

La organización del conocimiento en Internet

Pablo Lara Navarra y José Ángel Martínez Usero

Diseño de la colección: Editorial UOC
Diseño del libro y de la cubierta: Natàlia Serrano

Primera edición digital: marzo 2018

© Pablo Lara Navarra y José Ángel Martínez Usero, del texto
© Editorial UOC, de la edición
© Editorial UOC (Oberta UOC Publishing, SL) de esta edición, 2018
Rambla del Poblenou, 156, 08018 Barcelona
<http://www.editorialuoc.com>

Realización editorial: Oberta UOC Publishing, SL

Esta obra está sujeta –si no se indica lo contrario– a una licencia Creative Commons de Reconocimiento-No Comercial-Sin obra derivada 3.0 España. Puede copiar, distribuir y comunicar públicamente, siempre y cuando reconozca los créditos de las obras (autoría, Editorial UOC) de la manera especificada por los autores y la Editorial que la publica. No puede hacer uso comercial ni obra derivada sin el permiso del Editor y de los autores. La licencia completa se puede consultar en <https://creativecommons.org/licenses/by-nc-nd/3.0/es/deed.es>

Nuestro contrato

Este libro le interesará si quiere saber:

- Cúal es la diferencia entre datos, información y conocimiento.
- Cuáles son las principales características de la gestión del conocimiento electrónico y de la sociedad del conocimiento.
- Cómo utilizar los lenguajes de marcado para representar el conocimiento en los servicios web.
- Qué es la arquitectura de la información web y qué herramientas y vocabularios controlados se pueden utilizar para mejorar los sistemas de navegación y recuperación de la información electrónica.
- Qué son los metadatos, cómo se codifican y cuáles son sus principales aplicaciones.

Índex de continguts

Nuestro contrato	3
1. Introducción	7
2. La organización del conocimiento electrónico	9
2.1. Los datos, la información y el conocimiento	9
2.2. De la información al conocimiento	11
2.3. La gestión del conocimiento	16
2.4. La sociedad del conocimiento	17
3. Los lenguajes de marcado en la organización del conocimiento	21
3.1. La familia de normas SGML/XML	22
3.2. Las DTD de SGML	24
3.3. El lenguaje XML	27
3.4. Mecanismos para representación del conocimiento	29
3.5. Servicios de información web basados en XML	31
3.6. XML en la organización del conocimiento interno	33

4. Arquitectura de la información, navegación y recuperación	37
4.1. La arquitectura de la información	38
4.2. Sistemas de clasificación del conocimiento electrónico	41
4.3. Navegación y recuperación de la información web	45
5. Los metadatos y la recuperación del conocimiento	47
5.1. La función de los metadatos	47
5.2. Métodos para registrar y transferir metadatos	49
5.3. Las ventajas de RDF	57
Bibliografía	59

1. Introducción

En este libro se aclaran los conceptos de dato, información y conocimiento. Se especifican los tipos de conocimiento existente. Se introducen las características básicas de la gestión del conocimiento y se aclara la influencia de la sociedad del conocimiento en la producción de contenidos electrónicos y la necesidad de representar el conocimiento de éstos.

A continuación se analizan las posibilidades que ofrecen los lenguajes de marcado para la clasificación y organización del conocimiento. Se describen los tipos de lenguajes de marcado para estructurar la información. Se presenta la familia de normas SGML/XML y las relaciones con otros lenguajes de marcado. Se analizan las posibilidades de los lenguajes de marcado en la representación del conocimiento, para más tarde presentar algunos ejemplos de utilización de XML para el desarrollo de servicios de información y la propia gestión del conocimiento organizacional.

En el siguiente apartado se tratan las principales herramientas y vocabularios estructurados para organización del conocimiento electrónico: listas de cate-

gorías, taxonomías, tesauros, clasificaciones facetadas y ontologías. Se destaca la función de los vocabularios estructurados para la organización de contenidos, el fomento de la interoperabilidad de la información y el desarrollo de sistemas de navegación y recuperación de la información en los servicios electrónicos.

Finalmente, se expone la necesidad de incorporación de metadatos en los recursos de información web como método para facilitar la interoperabilidad de los datos y la integración de datos heterogéneos, así como mejorar la recuperación de la información. Se destaca el modelo de metadatos Dublin Core y se especifican los métodos para registrar y transferir metadatos: HTML, XML, RDF, y ciertos repositorios como bases de datos y los sistemas de gestión de contenidos web.

2. La organización del conocimiento electrónico

2.1. Los datos, la información y el conocimiento

Los datos son los resultados de observaciones humanas sobre el estado del mundo y están asociados a un objeto o hecho concreto. Esto es, los datos son expresiones mínimas de información, que aisladas no tienen sentido en sí mismas, pero que adquieren valor dentro de un contexto determinado. Los datos poseen tres características esenciales.

- En primer lugar, pueden ser identificados con claridad, sin posibilidad de confusión, por conjuntos de símbolos.
- En segundo lugar, tienen un nivel elevado de estructura, es decir, la posibilidad de error en la comunicación entre emisor y receptor es mínima.
- Finalmente, las tecnologías de la información son muy útiles cuando se aplican al procesamiento de los datos.

La información está compuesta de datos organizados, agrupados o clasificados en categorías que

les dotan de significado. Además, la información está asociada a un contexto que facilita su interpretación. Por tanto, información son datos elaborados con un significado para el receptor. De esta forma, la información puede entenderse como la significación que adquieren los datos como resultado de un proceso consciente e intencional de adecuación de tres elementos los datos del entorno, los propósitos y el contexto de aplicación, así como la estructura de conocimiento del sujeto.

El conocimiento es la información interpretada, personalizada, que tiene valor y que está orientada a la acción, esto es, que propicia la toma adecuada de decisiones. El conocimiento está asociado al individuo y a la acción, constituyendo un instrumento para la toma de decisiones en el marco de una organización.

Se pueden identificar dos tipos de conocimiento en una organización, y que es precisamente la conversión constante de un tipo al otro lo que explica la generación de nuevo conocimiento.

El **conocimiento tácito** es aquel que difícilmente puede ser formalizado, expresado en palabras (la destreza de un músico, la inspiración de un artista, y otras), así como las habilidades personales más simples que son muy difíciles de transmitir de una a otra persona (utilizar el ordenador, montar en bicicleta, conducir un coche, y otras). Por tanto, el conocimiento tácito reside individualmente en cada persona.

Las organizaciones disponen de conocimiento tácito "encapsulado" en forma de personas y difícilmente se puede almacenar en otras formas.

El **conocimiento explícito** es aquel que se puede expresar en forma de palabras y números, y que es compartido entre personas a través de datos, fórmulas, especificaciones, etc. (un programa de ordenador, una patente, un catálogo, un manual, etc.). Por tanto, el conocimiento explícito es más fácilmente transmisible entre personas. Las organizaciones disponen de grandes cantidades de conocimiento explícito, en forma de procesos estandarizados, procedimientos, manuales, guías, y otros documentos.

