=="si" ||
busquedaB[i]=="si"){//evalúo la búsqueda
        result[i]="<div style='text-
align: justify'>"+i+" - "+site[i][0]+site[i][1]+"</a><br>"+site[i][4]+"<br>"+<a
href='"+site[i][10]+'>"+site[i][10]+"</a><br><br></div>";
        }
        else{
            result[i]="&nbsp;";
        }
        break;

case "not"://busco not
    if(busquedaA[i]=="si" &&
busquedaB[i]!="si"){//evalúo la búsqueda
        result[i]="<div style='text-
align: justify'>"+i+" - "+site[i][0]+site[i][1]+"</a><br>"+site[i][4]+"<br>"+<a
href='"+site[i][10]+'>"+site[i][10]+"</a><br><br></div>";
        }
        else{
            result[i]="&nbsp;";
        }
        break;
    }
}
}

```

```

for (var j = 0; j<result.length;j++){//genero la variable con los resultados

    if(result[j]=="&nbsp;"){//si vale &nbsp; es porque no encontry nada,
    entonces la igualo a "" para poder hacer el if al final
        result[j]="";
    }

    recup=recup+""+result[j];
}

    if(recup==""){//si estó vacía es porque no se recupery
nada, sino la imprimo en la pantalla
        document.write("<DOCTYPE HTML PUBLIC '-//W3C//DTD HTML
4.01 Transitional//EN'><html><head><title>Buscador de Metadatos en
JS</title><meta name='author' content='Unregistered User'><meta name='keywords'
content=''><meta name='description' content=''><meta http-equiv='Content-Type'
content='text/html; charset=iso-8859-1'><link rel='stylesheet' type='text/css'
href='archivos/estilos.css'>"+ "<body bgcolor='#ffffff' text='#000000'
link='#000099' vlink='#996699' alink='#996699'><Center><font face='Arial'
size='3'><font face=Verdana size=12><b>DCMI</b><i>búsqueda
infructuosa</i></font><table border=0 width=550><tr><td align='center'
valign='middle'><h3><font color=green>no se encontry
<b>''"+document.miform.text.value+" "+bool+"
"+document.miform.text2.value+''</b></font></h3>"+
        "<p>El término -
<b>"+document.miform.text.value+" "+bool+" "+document.miform.text2.value+""</b> -
no se encuentra en la base de datos.</p>"+
        "<LI>Verifique que haya deletreado bien
la palabra.</li>"+
        "<LI>Intente nuevamente utilizando otra
palabra clave.</li>"+
        "<LI>Pruebe ingresando varias palabras en
un mismo campo.</li></span>"+
        "</td></tr></table>")
    }
else{
    document.write("<DOCTYPE HTML PUBLIC '-//W3C//DTD HTML
4.01 Transitional//EN'><html><head><title>Buscador de Metadatos en
JS</title><meta name='author' content='Unregistered User'><meta name='keywords'

```

```

content=''><meta name='description' content=''><meta http-equiv='Content-Type'
content='text/html; charset=iso-8859-1'><link rel='stylesheet' type='text/css'
href='archivos/estilos.css'>"+ "<body bgcolor='#ffffff' text='#000000'
link='#000099' vlink='#996699' alink='#996699'><Center><font face='Arial'
size='3'><font face=Verdana size=12><b>DCMI</b><i>resultados de
búsqueda</i></font><center><h3>Resultados para la fyrmula de búsqueda:
<b>"+document.miform.text.value+" "+bool+"
"+document.miform.text2.value+"</b>.</h3><table border=0 width=550>"+
        "<tr><td align='justify' valign='middle'
class='xresult'>"+recup+"</td></tr></table></Center>");//imprimo los resultados
que estan guardados en la variable recup
        }
}
//-->
</script>

</head>

<body bgcolor="#ffffff" text="#000000" link="#000099" vlink="#996699"
alink="#996699">
<Center><font face="Arial" size="3"><font face=Verdana
size=12><b>DCMI</b><i>búsqueda</i></font>
<table border=0 width=640><tr><td>
  <tr>
    <td align="center" valign="middle">
      <form name="miform" onSubmit="return
buscar(document.miform.text.value,document.miform.text2.value)">
        <table width="380" border="0" cellpadding="2" cellspacing="2">
          <tr bgcolor="#F5F5F5">
            <td align="right" valign="middle"><strong><font size="2" face="Arial,
Helvetica, sans-serif">T&eacute;rmino
              n&deg; 1 </font></strong></td>
            <td align="left" valign="middle"> <strong><font size="2" face="Arial,
Helvetica, sans-serif">
              <input name="text" type="text" size="30">
              </font></strong></td>
          </tr>
          <tr bgcolor="#F5F5F5">
            <td align="right" valign="middle" bgcolor="#FFFFFF">&nbsp;</td>
            <td align="center" valign="middle" bgcolor="#F5F5F5"><strong>

```

