

Los sistemas de gestión de documentos multimedia

Noemí Mases Blanch

PID_00202592



Los textos e imágenes publicados en esta obra están sujetos –excepto que se indique lo contrario– a una licencia de Reconocimiento-NoComercial-SinObraDerivada (BY-NC-ND) v.3.0 España de Creative Commons. Podéis copiarlos, distribuirlos y transmitirlos públicamente siempre que citéis el autor y la fuente (FUOC. Fundació para la Universitat Oberta de Catalunya), no hagáis de ellos un uso comercial y ni obra derivada. La licencia completa se puede consultar en <http://creativecommons.org/licenses/by-nc-nd/3.0/es/legalcode.es>

Índice

Introducción	5
Objetivos	6
1. Origen, funciones y arquitectura de los sistemas de gestión de contenidos (CMS, Content Management Systems) ..	7
2. Sistemas de gestión de contenidos y sistemas de gestión documental (DMS, Document Management Systems)	10
3. Flujo de trabajo, control y acceso de un sistema de gestión de documentos multimedia (MDMS, Multimedia Document Management Systems)	12
4. La gestión de la calidad en un sistema de gestión documental	14
5. Ejemplos de sistemas de gestión de documentos multimedia	15
5.1. Alfresco	15
5.2. Nuxeo	16
5.3. KnowledgeTree	16
5.4. OpenKM	16
6. El concepto de Web semántica	18
6.1. ¿Qué es la Web semántica?	18
6.2. Arquitectura y estándares de la Web semántica	19
6.3. La Web semántica y los sistemas de gestión de contenidos	20
6.4. Proyectos e iniciativas internacionales en torno a la Web semántica	21
6.5. El futuro de la Web semántica	22
Bibliografía	23

Introducción

Actualmente, las empresas e instituciones tienen la necesidad de gestionar los ciclos de vida de sus documentos, independientemente de los soportes y formatos, para conseguir sus objetivos y llevar a cabo sus actividades y procesos de una manera más eficiente. Muchas de estas organizaciones ya disponen de sistemas de gestión de contenidos, conocidos como *CMS (Content Management Systems)* que crean, editan, gestionan y publican contenidos digitales. Ahora bien, no son tantas las que contemplan otras tecnologías, métodos y técnicas que tienen que ver con la gestión documental, como por ejemplo la gestión del conocimiento, la gestión del ciclo de vida de los documentos o la preservación digital, tan necesarios para la buena difusión y recuperación de la información, así como también para el funcionamiento óptimo de los flujos de trabajo dentro del entorno organizativo.

Este módulo pretende incidir en el hecho de que la gestión de contenidos y la gestión de la información y la documentación en las organizaciones son dos conceptos cruciales y no pueden entenderse de manera separada. En los dos primeros apartados se explican las funcionalidades y la arquitectura de los sistemas de gestión de contenidos tradicionales y se enumeran las tecnologías que tendrían que tener para funcionar, también, como sistemas de gestión documental. En el tercer apartado se analizan a fondo los flujos de trabajo de estos sistemas de gestión documental, que evidentemente soportan y gestionan documentos multimedia y, en último lugar, se realiza una breve introducción de cómo se lleva a cabo la gestión de la calidad en estos entornos, y se muestran algunos ejemplos de aplicaciones que podemos encontrar actualmente en el mercado.

Por último, en el módulo se introduce el concepto de Web semántica, tan en expansión hoy en día. Por un lado, se explican el funcionamiento y los estándares creados en torno a esta tecnología, basada en ofrecer, en la web, datos bien definidos, estructurados y enlazados, y, por el otro, cómo los sistemas de gestión de contenidos actuales lo están aplicando de manera progresiva. Por último, se exponen algunas de las iniciativas que se están llevando a cabo a escala internacional y se describen algunos de los objetivos previstos a corto plazo.