2.2. De la información al conocimiento

Las diferencias esenciales entre datos e información se pueden sintetizar en dos. En primer lugar, la información es sólo aquello que el usuario considera que le informa; los datos, aunque sean relevantes, no le aportan nada nuevo o nada que se pueda predecir a partir de lo que ya sabía, el usuario no los considera información. En segundo lugar, las tecnologías de la información pueden manejar muy bien los datos, pero convertir los datos en información es un proceso básicamente humano, todavía no puede automatizarse.

La diferencia esencial entre información y conocimiento radica en que el conocimiento se refiere a

la habilidad de entender el contexto y descubrir las conexiones y el significado de la información. Esto es, información y conocimiento son en principio conceptos diferentes, pero no puede existir el uno sin el otro. La información no tiene valor por sí misma, su valor reside en la habilidad de un individuo para aportarle significado y crear nuevo conocimiento. El conocimiento, de igual forma, carece de valor por sí mismo, el valor aparece cuando el conocimiento se utiliza para solucionar un problema o generar nueva información. El conocimiento sería la experiencia y capacidad de los individuos, unidas a los procesos y a la memoria corporativa; y la información sería la materia prima que el conocimiento utiliza para generar nuevo conocimiento.

En el siguiente gráfico se representan las interrelaciones existentes entre los conceptos de dato, información y conocimiento.

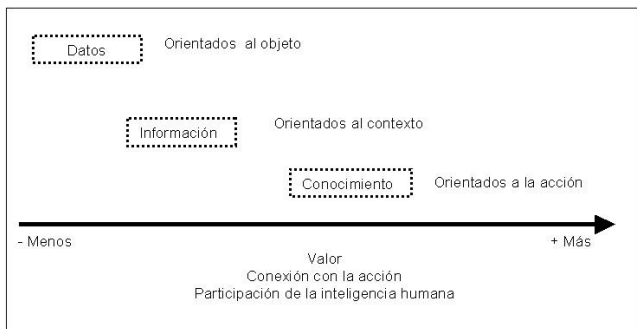


Figura 1. Relaciones entre datos, información y conocimiento

En un sentido genérico y siguiendo la tendencia en la evolución cualitativa del trinomio datos, información y conocimiento, se puede construir un modelo piramidal de esta evolución. Los datos, que son entidades independientes y sin significado intrínseco se convierten en información cuando son contextualizados (se conoce el propósito de los datos recogidos), categorizados (se conocen las unidades de análisis o los componentes clave de los datos), calculados (los datos son analizados matemática o estadísticamente), corregidos (se eliminan los errores existentes en los datos), o bien, condensados (los datos se resumen en una forma más concisa). Por tanto, usando uno o más de los cinco criterios anteriores, los datos se convierten en información.

De igual forma, la información puede transformarse en conocimiento cuando se producen una o varias de las siguientes situaciones: comparación (la información se compara con otras informaciones ya conocidas), consecuencia (la información supone ciertas implicaciones para la toma de decisiones o la acción), conexión (la información está o no relacionada con otros conocimientos), y conversación (la opinión de otras personas sobre la información). Se trata, por tanto, de una progresión lineal donde los datos se convierten en información y ésta en conocimiento mediante un proceso de agregación de valor, que se puede representar mediante un modelo piramidal.

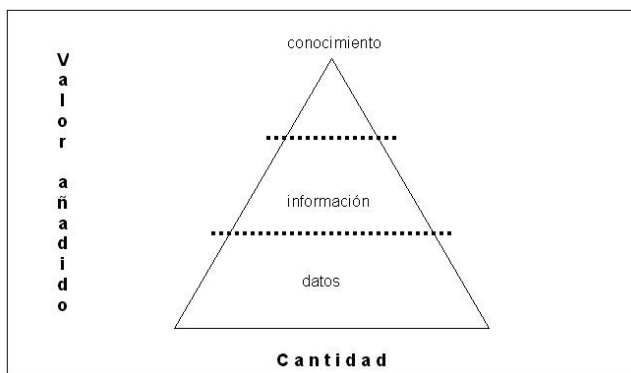


Figura 2. Modelo piramidal de la información

El conocimiento es la información internalizada, esto es, integrada en las estructuras cognitivas de la persona, por tanto, si no hay persona, no hay conocimiento. Pero, en los últimos años, con la influencia de las tecnologías de la información ha surgido un nuevo punto de vista que considera que el conocimiento puede ser generado mediante la información almacenada en la memoria humana o en la memoria artificial (electrónica).

Cuando el conocimiento se produce a partir del procesamiento de la información contenida en la memoria humana, se obtiene el denominado **conocimiento simbólico**.

Cuando el conocimiento se produce a partir del procesamiento de la información codificada y organizada documentalmente (técnicas de almacenamiento, procesamiento y recuperación de la información), se obtiene el **conocimiento figurativo**.

Un paso más allá, partiendo del conocimiento figurativo, cuando generamos nueva información por medio de técnicas de inteligencia artificial, se obtiene nuevo conocimiento, denominado **conocimiento productivo**, que podría ser descrito como un tipo de conocimiento artificial o información electrónica útil.

2.3. La gestión del conocimiento

La gestión del conocimiento se ha considerado desde una simple moda hasta que es el principio de la solución a todos los problemas de incompetencia organizacional. El análisis del concepto de gestión del conocimiento se inicia con el estudio de los tipos de conocimiento que se pueden dar en una organización. Una clasificación tradicional diferencia el conocimiento por su facilidad de comunicación y explotación. De esta forma, en una organización existen dos tipos de conocimiento: el conocimiento explícito, entendido como las habilidades, experiencias y hechos que están escritos o pueden escribirse, y se comunican con facilidad; y conocimiento tácito, entendido como las habilidades, experiencias e intuiciones de las personas, que se almacenan en su mente y que no pueden transmitirse con facilidad.

Otra forma de clasificar el conocimiento se fundamenta en su valoración económica, diferenciando entre tres tipos de capital:

- el **capital humano** se refiere al conocimiento útil para la empresa que poseen las personas y equipos de la misma, así como su capacidad de aprendizaje,
- el **capital estructural** está formado por conocimientos estructurados como sistemas de información y comunicación, tecnología, procedimientos de trabajo, patentes, y otros,

- el **capital relacional** hace referencia al valor que para la organización tienen el conjunto de relaciones establecidas con el exterior, como alianzas, contratos, colaboraciones, y otras.

Desde una perspectiva basada en la facilidad de comunicación y explotación, la gestión del conocimiento pretende identificar, organizar y explotar racionalmente el conocimiento explícito (el registrado o susceptible de serlo por la organización) y transformar la mayor cantidad posible de conocimiento tácito (el poseído por los individuos de la organización) en explícito. Por el contrario, desde una perspectiva basada en la valoración económica de los activos, la gestión del conocimiento es el arte de crear valor a partir de los activos intangibles de una organización.

2.4. La sociedad del conocimiento

La aparición de la sociedad y economía del conocimiento viene impulsada por cuatro factores.

1. El fenómeno de la globalización, que interrelaciona las economías de zonas geográficamente dispersas mediante la internacionalización de las empresas, el flujo de capitales, bienes, servicios y personas, y la apertura de nuevos mercados.
2. El fenómeno de las TIC, y en especial de Internet, que ha supuesto un incremento significativo en

las posibilidades de comunicación y transmisión de información y conocimiento.