```

        <select name="bool" id="bool">
            <option value="and" selected>and</option>
            <option value="or">or</option>
            <option value="not">not</option>
        </select>
    </strong></td>
</tr>
<tr bgcolor="#F5F5F5">
    <td align="right" valign="middle"><strong><font size="2" face="Arial,
Helvetica, sans-serif">T&eacute;rmino
    </font></strong></td>
    <td align="left" valign="middle"> <strong><font size="2" face="Arial,
Helvetica, sans-serif">
        <input name="text2" type="text" id="text23" size="30">
        </font></strong></td>
</tr>
<tr align="center" valign="middle" bgcolor="#F5F5F5">
    <td colspan="2"><strong><font size="2" face="Arial, Helvetica, sans-
serif">Seleccione
        los campos en los que desea buscar </font></strong></td>
</tr>
<tr valign="middle" bgcolor="#FFFFFF">
    <td align="left"> <strong><font size="2" face="Arial, Helvetica,
sans-serif">
        <input name="a" type="checkbox" id="a" checked>
        filename<br>
        <input name="b" type="checkbox" id="b" value="si" checked>
        DC_Title<br>
        <input name="c" type="checkbox" id="c" value="si" checked>
        DC_Creator<br>
        <input name="d" type="checkbox" id="d" value="si" checked>
        DC_Subject<br>
        <input name="e" type="checkbox" id="e" value="si" checked>
        DC_Description<br>
        <input name="f" type="checkbox" id="f" value="si" checked>
        DC_Publisher<br>
        <input name="g" type="checkbox" id="g" value="si" checked>
        DC_Contributor<br>
        <input name="h" type="checkbox" id="h" value="si" checked>
        DC_Date</font></strong></td>
    <td><strong><font size="2" face="Arial, Helvetica, sans-serif">
        <input name="i" type="checkbox" id="i" value="si" checked>

```

```

DC_Type<br>
<input name="j" type="checkbox" id="j" value="si" checked>
DC_Format<br>
<input name="k" type="checkbox" id="k" value="si" checked>
DC_Identifier<br>
<input name="l" type="checkbox" id="l" value="si" checked>
DC_Source<br>
<input name="m" type="checkbox" id="m" value="si" checked>
DC_Language<br>
<input name="n" type="checkbox" id="n" value="si" checked>
DC_Relation<br>
<input name="o" type="checkbox" id="o" value="si" checked>
DC_Coverage<br>
<input name="p" type="checkbox" id="p" value="si" checked>
DC_Rights</font></strong></td>
</tr>
<tr valign="middle" bgcolor="#FFFFFF">
  <td colspan="2" align="center"> <strong><font size="2" face="Arial,
Helvetica, sans-serif">
    <input type="submit" name="Submit" value="buscar">
    </font></strong></td>
</tr>
</table>
</form>
<hr size=1>
</td>
</tr>
</table>
</body>
</html>

```

B. Arrays de búsqueda para el Módulo de Procesamiento de la información

```
/*  
El array site contiene un elemento por cada archivo recuperable del sitio, en  
este caso es necesario indicar entre los paréntesis que definen la cantidad de  
elementos de site la cantidad de archivos indizados que contiene el sitio, que  
debe ser el mismo número de elementos que contiene el array site  
a su vez, cada uno de estos elementos tiene un array con el contenido de los 16  
elementos de metadatos  
0=file,1=título,2=creador,3=keywords,4=descripcion,5=publisher,6=contributor,7=da  
te,8=type,9=format,10=identifier,11=source,12=lenguaje,13=relation,14=coverage,15  
=rights  
*/  
site=new Array();  
  
site[0]=new Array("<a href='../proces/bibliopros.htm'>", "Módulo Procesamiento -  
Bibliografía", "Gustavo Liberatore, Marcela  

```

```
site[2]=new Array("<a href='../proces/bibliopros2.htm'>", "Mydulo Procesamiento - Bibliografha Eje 2", "Gustavo Liberatore, Marcela Coringrato", "tesauros, Madrid, lenguajes documentales, Tesouro sobre Reforma, Universidad, control del vocabulario, Bibliografha Eje, Mydulo Procesamiento, Bibliografha Eje, Mydulo Procesamiento", "Bibliografha del eje 2 de la asignatura Procesamiento de la Informaciyn de la carrera Licenciatura en Bibliotecologha y Documentaciyn a Distancia de la UNMDP.", "Universidad Nacional de Mar del Plata. Departamento de Documentaciyn", "A. Vuotto", "4/30/2002", "text/html; charset=windows-1252", "text/html", "../proces/bibliopros2.htm", "", "Spanish", "", "", "http://www.mdp.edu.ar/humanidades/documentacion/licad/rights.htm");
```