3. La tendencia de las empresas e instituciones a organizarse de forma más distribuida, fomentando la aparición de redes geográficamente dispersas y descentralizadas.
4. La creciente intensidad en la aplicación del conocimiento en la producción de bienes y servicios.

A partir de la segunda mitad de la década de los años noventa del siglo XX, las tecnologías digitales han permitido, fomentado y ampliado notablemente la importancia económica del conocimiento, mediante dos vías principales. La primera vía ha sido la espectacular mejora del acceso y la gestión de los flujos de información y conocimiento, que ha minimizado las barreras para la difusión de estos dos recursos, y, por consiguiente, se ha producido un notable incremento del conocimiento explícito. La segunda vía, estrechamente vinculada a la primera, ha sido la mejora de las posibilidades de acceso y difusión de los elementos que inciden en el conocimiento tácito, básicamente, las habilidades formativas y la experiencia.

En esta nueva sociedad/economía, el conocimiento pasa a ser un recurso tan significativo o incluso más que el capital y la mano de obra. Esto es, el aumento de conocimiento explícito, la transformación de conocimiento tácito en explícito y el desarrollo de

nuevas habilidades han generado un círculo virtuoso en la producción de conocimiento, que se constituye como uno de los recursos estratégicos de la actividad económica de la actualidad. En el nuevo contexto social y económico en torno al conocimiento, se otorga una gran importancia a la generación, difusión y uso de la información y conocimiento en las organizaciones. El buen uso del conocimiento determinará el nivel de éxito tanto de las organizaciones, como de las economías nacionales.

Las organizaciones se enfrentan al reto de proyectarse y adaptarse a un proceso de cambio que viene avanzando muy rápidamente. Se trata de un proceso dinámico, caracterizado esencialmente por el desarrollo de nuevas tendencias en la generación, difusión y utilización del conocimiento, que están demandando la revisión y adecuación de muchas organizaciones y procurando, asimismo, la creación de otras nuevas con capacidad para asumir y orientar el cambio.

En este contexto, la *sociedad del conocimiento* es una tendencia social y económica con capacidad para generar, apropiar, y utilizar el conocimiento para atender las necesidades de su desarrollo y así construir su propio futuro, convirtiendo la creación y transferencia del conocimiento en herramienta de la sociedad para su propio beneficio. En la sociedad del conocimiento, las comunidades, empresas y organizaciones

avanzan gracias a la difusión, asimilación, aplicación y sistematización de conocimientos creados u obtenidos localmente, o conseguidos del exterior. El proceso de aprendizaje se potencia en común, a través de redes, empresas, comunicación inter e intrainstitucional, entre comunidades y países.

3. Los lenguajes de marcado en la organización del conocimiento

Un lenguaje de marcado es un conjunto de reglas que establecen qué tipo de marcas han de ser utilizadas, de qué modo se distinguirán las marcas del texto del documentos y cómo se insertarán éstas (la gramática y su sintaxis), y cuáles son las marcas permitidas en cada una de las partes del texto. De forma genérica, se pueden distinguir dos tipos de lenguajes de marcado:

- Los lenguajes de marcado procedimentales: orientados a la presentación de los documentos, especifican cómo debe ser procesado el texto para su salida a través de diversos medios (pantalla ordenador, impresora, etc.) Estos lenguajes no aportan información de tipo semántico o estructura; son poco flexibles, dado que cualquier cambio en la presentación del documento implica modificar su marcado; y suelen ser lenguajes específicos de un sistema de procesamiento propietario, lo cual reduce la "portabilidad" de dichos documentos. Algunos ejemplos de estos lenguajes son RTF (Rich Text

Format) de Microsoft y PDF (Portable Document Format) de Adobe.

- Lenguajes de marcado descriptivos: orientados a la descripción formal y de contenido de los documentos. Estos lenguajes aportan información sobre la estructura del documento y describen el contenido informacional del mismo, además, son lenguajes más flexibles, que diferencian entre el contenido real del documento y su representación. Algunos ejemplos son SGML, HTML, XML, etc.

3.1. La familia de normas SGML/XML

El lenguaje de marcado SGML comenzó a gestarse en 1969, cuando un equipo de investigadores de IBM desarrollaron GML (Generalized Markup Language). GML se desarrolló como una vía para crear la documentación básica de la empresa en un formato electrónico transferible y fácil de gestionar. Se trata del primer lenguaje no propietario de marcado de texto (independiente del sistema en el que se crean los documentos y de la plataforma en la que circulan) capaz de definir las estructuras lógicas de cualquier tipo de documento basándose en un serie de normas.

GML consiguió el estatus de norma ANSI (American National Standards Institute) y bajo la denominación de SGML (Structured Generalized Mark-up Language) en 1978. Esta norma fue adoptada como

norma ISO 8879 en 1986 y forma parte de un conjunto más amplio de normas bajo la denominación genérica de Information processing-text and office systems-standard generalized markup language. Además de este conjunto de normas genéricas, la literatura científica reciente ha empezado a referirse a la familia de normas sgml/xml como al conjunto de normas y/o especificaciones subsidiarias y/o complementarias que constituyen un grupo normativo en torno a la ISO 8879 de 1986. Dos ejemplos muy claros de esta familia y ligados directamente al desarrollo de SGML son, por una lado, la norma ISO 10744-1992 Information technology hypermedia/time-based structuring language, que proporciona normas específicas para establecer enlaces hipertexto en documentos sgml; por otro, la norma ISO 10179-1996 document style semantics and specification language, que especifica el sistema de definición de estilos.

Hay tres características de sgml que lo distinguen de otros lenguajes de marcado. Por un lado, sgml pone un mayor énfasis en el marcado descriptivo que en el procedimental. Por otro, introduce el concepto de tipo de documento (document type) y, por extensión, el concepto de definición de tipo de documento (DTD—document type definition), que se emplea para definir un tipo de documento de acuerdo con sus partes constituyentes y la estructura lógica que adoptan dichas partes. Finalmente, sgml es independiente

de cualquier sistema de representación del alfabeto en que esté escrito el texto.

3.2. Las DTD de SGML

Una DTD (definición de tipo de documento) define la estructura de un tipo de documento específico, que abarca varios aspectos: los elementos que pueden formar parte del tipo de documento, el nombre de los elementos y si son repetibles, el orden de los elementos, los contenidos de los elementos, que tipos de marcado pueden ser omitidos, los atributos y su valores por defecto y los nombres de las entidades permitidas.

Una DTD puede desarrollarse para un documento en concreto o para muchos documentos. La creación de una DTD para un solo documento no es eficiente, por tanto, tiene mucho más sentido crear DTDs que puedan ser usadas para muchos documentos. Desde la aparición de SGML se han creado numerosas DTDs para uso general, algunos ejemplos de las DTD de SGML que han tenido mayor repercusión son:

HTML (HyperText Markup Language)

El W3C ha desarrollado una DTD de SGML destinada a la presentación de la información a través de Internet. HTML iba destinado al gran público por su

sencillez, por tanto, se primó la apariencia visual, sin tener en cuenta la importancia de marcar la estructura lógica de un documento electrónico. A partir de HTML 4.0 se incorporan los elementos META de HTML que pueden ser usados para describir las propiedades de un documento y posibilitan la incorporación de los metadata Dublin Core.

TEI (Text Encoding Initiative)

Surge en 1987 como un proyecto del área de las humanidades, promovido en un congreso de la Association for Computers in the Humanities (ACH), pero su publicación efectiva para la codificación de textos, fundamentalmente literarios, es posterior (1994). Se trata de una DTD (definición del tipo de documento) madura y bien formada de SGML para asegurar un formato estándar con amplias capacidades de marcado para la indización y el intercambio de información textual.

El esquema TEI presenta algunas ventajas con relación a las particularidades de la metainformación necesaria para la documentación de recursos informativos electrónicos debido a su control documental. TEI se compone de un encabezamiento seguido del texto en sí. El encabezamiento consta, a su vez, de cuatro elementos que proporcionan información

(principalmente bibliográfica) relativa al documento como fichero electrónico.

El encabezamiento TEI constituye una de las principales contribuciones para la codificación sgml de información bibliográfica. Además, las directrices TEI incluyen un a sección especial (24.3) sobre los elementos de este encabezamiento y su relaciones con los registros MARC.

EAD (Encoded Archival Description) DTD.

La DTD de EAD define una clase de documentos, que, en términos generales, constan de una página de título opcional, la descripción de una unidad de material archivístico y unas apéndices también opcionales. Esta DTD posibilita el empleo de xml a partir de su versión 1.0 y ha sido diseñada para reflejar la jerarquía natura que presenta la organización de los fondos, en conjunción con el orden intelectual que imponen los archiveros con sus prácticas descriptivas.

MARC (Machine Readable Catalogue) DTD.

La DTD de MARC, trata el formato MARC como un tipo de documento específico. La DTD define todos los elementos que pueden aparecer en un

registro MARC y especifica como se deben codificar y representar en sgml. El objetivo principal de MARC DTD fue que los registros se pudieran importar y exportar a sgml automáticamente, para ello se ha seguido la estructura de un registro MARC, incluyendo las etiquetas, indicadores y códigos de subcampo.

3.3. El lenguaje XML

XML es un proyecto del World Wide Web Consortium (W3C) y su desarrollo está coordinado por el XML working group. Se trata de un subconjunto (o simplificación) adaptado de SGML que tiene la intención de aprovechar sus máximas ventajas posibles, pero permitiendo su implementación en Internet.

El lenguaje XML consta de cuatro especificaciones (recomendaciones de W3C):

DTD (Document Type Definition): Definición de tipo de documento. Se trata de un archivo que encierra una definición formal de un tipo de documento y, a la vez, especifica la estructura lógica de cada documento. El DTD en XML es opcional. En tareas sencillas no es necesario construir una DTD, entonces se trataría de un documento "bien formado" (well-formed). Si se construye una DTD, entonces será un documento "validado" (valid).

XSL (eXtensible Stylesheet Language): Lenguaje de estilo para XML. Se trata de un lenguaje para elaborar hojas de estilo. Consta de tres partes: XSL

Transformations (XSLT), que es un lenguaje para transformar documentos XML; XML Path Language (XPath), que es un lenguaje de expresión usado por XSLT para acceder o referir partes de un documento XML (XPath se usa también en la especificación XML Linking); XSL Formatting Objects, es un vocabulario para especificar la semántica del formato.

XLL (eXtensible Linking Language): Lenguaje de enlaces en XML. Define el modo de actuación entre diferentes enlaces. Se considera un subconjunto de HyTime (ISO 10744) y sigue algunas especificaciones de TEI. XLL tiene dos importantes componentes: Xlink y Xpointer, con los que se va mucho más allá de los enlaces simples soportados por HTML, los enlaces pueden ser bidireccionales, múltiples (anillos, múltiples ventanas, etc.), agrupados (múltiples orígenes).

XUA (XML User Agent): Estandarización de navegadores XML. Se aplica a los navegadores para que sean capaces de reconocer todas las especificaciones XML.

Las características esenciales de XML son:

- XML es una arquitectura abierta y extensible. No necesita versiones para funcionar en futuros navegadores.
- XML posee mayor consistencia, homogeneidad y amplitud de los identificadores descriptivos del

documento (RDF), en comparación con los atributos de la etiqueta de HTML.

- XML permite agrupar una amplia variedad de datos y aplicaciones, desde páginas web hasta bases de datos.
- XML consigue que la estructura de la información resulte más accesible, por tanto, los motores de búsqueda devolverán respuestas más adecuadas y precisas.
- XML permite el desarrollo de búsquedas personalizables para robots y agentes inteligentes.
- XML desarrolla ampliamente el concepto de "hipertexto" mediante enlaces bidireccionales, enlaces que pueden especificarse y gestionarse desde fuera del documento, hiperenlaces múltiples, enlaces agrupados, atributos para enlaces, etc.
- XML facilita la exportabilidad a otros formatos de publicación (papel, web, cd-rom, etc.)

3.4. Mecanismos para representación del conocimiento

Cada lenguaje de marcado de información utiliza mecanismos diferentes para la representación del conocimiento contenido en un documento. En este apartado vamos a tratar de los dos lenguajes de marcado más extendidos para la publicación de contenidos en Internet: HTML y XML. Para ambos lengua-

jes determinaremos los mecanismos que podríamos denominar de representación del conocimiento documental o codificación de la meta información documental.

Dublin Core (DC)

La Iniciativa de Metadatos Dublin Core es actualmente una iniciativa consorciada para el desarrollo del modelo de metadatos Dublin Core, surgido originariamente en 1995 en el ámbito bibliotecario. Su desarrollo, acorde con la evolución de Internet, lo ha convertido en un formato altamente normalizado y utilizado en distintos sectores. El modelo DC se compone de un conjunto de 15 elementos que describen el conocimiento contenido en un recurso de una forma estructurada.

Resource Description Framework (RDF)

RDF es sin duda alguna el estándar en desarrollo más importante para la descripción de contenidos web. Su importancia reside, por un lado en la entidad Consorcio Web (W3C) quien fomenta el proyecto, y por otro, el hecho de ser una aplicación de metadatos que utiliza XML a fin de proporcionar un marco estándar para la interoperabilidad entre distintos modelos de metadatos para la representación

de los conocimiento (metainformación) contenido en un documento (como por ejemplo el DC, que admite las especificaciones RDF a partir de la versión DC5).

Desde que RDF se convirtió en una recomendación de W3C en febrero de 1999, se han desarrollado un buen número de herramientas que permiten trabajar con RDF de una forma más eficiente.

RDF ofrece una variedad de aplicaciones, tales como: catálogos de biblioteca y directorios web (ejemplos: Dublin Core Metadata Initiative, OCLC Connexion, Open Directory Project, etc.); categorización y gestión de listas de noticias, software y contenidos (PICS, XMLNew, UK Mirror Service), así como, colecciones de música, fotos y eventos (MusicBrains metadata initiative, RDFPic, etc.); usando XML como sintaxis de intercambio.