```
site[3]=new Array("<a href='../proces/bibliopros3.htm'>", "Mydulo Procesamiento - Bibliografha Eje 3", "Gustavo Liberatore, Marcela Coringrato", "Madrid, documentaciyn, informaciyn, los, procesamiento, lenguajes de indizaciyn, an6lisis documental, Bibliografha Eje, documentaciyn, informaciyn, lenguajes de indizaciyn, an6lisis documental, Bibliografha Eje", "Bibliografha del eje 3 de la asignatura Procesamiento de la Informaciyn de la carrera Licenciatura en Bibliotecologha y Documentaciyn a Distancia de la UNMDP.", "Universidad Nacional de Mar del Plata. Departamento de Documentaciyn", "A. Vuotto", "4/30/2002", "text/html; charset=windows-1252", "text/html", "../proces/bibliopros3.htm", "", "Spanish", "", "", "http://www.mdp.edu.ar/humanidades/documentacion/licad/rights.htm");
```

```
site[4]=new Array("<a href='../proces/bibliopros4.htm'>", "Mydulo Procesamiento - Bibliografha Eje 4", "Gustavo Liberatore, Marcela Coringrato", "recuperaciyn de informaciyn, b6squeda, Bibliografha Eje, los sistemas, Eyrolles, Barcelona, Motores de b6squeda, Leloup, recuperaciyn de informaciyn, b6squeda, Bibliografha Eje, Motores de b6squeda", "Bibliografha del eje 4 de la asignatura Procesamiento de la Informaciyn de la carrera Licenciatura en Bibliotecologha y Documentaciyn a Distancia de la UNMDP.", "Universidad Nacional de Mar del Plata. Departamento de Documentaciyn", "A. Vuotto", "4/30/2002", "text/html; charset=windows-1252", "text/html", "../proces/bibliopros4.htm", "", "Spanish", "", "", "http://www.mdp.edu.ar/humanidades/documentacion/licad/rights.htm");
```

```
site[5]=new Array("<a href='../proces/buscarpros.htm'>", "Mydulo Procesamiento - Buscador", "A. Vuotto", "", "Buscador de la asignatura Procesamiento de la Informaciyn de la carrera Licenciatura en Bibliotecologha y Documentaciyn a Distancia de la UNMDP.", "Universidad Nacional de Mar del Plata. Departamento de Documentaciyn", "A. Vuotto", "4/30/2002", "text/html; charset=windows-1252", "text/html", "../proces/buscarpros.htm", "", "Spanish", "", "", "http://www.mdp.e
```


du.ar/humanidades/documentacion/licad/rights.htm");

```
site[6]=new Array("<a href='../proces/procesamiento.htm'>", "Mydulo Procesamiento  
- P6gina principal", "Gustavo Liberatore, Marcela Coringrato", "recuperaciyn de  
informaciyn, estrategia de bsqueda, sistemas `de clasificaciyn, informaciyn  
textual, informaciyn referencial, lgica de Boole, lenguaje libre, lenguajes  
documentales, los lenguajes, recuperaciyn de informaciyn, estrategia de  
bsqueda, sistemas de clasificaciyn, informaciyn textual, informaciyn  
referencial, lgica de Boole", "Presentaciyn de la asignatura Procesamiento de la  
Informaciyn de la carrera Licenciatura en Bibliotecologha y Documentaciyn a  
Distancia de la UNMDP.", "Universidad Nacional de Mar del Plata. Departamento de  
Documentaciyn", "A. Vuotto", "4/30/2002", "text/html; charset=windows-  
1252", "text/html", "../proces/procesamiento.htm", "", "Spanish", "", "", "http://www.md  
p.edu.ar/humanidades/documentacion/licad/rights.htm");
```

```
site[7]=new Array("<a href='../proces/progprosec.html'>", "Mydulo Procesamiento -  
Programa", "Gustavo Liberatore, Marcela Coringrato", "trabajo  
pr6ctico, los contenidos, documentaciyn, evaluaciyn, bibliografha, procesamiento, infor  
macion, trabajo  
pr6ctico, documentaciyn, evaluaciyn, bibliografha, informaciyn", "Programa de la  
asignatura Procesamiento de la Informaciyn de la carrera Licenciatura en  
Bibliotecologha y Documentaciyn a Distancia de la UNMDP.", "Universidad Nacional  
de Mar del Plata. Departamento de Documentaciyn", "A.  
Vuotto", "4/30/2002", "text/html; charset=windows-  
1252", "text/html", "../proces/progprosec.html", "", "Spanish", "", "", "http://www.mdp.  
edu.ar/humanidades/documentacion/licad/rights.htm");
```