3.5. Servicios de información web basados en XML

Las organizaciones que prestan servicios de información deben diseñar sistemas eficaces para gestionar y difundir la información. Con la aplicación de las tecnologías de la información en el desarrollo e implementación de sistemas de información y la utilización de la tecnología web, un porcentaje muy alto de los servicios de información se fundamentan en información electrónica. En este contexto la familia de normas sgml resulta clave, puesto que asegura la

consistencia (en la producción, el procesamiento, el almacenamiento y la distribución) y proporciona una enorme flexibilidad (en la presentación y en el formato, por ejemplo).

Los profesionales de la información están muy interesados en el potencial de XML para la organización de la información. Uno de los campos de mayor actividad consiste en la sustitución de MARC usando tecnología XML, de hecho la Library of Congress y otras organizaciones ya han llevado a cabo este tipo de proyectos.

Cada día son más numerosos los proyectos xml que se están emprendiendo en el ámbito de los servicios de información comercial. Se trata de aplicaciones relacionadas con dominios muy variados: banca, telecomunicaciones, revistas electrónicas, etc. Esta proliferación de aplicaciones xml no significa que éste vaya a sustituir a sgml. En principio, mantener una o varias bases de datos sgml y filtrarlas a xml para las operaciones de indización, otros tratamientos de los datos y distribución en la web, se conforma como una estrategia a largo plazo con gran posibilidad de éxito. De hecho, la mayoría de las bibliotecas digitales mantienen sus datos en sgml, pero distribuyen la información en xml.

3.6. XML en la organización del conocimiento interno

La literatura científica reciente se centra en las ventajas de XML para el desarrollo de sitios web y su aplicación a las tecnologías de búsqueda y recuperación de información web. En esta ocasión es necesario tener en cuenta los usos de XML para la gestión del conocimiento organizacional. XML es un lenguaje que presenta simultáneamente el contenido para ser publicado en Internet y describe este contenido de forma que otro software puede entender y usar este conocimiento. La clave de XML es que, en oposición a HTML, proporciona información sobre el significado de los datos, lo que va a facilitar el procesamiento automático de la información contenida en la información (meta información) y la consiguiente gestión del conocimiento.

XML se empieza a considerar como "la próxima revolución en la gestión del conocimiento", y las organizaciones están empezando a entender el potencial de esta tecnología para el desarrollo de arquitecturas de información corporativas.

La tecnología XML, por sí misma no reportan ningún valor a la organización, su valor depende de cómo se use esta tecnología dentro de la organización. Su implementación no debe ser departamental e, idealmente, debería incluir a partners estratégicos y otras

organizaciones con las que se necesita compartir datos e información. Un ejemplo claro de los beneficios de XML en la gestión de la información corporativa consiste en la creación de portales corporativos que ofrecen prestaciones para diferentes funciones de negocio, tales como: ERP, datwarehousing, sistemas de apoyo a la decisión y gestión del conocimiento organizacional.

Entre las aplicaciones estratégicas que XML puede tener dentro de una organización podemos destacar:

Cadena de suministro: La gestión eficiente del conocimiento organizacional depende del acceso a la información externa. XML puede ayudar a mejorar el funcionamiento de la cadena de suministro y la extranet de la organización. Muchas organizaciones necesitan compartir información y no siempre tienen una vía para hacerlo, puesto que tienen infraestructuras tecnológicas diferentes y no compatibles con algunos agentes de la cadena de suministro o con los clientes. XML puede proporcionar una forma estándar para intercambiar información con proveedores, distribuidores y clientes, lo cual se convierte en una diferencia competitiva para la organización capaz de gestionar la cadena de suministro mediante XML.

Retorno de Inversiones (ROI): Los documentos XML pueden producir diferentes outputs y una vez que la información organizacional está organizada en documentos XML, se pueden elaborar multitud de productos y servicios de información vía web, así como en diferentes plataformas: PDA, telefonía móvil, televisión digital, etc. La creación de un corpus de conocimiento organizacional en XML supone un gran esfuerzo de inversión, pero pronto reporta un ahorro en los costes directos relacionados con la inversión en tecnologías de la información, la formación de los empleados y los servicios de soporte.

Gestión integral del conocimiento: XML posee un conjunto de características que lo convierten en un aliado para la gestión del conocimiento organizacional. Si se combina el potencial de XML y las posibilidades de RDF, como herramienta para la gestión del conocimiento que permite organizar, interrelacionar, clasificar y anotar el conocimiento contenido en los documentos XML, se puede obtener un incremento considerable del valor añadido de los datos almacenados en los documentos XML y conlleva una gestión automática del conocimiento explícito de la organización.

4. Arquitectura de la información, navegación y recuperación

Los servicios electrónicos generalmente adoptan una estructura hipertextual y un formato web que pueden ocasionar una serie de problemas en cuanto a la navegación por los contenidos y la recuperación de información útil. Estos problemas suelen estar relacionados con la sobrecarga cognitiva y la desorientación. Para superar tales problemas los sistemas de clasificación, como herramientas para la organización de contenidos web, contribuyen a facilitar la recuperación de información y el aprendizaje de estrategias de navegación. A este respecto, la característica principal del aprendizaje que se produce a través de la navegación hipertextual estriba en la posibilidad de organizar determinados conocimientos según estructuras diferentes que permitan al usuario conseguir su objetivo.

Tradicionalmente, un sistema de clasificación se refiere a un lenguaje documental que organiza lógicamente una estructura de conceptos y/o notaciones, y que está destinado a permitir la clasificación de documentos conforme a sus respectivas temáticas. El sis-

tema de clasificación, como todo lenguaje documental, establece tres tipos de relaciones básicas entre los conceptos que lo integran: de sinonimia o identidad, jerárquicas o asociativas. En cuanto a su tipología pueden ser muy diversos: por su cobertura temática pueden ser universal o especializado; pueden tener bajo, medio o alto nivel de especificidad según el detalle y la profundidad de su desarrollo; y pueden clasificarse como enumerativo, prefacetado o facetado, de acuerdo con sus características estructurales. En general, los sistemas de clasificación son aplicables a un universo indeterminado de objetos y por ello favorecen la organización lógica de un ámbito de conocimiento.

4.1. La arquitectura de la información

En el ámbito de la información electrónica y la arquitectura de contenidos web se entiende por clasificar la actividad de agrupar los elementos de información de acuerdo a atributos o propiedades comunes entre ellos. Los elementos de información en realidad son contenidos, esto es, aquellos 'trozos' de información u objetos informativos a organizar, estructurar y clasificar, que pueden ser: textos, imágenes, videos, y otros. Por tanto, un sistema de clasificación consiste en elegir en base a qué atributos se agrupan los contenidos y cómo se organizan estos atributos.

Siguiendo a Rosenfield y Morville, los sistemas de clasificación pueden ser exactos o ambiguos. En los sistemas de clasificación exactos, los contenidos están perfectamente definidos y diferenciados del resto; un ejemplo de este tipo de sistemas son la clasificación alfabética, la cronológica y la geográfica. En los sistemas de clasificación ambiguos, los contenidos están organizados en categorías no definidas de forma exacta y precisa; un ejemplo de este tipo de sistemas son la clasificación temática o por categorías, la clasificación orientada a tareas, la clasificación orientada a la audiencia o tipología de usuarios, y la clasificación metafórica. Todos ellos son sistemas válidos para organizar los contenidos web de una organización y se utilizan actualmente para facilitar el acceso a la información corporativa, y a los servicios y productos de las organizaciones públicas.

La visión tradicional de sistemas de clasificación enlaza con la actual perspectiva de la disciplina de Arquitectura de la Información. El término arquitectura de la información fue utilizado por primera vez por Richard Saul Wurman en 1975, quién la define como el estudio de la organización de la información con el objetivo de permitir al usuario encontrar su vía de navegación hacia el conocimiento y la comprensión de la información. Por tanto, la arquitectura de la información restringida al ámbito web sería el arte y la ciencia de estructurar y clasificar sitios web con el fin de

ayudar a los usuarios a encontrar y manejar la información. De esta definición se desprende que la arquitectura de la información de un sitio web, comprende los sistemas de organización y estructuración de los contenidos, los sistemas de rotulado o etiquetado de dichos contenidos, y los sistemas de recuperación de información y navegación que provea el sitio web.

Teniendo en consideración la visión tradicional de los sistemas de clasificación y combinándolos con las nuevas tendencias en cuanto a sistemas de clasificación en el entorno web, las herramientas tradicionales, como las listas de términos, las taxonomías, los tesauros, y otros, se pueden utilizar para proceder a la clasificación y organización del conocimiento electrónico, fundamentalmente en los sistemas de clasificación ambiguos. La arquitectura de la información incluye muchos aspectos diferentes sobre la creación y organización de un sitio web, pero sus herramientas principales son las técnicas de organización de información desarrollada en otras disciplinas. De hecho, la mayoría de estas técnicas provienen de la Documentación, en particular en lo referente a vocabularios estructurados.

4.2. Sistemas de clasificación del conocimiento electrónico

Para realizar las tareas de clasificación de los contenidos en un sitio web corporativo, se pueden utilizar diferentes herramientas de clasificación temática o por materias, que van a contribuir a la orientación del usuario de los servicios web y a la mejoría en la recuperación de información, puesto que ésta ha sido organizada a priori mediante listas de categorías, taxonomías, tesauros o clasificaciones facetadas, de acuerdo con las características del servicio o producto de información web. Todas las herramientas mencionadas son vocabularios controlados y estructurados que se utilizan para etiquetar contenidos y para la posterior localización de estos contenidos mediante sistemas de navegación o búsqueda. Sin embargo la definición específica de estas herramientas es muy difusa y no existe una definición única y consensuada, que permita diferenciar claramente, por ejemplo, entre lista de categorías y taxonomía, dado que las definiciones existentes en la literatura científica tienden a ser muy heterogéneas entre sí. Por tanto, se aporta una definición muy escueta y genérica que se adapte al propósito de este artículo.

Una *lista de categorías* es una lista controlada de términos que denota una materia y que se utiliza para clasificar, indizar y recuperar recursos de información.

La función de la lista de categorías consiste en evitar al usuario definir términos sin un significado específico, términos demasiado amplios o excesivamente restrictivos, o términos muy semejantes en cuanto a la forma.

El término *taxonomía* proviene del área de las Ciencias Naturales, en particular de la Biología, que estudia la clasificación de los seres vivos en estructuras jerárquicas. Se trata de un término muy utilizado en Arquitectura de la Información, y según en qué contexto se utilice, se le podría asignar un significado u otro. De forma muy esquemática, una taxonomía es una lista estructurada y jerárquica de elementos o grupos de elementos, que presenta una forma arbórea. Una taxonomía es la organización jerárquica del conjunto de categorías (palabras clave) bajo las que clasificar las unidades de contenido. Por ejemplo, en el caso de un servicio de administración electrónica basado en eventos de la vida, se identifican bajo la taxonomía "estudios", las categorías: "estudios primarios", "estudios secundarios", "estudios universitarios", y "otros estudios". Por tanto, la función de la taxonomía en la organización de los contenidos web reside en la posibilidad de agrupar términos jerárquicamente relacionados bajo una taxonomía que facilite encontrar el término correcto para la búsqueda o descripción de un contenido específico.

Los *tesauros* se pueden considerar una taxonomía con extras, ya que permiten representar la realidad mediante términos no sólo organizados de forma jerárquica, sino permitiendo otro tipo de relaciones entre ellos, como son la relación de equivalencia y de asociación. La tesauros están normalizados mediante la norma ISO 2788, en la que se define tesoro como "un vocabulario controlado y dinámico, compuesto por términos que tienen entre ellos relaciones semánticas y genéricas y que se aplica a un dominio particular del conocimiento". Además de la mencionada función de control del vocabulario, el tesoro constituye una herramienta que representa, por una parte, la estructura de los conocimientos transmitidos por los creadores de la información (arquitectos de la información), y, por otra, la estructura cognitiva de los usuarios que formulan la expresión de búsqueda.

La *clasificación facetada* parte de la premisa de que una unidad de contenido puede ser descrita a través de varias dimensiones o facetas, cada una de las cuales contiene su propia relación de posibles valores o categorías. La clasificación por facetas fue propuesta en 1933 por Ranganathan, y es también conocida como clasificación colonada. Las clasificaciones facetadas se utilizan para organizar conjuntos de objetos (sitios web, productos, documentos, y otros), con una homogeneidad suficiente como para que puedan ser descritos por un número determinado de atributos o

propiedades (facetas y categorías) y sus valores (pertenencia a categorías). En principio, la clasificación por facetas parece muy diferente de un tesoro, pero, de hecho, una clasificación facetada puede ser considerada como una vía adecuada para construir un tesoro o bien para ser utilizada con propósitos de clasificación de contenidos. En el ámbito de los servicios web existe un lenguaje con sintaxis XML para definir, distribuir e intercambiar metadatos en forma de taxonomías facetadas. Este lenguaje ha sido desarrollado por Peter van Dijck y se denomina XFML (eXchangeable Faceted Metadata Language).

Ontologías son los instrumentos o modelos de estructuras con una traducción tecnológica que pueden usarse para la presentación y uso en web de catálogos, bases de datos, listas de encabezamiento, glosarios, tesoros, por ejemplo; e implican una especificación, conceptualización y modelización del conocimiento que se recoge en cada uno de los lenguajes documentales. Como se observa por la definición aportada, se puede establecer cierto paralelismo entre ontología y tesoro; fundamentalmente, en la delimitación de los términos que deben formar parte de una construcción de representación del conocimiento, y de las relaciones que se establecen entre los términos.

4.3. Navegación y recuperación de la información web

Existen dos formas básicas de arquitectura de la información en las que se utiliza control del vocabulario para la clasificación de contenidos con el objeto de organizar el conocimiento electrónico: los esquemas de navegación y los sistemas de recuperación. Los esquemas de navegación no pueden ser ambiguos, y se utilizan para orientar a los usuarios sobre la tipología de los recursos de información disponibles en un sitio web. En los sistemas de recuperación de información, el usuario utiliza términos o frases para expresar su necesidad de información, en general, utilizando un formulario de consulta que puede incluir uno o más campos. La cuestión radica en que ambas formas de acceso a la información son necesarias, puesto que son complementarias y para ambas es necesario cierto control del vocabulario.