```
site[8]=new Array("<a href='../proces/prosejel.htm'>", "Mydulo Procesamiento -  
Ejel: El tratamiento documental", "Gustavo Liberatore, Marcela  
Coringrato", "documentaciyn, ling7stica documental, an6lisis  
documental, procesamiento de la informaciyn, tratamiento documental, an6lisis  
formal, an6lisis de contenido, catalogaciyn, indizaciyn, resumen, asiento  
bibliogr6fico, clasificaciyn, cat6logo, ndice", "Eje 1 de la asignatura  
Procesamiento de la Informaciyn de la carrera Licenciatura en Bibliotecologha y  
Documentaciyn a Distancia de la UNMDP.", "Universidad Nacional de Mar del Plata.  
Departamento de Documentaciyn", "A. Vuotto", "4/30/2002", "text/html;  
charset=windows-  
1252", "text/html", "../proces/prosejel.htm", "", "Spanish", "", "", "http://www.mdp.edu  
.ar/humanidades/documentacion/licad/rights.htm");
```

```
site[9]=new Array("<a href='../proces/proseje2.htm'>", "Mydulo Procesamiento - Eje  
2: Los lenguajes documentales", "Gustavo Liberatore, Marcela
```

Coringrato", "lenguajes documentales, tesauros, ciclo documental, indizaciyn, lenguaje natural", "Eje 2 de la asignatura Procesamiento de la Informaciyn de la carrera Licenciatura en Bibliotecologha y Documentaciyn a Distancia de la UNMDP.", "Universidad Nacional de Mar del Plata. Departamento de Documentaciyn", "A. Vuotto", "4/30/2002", "text/html; charset=windows-1252", "text/html", "../proces/proseje2.htm", "", "Spanish", "", "", "http://www.mdp.edu.ar/humanidades/documentacion/licad/rights.htm");

site[10]=new Array("", "Mydulo Procesamiento - Eje 3: Indizaciyn y resumen", "Gustavo Liberatore, Marcela Coringrato", "palabras clave, indizaciyn, descriptores, tesauros, indizaciyn autom6tica, lenguaje controlado, lenguaje natural, indizaciyn humana, indizaciyn exhaustiva, resumen, concepto, traducciy documenta", "Eje 3 de la asignatura Procesamiento de la Informaciyn de la carrera Licenciatura en Bibliotecologha y Documentaciyn a Distancia de la UNMDP.", "Universidad Nacional de Mar del Plata. Departamento de Documentaciyn", "A. Vuotto", "4/30/2002", "text/html; charset=windows-1252", "text/html", "../proces/proseje3.htm", "", "Spanish", "", "", "http://www.mdp.edu.ar/humanidades/documentacion/licad/rights.htm"); ,

site[11]=new Array("", "Mydulo Procesamiento - Eje 4: Búsqueda y recuperaciyn de informaciyn", "Gustavo Liberatore, Marcela Coringrato", "búsqueda de informaciyn, recuperaciyn de informaciyn, operadores booleanos, motores de búsqueda, directorios, operadores sint6cticos, operadores de proximidad, internet, bases de datos, buscadores, ecuaciyn de búsqueda, estrategia de búsqueda, ndices", "Eje 4 de la asignatura Procesamiento de la Informaciyn de la carrera Licenciatura en Bibliotecologha y Documentaciyn a Distancia de la UNMDP.", "Universidad Nacional de Mar del Plata. Departamento de Documentaciyn", "A. Vuotto", "4/30/2002", "text/html; charset=windows-1252", "text/html", "../proces/proseje4.htm", "", "Spanish", "", "", "http://www.mdp.edu.ar/humanidades/documentacion?/licad/rights.htm");

site[12]=new Array("", "Mydulo Procesamiento - Trabajos pr6cticos", "Gustavo Liberatore, Marcela Coringrato", "tratamiento documental, an6lisis documental, lenguajes documentales, tesauros, ling6stica documental, indizaciyn, búsqueda de informaciyn, recuperaciyn de informaciyn", "Trabajos pr6cticos de la asignatura Procesamiento de la Informaciyn de la carrera Licenciatura en Bibliotecologha y Documentaciyn a Distancia de la UNMDP.", "Universidad Nacional de Mar del Plata. Departamento de Documentaciyn", "A. Vuotto", "4/30/2002", "text/html; charset=iso-8859-1", "text/html", "../proces/tpprosec.htm", "", "Spanish", "", "", "http://www.mdp.edu.ar/humanidades/documentacion/licad/rights.htm");

```
site[13]=new Array("<a href='../proces/whopros.htm'>", "Mydulo Procesamiento -  
Curricula vitarum", "Gustavo Liberatore, Marcela Coringrato", "curriculum vitae,  
marcela coringrato, gustavo liberatore", "Curricula vitarum de los profesores de  
la asignatura Procesamiento de la Informaciyn de la carrera Licenciatura en  
Bibliotecologna y Documentaciyn a Distancia de la UNMDP.", "Universidad Nacional  
de Mar del Plata. Departamento de Documentaciyn", "A.  
Vuotto", "4/30/2002", "text/html; charset=windows-  
1252", "text/html", "../proces/whopros.htm", "", "Spanish", "", "", "http://www.mdp.edu.  
ar/humanidades/documentacion/licad/rights.htm");
```