La información debe organizarse de una forma coherente, de tal manera que el usuario pueda acceder a la información deseada e importante para él/ella de una manera rápida y eficaz. La organización de la información es esencial para que el sistema de información sea útil.

La arquitectura de la información de un sitio web comprende los sistemas de organización y estructuración de los contenidos, los sistemas de rotulado o

etiquetado de dichos contenidos, y los sistemas de recuperación de información y navegación que provea el sitio web.

Un buen sistema de información requiere algo más que la materia prima. Y esta necesidad de organización es cada vez más clara conforme el usuario dispone de menos tiempo para poder dedicar a moverse por las fuentes. La organización de la información es esencial para que el sistema de información sea útil. La organización de la información en un sistema, por ejemplo en una página web, es responsabilidad del arquitecto de información.

5. Los metadatos y la recuperación del conocimiento

Metadato es toda aquella información descriptiva sobre el contexto, calidad, condición o características de un recurso que tiene la finalidad de facilitar su recuperación, autenticación, evaluación, preservación o interoperabilidad. Para la aplicación de metadatos se han desarrollado distintos modelos, esquemas o estándares de metadatos, que si bien comparten una sintaxis y estructura de la información XML, difieren atendiendo a los propósitos de la información que describen y a sus necesidades de especificidad y gestión remota de los recursos en cuestión

5.1. La función de los metadatos

La aplicación de metadatos supone una mejora en la organización y recuperación de la información, tanto de forma humana como automatizada. La gran incógnita en este sentido consiste en determinar los beneficios específicos que aportan los metadatos en la búsqueda y recuperación de la información web, sobre todo cuando muchos motores de búsqueda no utilizan los metadatos como un criterio en la inde-

xación de los recursos electrónicos y, por tanto no se utilizan para la búsqueda. Si embargo, existe una amplia gama de software de motor de búsqueda para la indexación de los recursos del sitio web, la Intranet y los productos electrónicos (CD-ROMs, DVDs y otros productos que utilicen tecnología web) de las organizaciones que indexan y gestionan metadatos.

Algunos motores de búsqueda (como Convera, Harvest, Blue Angel, Microsoft Site Index, etc.) son capaces de utilizar los metadatos y otras herramientas de representación del conocimiento (como ontologías y topic maps) para obtener mejores resultados en la recuperación. Por tanto, aunque las organizaciones públicas pueden sentirse reticentes ante la incorporación de metadatos en sus recursos de información debido al esfuerzo económico que ello supone (coste de personal y tecnología), es importante que las organizaciones que desarrollan servicios y productos web tomen conciencia de la importancia de los metadatos para mejorar la relevancia de los sistemas de recuperación de información, así como para facilitar la integración y combinación de recursos heterogéneos en el desarrollo de servicios electrónicos y mejorar el acceso de los usuarios a los recursos.

Los sistemas de recuperación de la información en Internet de propósito general (motores de búsqueda) se basan en la extracción automática de la información y utilizan sencillas técnicas para representar

el conocimiento contenido en los recursos electrónicos. Por tanto, no pueden dar una respuesta precisa a una pregunta concreta sobre el contenido semántico de los documentos y por ello, recuperan mucho ruido. Sin embargo, los sistemas de recuperación en sectores específicos, como la información pública, dado que todos los recursos de información son objeto de descripción, organización y control del vocabulario, ofrecen mayor relevancia en la recuperación.

La existencia de un compromiso para que la información de carácter público se adapte a unos estándares, y contemple el uso de metadatos en todos los recursos electrónicos y digitales, favorece la recuperación de la información en este ámbito de conocimiento. La clave esencial reside en la aplicación de metadatos de forma sistemática, normalizada y coherente. Con este proceso se facilita la descripción de todos los recursos de la organización (aplicación sistemática), el intercambio de información (mediante la normalización), y su adaptación a nuevas formas tecnológicas (aplicación coherente).

5.2. Métodos para registrar y transferir metadatos

Los metadatos, por el momento, son la herramienta más extendida de representación del conocimiento web para favorecer la recuperación de información. El modelo de metadatos más difundido y utiliza-

do es Dublin Core, que originalmente se aplicaba sobre el lenguaje de marcado HTML, y que actualmente puede aplicarse a diferentes entornos tecnológicos. La situación ha variado considerablemente en los últimos años, de forma que para garantizar la interoperabilidad entre los propios metadatos se han desarrollado múltiples métodos para registrar y transferir metadatos Dublin Core. Los más comunes son: HTML, XML, RDF y las bases de datos relacionales.

La Iniciativa de Metadatos Dublin Core es actualmente una iniciativa consorciada para el desarrollo del modelo de metadatos Dublin Core, surgido originariamente en 1995 en el ámbito bibliotecario. Su desarrollo, acorde con la evolución de Internet, lo ha convertido en un formato altamente normalizado y utilizado en distintos sectores. El modelo DC se compone de un conjunto de 15 elementos que describen el conocimiento contenido en un recurso de una forma estructurada. En la siguiente tabla se agrupan los elementos según su función: descripción de contenido, determinación de la propiedad intelectual o la instanciación de recursos.

Tabla de los 15 elementos DC

Contenido:

- Title (título)
- Subject (tema)
- Description (descripción)
- Source (fuente)
- Language (lenguaje)
- Relation (relación)
- Coverage (cobertura).

Propiedad intelectual:

- Creator (autor)
- Publisher (editor) y, otras colaboraciones
- Contributor (otros autores/colaboradores)
- Rights (derechos).

Instanciación:

- Date (fecha)

- Type (tipo de recurso)
- Format (formato)
- Identifier (identificador)

Para la codificación de los metadatos se pueden utilizar diferentes métodos dependiendo del nivel de desarrollo tecnológico de la organización. En principio la codificación de los elementos DC se realizaba en la cabecera de un documento HTML, en el contexto de los elementos META y embebido en el contenido del propio documento.

```

<head>
<meta http-equiv="content-type" content="text/html; charset=iso-8859-1" />
<meta name="author" content="e-Government Unit" />
<meta name="robots" content="all" />
<meta name="DC.creator" content="e-Government Unit" />
<meta name="DC.title" content="XML Schemas" />
<meta name="eGMS.subject.category" scheme="GCL" content="e-Government" />
<meta name="DC.date" content="30/04/2001" />
<meta name="DC.identifier" content="http://www.govtalk.gov.uk/schemasstandards/xmlschema.asp" />
<meta name="DC.publisher" content="e-Government Unit" />
<meta name="DC.description" content="The e-Government Unit is part of the Cabinet Office and works to ensure that IT supports the business transformation of government itself so that government can provide better and more efficient public services." />
<meta name="DC.subject" content="access, business, cabinet office, channels, civil service, e-commerce, e-envoy, e-gif, e-gms, e-government, e-government unit, e-minister, gdsc, government information, govtalk, great britain, information, information age, internet, online, interoperability, metadata, public sector, schema, schemas, standards, uk, uk govtalk, xml, xml schemas" />
</head>

```

Ejemplo de registro DC en HTML.