```
site[14]=new Array("0-file", "1-titulo", "2-creador", "3-palabrasclave", "4-  
descripcion", "5-publisher", "6-contributor", "7-date", "8-type", "9-format", "10-  
identifier", "11-source", "12-lenguaje", "13-relation", "14-coverage", "15-rights");
```

9. Glosario

- ✓ **Aplicación de metadatos:** proceso de incorporación de elementos de descripción a un documento electrónico.

- ✓ **Array:** En programación, una **matriz** o **vector** (llamados en inglés **arrays**) es una zona de almacenamiento contiguo, que contiene una serie de elementos del mismo tipo, los elementos de la matriz. Desde el punto de vista lógico una matriz se puede ver como un conjunto de elementos ordenados en fila (o filas y columnas si tuviera dos dimensiones). En principio, se puede considerar que todas las matrices son de una dimensión, la dimensión principal, pero los elementos de dicha fila pueden ser a su vez matrices (un proceso que puede ser recursivo), lo que nos permite hablar de la existencia de matrices multidimensionales, aunque las más fáciles de imaginar son los de una, dos y tres dimensiones. Estas estructuras de datos son adecuadas para situaciones en las que el acceso a los datos se realice de forma aleatoria e impredecible. Por el contrario, si los elementos pueden estar ordenados y se va a utilizar acceso secuencial sería más adecuado utilizar una lista, ya que esta estructura puede cambiar de tamaño fácilmente durante la ejecución de un programa.

- ✓ **Base de datos:** Una colección de información que existe durante un periodo largo de tiempo. En informática, con la expresión *base de datos* se designa a una colección de datos estructurada que es administrada por un sistema de administración de bases de datos. Un sistema de base de datos consiste en un conjunto de datos cuya manipulación está a cargo del Sistema Manejador de Base de Datos (Database Management System, DBMS), el cual está constituido, a su vez, por tres partes:
 - a. Un lenguaje de definición de datos (Data Definition Language, DDL), cuya función es la definición de la estructuración general de los datos;
 - b. Un lenguaje de manipulación (Database Manipulation Language, DML) que permite cambiar y actualizar los datos y,
 - c. Un lenguaje de consulta (Database Query Language, DQL).

- ✓ **Descripción de información:** suministro de información representativa acerca de un documento, de forma que permita discernir sobre su utilidad.
- ✓ **Documento electrónico:** registro de información legible por computadora, que puede estar almacenado en uno o varios archivos.
- ✓ **HTML (Hypertext Mark-up Language):** conjunto de etiquetas y normas derivados de SGML para crear documentos de hipertexto para la World Wide Web. Oficialmente, una Recomendación del W3C.
- ✓ **Hipertexto:** Representación de la información en unidades que permite que el usuario, a través de asociaciones, se conecte con otras unidades de información. Cada instancia de asociación es llamada link (enlace).
- ✓ **Internet:** Red de computadoras mundial que permite comunicación y transferencia de datos, noticias y opiniones entre máquinas y usuarios conectadas a ella, a través de un sistema común de direcciones y comunicación llamado TCP/IP (Transmission Control Protocol / Internet Protocol). Los servicios más utilizados actualmente son: World Wide Web (Hyper Text Transfer Protocol [HTTP]), Correo Electrónico (Simple Mail Transfer Protocol [SMTP]), Transferencia de archivos (File Transfer Protocol [FTP]),
- ✓ **Interoperabilidad:** Capacidad para intercambiar información de modo preciso que tienen múltiples sistemas con diferente hardware, plataformas de software, estructuras de datos e interfaces. Se pueden definir cuatro capas de interoperabilidad, que enumeradas desde un nivel más concreto hacia el más abstracto son: técnica/básica, sintáctica, funcional/pragmática y semántica.
- ✓ **Lenguaje de marcado:** forma de codificar un documento que, junto con el texto, incorpora etiquetas o marcas que contienen información adicional acerca de la estructura del texto o su presentación. Se pueden distinguir tres clases:

- ✓ El **marcado de presentación** es aquel que indica el formato del texto. Este tipo de marcado es útil para maquetar la presentación de un documento para su lectura, pero resulta insuficiente para el procesamiento automático de la información.