La evolución de las tecnologías web ha permitido que en XML, los elementos de metadatos se registren de una forma más estructurada, por ejemplo en forma de *namespaces* o *esquemas* XML (XML Schemes o XMLS). Por su parte, RDF (Resource Description

Framework / Marco para la Descripción de Recursos) es el estándar más importante para la descripción de contenidos web y posee el mayor potencial para la expresión de metadatos. De forma similar, la gestión de metadatos también puede realizarse en el contexto de un sistema de gestión de bases de datos e incluso en un sistema de gestión de contenidos web. Ello asegura que se utilicen procedimientos para garantizar la consistencia y sincronización entre los metadatos y los recursos web que describen.

La noción de namespaces (espacios de nombres) es parte fundamental de la infraestructura web y particularmente de XML. Son una parte crítica de la infraestructura necesaria para desplegar sistemas de metadatos modulares en la web. Las declaraciones de namespace permiten al diseñador de esquemas de metadatos definir el contexto para un término particular, asegurando que el término tiene una única definición dentro de los límites del namespace declarado.

```

<?xml version="1.0"?>
<!DOCTYPE rdf:RDF PUBLIC "-//DUBLIN CORE//DCMES DTD 2002/07/31//EN"
    "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="http://dublincore.org/">
    <dc:title>Dublin Core Metadata Initiative - Home Page</dc:title>
    <dc:description>The Dublin Core Metadata Initiative Web
    site.</dc:description>
    <dc:date>2001-01-16</dc:date>
    <dc:format>text/html</dc:format>
    <dc:language>en</dc:language>
    <dc:contributor>The Dublin Core Metadata
    Initiative</dc:contributor>
    <!-- guesses for the translation of the above titles -->
    <dc:title xml:lang="fr">L'Initiative de mitadonnies du Dublin
    Core</dc:title>
    <dc:title xml:lang="de">der Dublin-Core Metadata-
    Diskussionen</dc:title>
</rdf:Description>
</rdf:RDF>

```

Ejemplo de registro DC en RDF

```

<?xml version="1.0"?>

<metadata
  xmlns="http://example.org/myapp/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://example.org/myapp/
http://example.org/myapp/schema.xsd"
  xmlns:dc="http://purl.org/dc/elements/1.1/">

  <dc:title>
    UKOLN
  </dc:title>

  <dc:description>
    UKOLN is a national focus of expertise in digital information
    management. It provides policy, research and awareness services
    to the UK library, information and cultural heritage
    communities.

    UKOLN is based at the University of Bath.
  </dc:description>

  <dc:publisher>
    UKOLN, University of Bath
  </dc:publisher>

  <dc:identifier>
    http://www.ukoln.ac.uk/
  </dc:identifier>

</metadata>

```

Ejemplo de registro DC en XML

5.3. Las ventajas de RDF

De los sistemas de codificación de metadatos destaca RDF, que es un sistema de metadatos que sirve tanto para la descripción por sí mismo de recursos electrónicos, como para envolver otros sistemas de metadatos con el fin de lograr un marco genérico de trabajo, facilitando de esta forma la interconexión entre diferentes métodos de descripción. Su importancia reside en que una organización como el W3C fomenta el proyecto, y en el hecho de ser una aplicación de metadatos que utiliza XML, a fin de proporcionar un marco estándar para la interoperabilidad entre distintos modelos de metadatos para la representación del conocimiento (metainformación) contenido en un recurso. RDF está constituido por un conjunto de especificaciones técnicas desarrolladas por W3C para la normalización de los metadatos en XML. Desde que RDF se convirtió en una recomendación de W3C, en febrero de 1999, se han desarrollado un buen número de herramientas que permiten trabajar con RDF de una forma más eficiente.

El sistema de metadatos RDF proporciona una herramienta genérica para la creación, gestión y búsqueda de datos inteligibles por máquinas en la web; y se trata de una capa adicional (semántica) sobre XML (sintáctica) que simplifica la reutilización de términos de vocabulario entre namespaces. Las ventajas esenci-

ales de RDF son las siguientes: proporciona las bases para la interoperabilidad en los metadatos; proporciona una semántica para la creación de metadatos inteligibles por máquinas; ofrece mayor precisión en la recuperación de recursos que las búsquedas a texto completo; y permite la automatización del descubrimiento de recursos (como los motores basados en robots) y la indexación de los recursos inteligentemente (como los motores basados en directorios).

Bibliografía

- **CORKERN, Carla.** *XML application integration and knowledge management.* XML Europe 2001. How XML powers industry applications. 21-25 May 2001. Berlin.
- **DÍAZ ORTUÑO, Pedro Manuel.** "Problemática y tendencias en la arquitectura de metadatos web". En: Anales de documentación, 2003, n. 6, pp. 35-58.
- **EDEN, Brad.** *Metadata standards.* Library technology reports, September-October 2002.
- **GARCÍA JIMÉNEZ, Antonio.** *Instrumentos de representación del conocimiento: tesauros versus ontologías.* Anales de documentación, 2004, num. 7. p. 79-95.
- **GARDNER, John Robert.** *Information architecture planning with XML.* Library Hi Tech, vol. 19, num. 2, 2001, p. 231-241
- **MARTÍN GALAN, Bonifacio; RODRÍGUEZ MATEOS, David.** *Estructuración de la información mediante XML: un nuevo reto para la gestión documental.* VII Jornadas Españolas de Documentación: la gestión del conocimiento, retos y soluciones de los profesionales de la información, Bilbao, octubre de 2000. p.113-123

- **MARTINEZ USERO, José Ángel.** *El uso de metadatos para mejorar la interoperabilidad del conocimiento en los servicios de administración electrónica.* El profesional de la información, 2006, vol.15, n. 2, pp. 114 -126. <http://www.ucm.es/eprints/5638/>
- **MARTÍNEZ USERO, José Ángel; BELTRÁN ORENES, María Pilar.** *Ontologies in the context of knowledge organization and interoperability in e-government services.* EN: Conference on Digital Divide, Global Development and the Information Society, 2005,Túnez. <http://www.ucm.es/eprints/5631/>
- **MENDOZA, Cornelius.** *Another use of XML: internal knowledge management.* Serverworld Magazine, June 2000, Issue 6. <http://www.serverworldmagazine.com/compaqnet/2000/06/anotheru.shtml>
- **MONTERO, Yusef Hassan; NUÑEZ PEÑA, Ana.** (2005).*Diseño de arquitecturas de información: descripción y clasificación.*
- **PEIS, Eduardo; MOYA, Felix de.** *SGML y servicios de información.* El profesional de la información, vol. 9, num. 6, junio 2000, p. 4-17
- Resource Description Framework (RDF). W3C Semantic Web Activity. <http://www.w3c.org/RDF/>
- **ROSENFELD, Louis; Morville, Peter.** *Information architecture for the World Wide Web: designing large-scale web sites.* O'Reill, 2002.

- **SAN SEGUNDO, Rosa; BELTRÁN, Pilar.** *Aplicación de ontologías en la organización de información en Internet.* En: CИСCI 2003. Memorias. 2da. Conferencia iberoamericana en sistemas, cibernética e informática. Orlando (Florida).
- **SERRADELL LÓPEZ, Enric; JUAN PÉREZ, Ángel A.** *La gestión del conocimiento en la nueva economía.* (2003). <http://www.uoc.edu/dt/20133/index.html>.