- ✓ El **marcado de procedimientos** está enfocado hacia la presentación del texto, sin embargo, también es visible para el usuario que edita el texto. El programa que representa el documento debe interpretar el código en el mismo orden en que aparece. Por ejemplo, para formatear un título, debe haber una serie de directivas inmediatamente antes del texto en cuestión, indicándole al software instrucciones tales como centrar, aumentar el tamaño de la fuente, o cambiar a negrita. Inmediatamente después del título deberá haber etiquetas inversas que reviertan estos efectos.
- ✓ El **marcado descriptivo o semántico** utiliza etiquetas para describir los fragmentos de texto, pero sin especificar cómo deben ser representados, o en que orden. Los lenguajes expresamente diseñados para generar marcado descriptivo son el SGML y el XML.

- ✓ **Metadatos:** información acerca de la información; elementos de identificación, descripción y localización de recursos electrónicos.
- ✓ **Multiplataforma:** servicios de software que permiten su operatividad en diversas plataformas.
- ✓ **Organización de la información:** conjunto de técnicas y habilidades complejas, que permiten gestionar la interrelación entre la oferta de servicios de información y los contenidos a usuarios en contextos determinados.
- ✓ **Parseo:** Proceso de analizar una secuencia de símbolos a fin de determinar su estructura gramatical con respecto a una gramática formal dada. Formalmente es llamado análisis de sintaxis. Un parseador (parser) es un programa de computación que lleva a cabo esta área. El parseo transforma una entrada de texto en una estructura de datos (usualmente un árbol)

que es apropiada para ser procesada. Generalmente los parseadores primero identifican los símbolos de la entrada y luego construyen el árbol de parseo para esos símbolos.

- ✓ **Plataforma:** En informática, determinado software y/o hardware con el cual una aplicación es compatible y permite ejecutarla. Una plataforma es, por ejemplo, un sistema operativo, un gran software que sirve como base para ejecutar determinadas aplicaciones compatibles con este. También son plataformas la arquitectura de hardware, los lenguajes de programación y sus librerías en tiempo de ejecución, las consolas de videojuegos, etc.
- ✓ **Recuperación de información:** procedimiento que permite la localización de documentos pertinentes a una inquietud determinada. Podría decirse que consta de las siguientes etapas: definición del problema de información, determinación de una estrategia de búsqueda y de fuentes prioritarias, localización y acceso de las fuentes de información, uso de la información, síntesis y evaluación de los resultados.
- ✓ **Sitio web:** conjunto de páginas que pertenecen a un dominio determinado.
- ✓ **WYSIWYG (What You See Is What You Get, lo que ves es lo que obtienes):** aplicación que permite ver el documento en su presentación final al momento de su procesamiento/creación, al posibilitar que cualquier usuario produzca y publique contenidos, sin que estos necesariamente cumplan con normas bien definidas.
- ✓ **World Wide Web:** Servicio de Internet formado por documentos HTML interconectados entre sí y distribuidos entre servidores en el mundo entero, que utiliza el protocolo de transferencia de hipertexto (HTTP).
- ✓ **XML (*Extensible Markup Language*):** un perfil de aplicación de SGML diseñado para su uso en aplicaciones Web. Oficialmente, una Recomendación del W3C. XML fue concebido como un medio para recuperar la potencia y flexibilidad de SGML, eliminando la mayor parte de su complejidad. Aunque es una forma restringida de SGML, XML conserva la mayor parte de su potencial y riqueza, manteniendo todas las características de uso general del SGML.

✓ **XHTML (eXtensible Hypertext Markup Language):** Se trata de la reformulación de los documentos de tipo HTML 4 como aplicaciones de XML 1.0. Su propósito es ser utilizado como un lenguaje de contenido que es a la vez compatible con XML y trabajar con los agentes de usuario en el cumplimiento de HTML 4, si se siguen algunas simples directrices. Los desarrolladores que migren sus aplicaciones a XHTML 1.0 se darán cuenta de los siguientes ventajas:

- ✓ Los documentos XHTML se ajustan a XML. Como tales, son fáciles de visualizar, editar, y pueden ser validados con herramientas XML estándar.
- ✓ Los documentos XHTML pueden escribirse para trabajar igual o mejor que antes con los agentes de usuario que se ajusten a HTML 4, así como en las aplicaciones de usuario nuevas que se ajusten a XHTML 1.0.
- ✓ Los documentos XHTML pueden utilizar las aplicaciones (por ejemplo, scripts y applets) que se basan ya sea en el modelo de objetos de documento HTML o el XML Document Object Model [DOM].

A medida que la familia XHTML evolucione, los documentos conformes a XHTML 1.0 estarán más dispuestos a interactuar dentro de y entre los distintos entornos XHTML.

10. Bibliografía

- ✓ Baca, Murtha. Introduction to Metadata : Pathways to Digital Information. Los Ángeles: Getty Research Institute, mayo 2000. [Versión HTML] Disponible en: <http://www.getty.edu/research/institute/standards/intrometadata/pdf/swetland.pdf>
- ✓ Berners-Lee, Tim. Metadata architecture: document, metadata and links. [Versión HTML]. En Design Issues. W3C, 6/ENE/1997. Disponible en: <http://www.w3.org./DesignIssues/Metadata.html>.
- ✓ Berners-Lee, Tim. Tejiendo la red: el inventor del World Wide Web nos descubre su origen. Madrid: Siglo XXI, 2000. 237 p.
- ✓ Berners-Lee, Tim. The Evolution of a specification : Commentary on Web architecture EN: The Evolution of a specification – Commentary on Web architecture.htm , Marzo.1998.
- ✓ Bodoff, D.; Hung, P. C. K. y Ben-Menachem, M. Web metadata standards: observations and prescriptions. EN IEEE Software, Ene/Feb. 2005, vol. 22, no. 1, pp. 78-85.
- ✓ Bosch, Mela. Documentos y lenguaje de marcado: conceptos, problemas y tendencias. En: El profesional de la información, Barcelona, nov. 2001, v. 10, n. 11, pp. 4-9.
- ✓ Bosch, Mela. Qué es una base de datos, modelos de las bases de datos, el modelado de las bases de datos. En: Seminario B - LICAD. Mar del Plata: UNMDP, 2003.
- ✓ Burns, Joe y Growney, Andree. Descubre JavaScript. Madrid: Pearson Educación, 2000. 392 p.
- ✓ Dempsey, Lorcan and Heery, Rachel. A review of metadata: a survey of current resource description formats. Work Package 3 of Telematics for Research project DESIRE (no. 1004). URL: <http://www.ukoln.ac.uk/metadata/desire/overview/> PDF: <http://www.ifla.org/documents/libraries/cataloging/metadata/d32p1.pdf> 1997

- ✓ Dempsey, Lorcan y Rachel Heery. Metadata: a current view of practice and issues. En: *Journal of documentation*, Mar 1998, vol. 54, N°2. Pp. 145-172
- ✓ Dertouzos, Michael L. *La revolución incompleta : las computadoras centradas en el hombre y qué pueden hacer por nosotros*. Buenos Aires : Fondo de Cultura Económica, 2003. 220 p.
- ✓ Diché, Hill. *E-Data : transformando datos en información con Data warehousing – 1ª. Ed. –* Buenos Aires : Prentice Hall, 2001. 374 p.
- ✓ Frentzen, Jeff; Sobotka, Henry. *Superutilidades para JavaScript*. Madrid: Osborne - McGraw-Hill, 1999. 549 p.
- ✓ Galea, J., Romero, M., Rallo, R. Los procesos de Estandarización de Contenidos para la Formación. Actas de EDUTEC'01, Murcia, España, 2001. EN: <http://www.etse.urv.es/~rrallo/papers/edutec.pdf>. [Visitado en Oct. 2005]
- ✓ García Martínez, Ana María. Definición y estilos de los objetos de información digitales y metadatos para la descripción. En: *Boletín de la Asociación Andaluza dfe Bibliotecarios*, Jun 2001, vol. 16, N°63. Pp. 23-47.
- ✓ Gilliland-Swetland, Anne J.. Setting the Stage. EN: *Introduction to Metadata : Pathways to Digital Information*. Baca, Murtha (Ed.). Los Ángeles: Getty Research Institute, mayo 2000. [Versión HTML]. Disponible en: <http://www.getty.edu/research/institute/standards/intrometadata/pdf/swetland.pdf>.
- ✓ Gilliland-Swetland, Anne J. La definición de los metadatos. En: *Introducción a los metadatos: vías a la información digital*. Murtha Baca (ed.) Los Ángeles: Getty Institute Trust, 1999. pp. 1-9.

- ✓ Greenberg J., Spurgin, K., Crystal, A., Cronquist, M., Wilson, A.. Final Report for the AMeGA (Automatic Metadata Generation Applications) Project, UNC School of information and library science, 2005. http://www.loc.gov/catdir/bibcontrol/lc_amega_final_report.pdf.
- ✓ Greenberg, Jane. Metadata Generation : Processes, People and Tools. En: Bulletin of the American Society for Information Science and Technology, December/January 2003. pp. 16-19.
- ✓ Heery, R. Metadata Formats. BIBLINK Study of Metadata. URL: <http://www.ukoln.ac.uk/metadata/review.html> PDF: <http://www.ifla.org/documents/libraries/cataloging/metadata/biblink1.pdf>
- ✓ Heery, R. Resource description: initial recommendations for metadata formats. Peer review draft of deliverable for Work Package 3 of Telematics for Research project DESIRE). July 1996. URL: <http://www.ukoln.ac.uk/metadata/desire/recommendations/>
- ✓ Heery, R. Review of metadata formats. Pre-publication draft of article published in: Program, Vol. 30, no.4, October 1996, pp. 345-373. URL: <http://www.ukoln.ac.uk/metadata/review.html>
- ✓ Heery, Rachel. D1.1 Metadata Formats. [PDF], WP Study of Metadata, N°1, Bath: University, UKOLN, 1996, disponible en: <http://hosted.ukoln.ac.uk/biblink/wp1/d1.1.pdf>
- ✓ Hodge, Gail. Metadata Made Simpler: A guide for libraries. Bethesda: NISO Press, 2001.
- ✓ Introducción a los metadatos: vías a la información digital. Murtha Baca (ed.) Los Ángeles: Getty Institute Trust, 1999.
- ✓ Introduction to Metadata : Pathways to Digital Information. Baca, Murtha (Ed.). Los Ángeles: Getty Research Institute, mayo 2000. [Versión HTML] Disponible en: <http://www.getty.edu/research/institute/standards/intrometadata/pdf/swetland.pdf>

- ✓ Kwong Bor Ng, Soyeon Park; Burnett, Kathleen. Control or Management: A Comparison of the Two Approaches for Establishing Metadata Schemes in the Digital Environment. 1997. [Versión HTML] Disponible en: <http://www.scils.rutgers.edu/~sympark/asis.html> .
- ✓ La sociedad de la información: política, tecnología e industria de los contenidos. Mercedes Caridad Sebastián (Coord.). Madrid: Fundación Ramón Areces, 1999.
- ✓ Landow, George P. Hipertexto: la convergencia de la teoría crítica contemporánea y la tecnología. Barcelona: Paidós, 1995.
- ✓ Leloup, Catherine. Motores de búsqueda e indexación: entornos cliente servidor, internet e intranet. Barcelona: Ediciones Gestión 2000, 1998.
- ✓ Lupovici, Catherine. Identification des ressources sur Internet et métadonnées: diversité des standards. EN Documentaliste – Sciences de l'information, 1999, vol. 36, N°6, p. 221-325.
- ✓ Méndez Rodríguez, Eva y Senso, José A. Introducción a los metadatos ; estándares y aplicación. Madrid : Sedic, 2004. EN <http://www.sedic.es/autoformacion/metadatos/index.htm>.
- ✓ Méndez Rodríguez, Eva. Metadatos y recuperación de información : estándares, problemas y aplicabilidad en bibliotecas digitales. Gijón : Ediciones Trea, 2002. 419 p.
- ✓ Méndez Rodríguez, Eva; Merlo, José A. Localización, identificación y descripción de documentos web: tentativas hacia la normalización. (Jornadas Españolas de Documentación, Bilbao, Jul/2000). EN: La gestión del conocimiento: retos y soluciones de los profesionales de la información. Bilbao: Universidad del País Vasco, 2000. Pp. 221-231.
- ✓ Nielsen, Jakob. Usabilidad: diseño de sitios web. Madrid: Pearson Educación, 2000.

- ✓ Noufal, P.. Metadata: Automatic generation and extraction. In *7th MANLIBNET Annual National Convention on Digital Libraries in Knowledge Management: Opportunities for Management Libraries*. Indian Institute of Management Kozhikode, May 2005.

- ✓ Search Engine Watch. <http://www.searchenginewatch.com/reports/directories.html>. [Visitado en Dic. 2002]

- ✓ Sha, Vianne T. Cataloguing internet resources: the library approach. En *The electronic Library*, oct. 2005, Vol. 13, N°5, pp. 467-476.

- ✓ Siri, Laura. *Internet: búsquedas y buscadores*. Buenos Aires: Norma, 2000.

- ✓ Sorlí Rojo, Ángela y Merlo Vega, José Antonio. Organizaciones de normalización en Internet. *Revista Española de Documentación Científica*, jul.-sept. 2000, vol. 23, n. 2, p. 327-340.

- ✓ Tutorial de Digitalización de imágenes – Metadatos. Departamento de preservación y conservación, Biblioteca de la Universidad de Cornell. En: <http://www.library.cornell.edu/preservation/tutorial-spanish/metadata/table5-1.html> [Visitado en Nov. 2000]

- ✓ Ullman, Jeffrey D. y Widom, Jennifer. *Introducción a los sistemas de bases de datos*. México: Prentice Hall, 1999. 470 p.

- ✓ *Understanding Metadata*. Bethesda: NISO Press, 2004.

- ✓ Vellucci, Sherry L. Metadata. En: *Annual Review of Information Science and Technology (ARIST)*, 1998, vol. 33. Pp. 187-222.

- ✓ Wason, Thomas D.; Neely, Gary. *Dr. Tom's Guide to a simple path to Meta-Data implementation*. IMS : Indiana, 2000. EN: <http://wason.home.mindspring.com/TDW/drtomimplement.html>. [Visitado en Oct. 2005]

- ✓ Weibel, Stuart y Traugott Koch. The Dublin Core Metadata Initiative: mission, current activities and future directions. [Versión HTML]. En: D-Lib Magazine, Nov/Dic 2000, vol. 6, N°12. Disponible en: <http://www.dlib.org/dlinb/december00/weibel/12weibel.html>.
- ✓ Weibel, Stuart. Metadata: the foundations of resource description. [Versión HTML] En: D-Lib Magazine, Jul/Ago 1995, vol. 1, N°7/8. Disponible en: <http://www.dlib.org/dlib/July95/07weibel.html>.
- ✓ Weibel, Stuart. RE: metadata schemes. EN: OCLC Office of Research Listserv Archives List_Serv: Metamarda-L METAMARDA-L@cornell.edu, <http://orc.dev.oclc.org:5103/metamarda-l/msg00104.html>. [99/04/01] Copyright © 1998 OCLC Online Computer Library Center, Inc.