Objetivos

Con el estudio de este módulo, lograréis los objetivos siguientes:

- 1.** Entender la gestión documental como un beneficio a corto y largo plazo por parte de las empresas e instituciones que producen, difunden y comercializan contenidos y documentos.
- 2.** Ver cómo las herramientas actuales de gestión documental van más allá del tratamiento de la información, y se extienden desde el proceso de creación hasta el proceso de publicación y personalización de contenidos digitales.
- 3.** Tomar conciencia de que no todos los sistemas de gestión de contenidos actuales ofrecen servicios documentales, ni todos los sistemas de gestión documental ofrecen servicios de contenidos.
- 4.** Tener presente que tanto la gestión de contenidos como la gestión documental se tienen que llevar a cabo en entornos de trabajo colaborativos, para aprovechar al máximo todas las potencialidades de ambos sistemas.
- 5.** Conocer el flujo de trabajo y las tareas de control y acceso de un sistema de gestión de documentos multimedia.
- 6.** Conocer algunos ejemplos de sistemas de gestión documental con versiones de código abierto y de pago.
- 7.** Entender el concepto de Web semántica como un salto cualitativo de la web convencional.
- 8.** Comprender tanto su arquitectura como sus estándares, así como también conocer las principales iniciativas llevadas a cabo a escala internacional.
- 9.** Tomar conciencia de los enormes beneficios de esta nueva manera de entender la web, que pasan por la interoperabilidad, la optimización de recursos y el desarrollo de tecnologías abiertas.

1. Origen, funciones y arquitectura de los sistemas de gestión de contenidos (*CMS, Content Management Systems*)

La aparición de los sistemas de gestión de contenidos, conocidos como *CMS*, del inglés *Content Management Systems*, tiene lugar a partir del año 2000, en un momento en que hay una importante evolución tecnológica y una voluntad de explotación del entorno internet, tanto por parte de los usuarios como de las empresas. Es en este momento cuando las empresas ven en la Red un recurso tecnológico idóneo para mejorar su estructura y hacer más eficientes sus procesos internos. Simultáneamente, también es el momento en que las tecnologías web que permiten estos usos se están volviendo accesibles para empresas e instituciones. Las páginas web estáticas y poco articuladas empiezan a verse como insuficientes. Es por eso por lo que surge la necesidad de utilizar entornos más dinámicos y escalables, es decir, adaptables a los cambios. Ahora bien, además de una necesidad tecnológica, también hay una necesidad económica. Las empresas desean aplicar estas nuevas herramientas de manera rápida, sencilla y con un bajo coste.

Es en este contexto, pues, que aparecen los sistemas de gestión de contenidos. Un contexto en el que hay que añadir la gran variedad de documentos digitales que generan y contienen las empresas e instituciones. Documentos que a la vez contienen un gran volumen de datos textuales, numéricos y gráficos, de imágenes, de sonidos, de interacciones, etc., y que se presentan en varios formatos. Ante esta diversidad, los productores muchas veces tienen que optar por gestionar de manera segmentada, utilizando varios sistemas de gestión de contenidos.

Pero ¿qué es exactamente un sistema de gestión de contenidos o *CMS*?

Un *CMS* es una aplicación web que se usa para crear, editar, gestionar y publicar contenidos digitales (Lundgren, 2012).

Cabe decir que un *CMS* es un conjunto de aplicaciones informáticas muy articuladas que integran documentos en varios formatos convertidos en uno solo, como por ejemplo el XML; y también crean nuevos en este mismo formato. Estos documentos o recursos digitales resultantes reciben el nombre genérico de *contenidos*. Una de las ventajas de estos sistemas es que no son necesarios los conocimientos informáticos avanzados para gestionar los contenidos del

sitio web. Por otro lado, mantienen separados el contenido de su presentación final, hecho que permite obtener importantes beneficios en su explotación. Veamos cuáles son sus funciones principales (Tramullas, 2006):

a) Funciones de edición de contenidos

- Creación de contenidos
- Gestión de contenidos
- Publicación de contenidos
- Presentación de contenidos

b) Funciones de explotación de la web

- Entornos colaborativos
- Perfiles de usuarios
- Sindicación de contenidos
- Articulación de funcionalidades
- Integración de aplicaciones

Por otro lado, un CMS tiene que ofrecer, como mínimo, una arquitectura que soporte los procesos de publicación, de flujo de trabajo y de gestión de repositorios de información, así como también un repositorio de este tipo, herramientas de integración de información externa, y modelos y plantillas para los productos finales. Sería deseable, además, que los sistemas de gestión de contenidos incorporaran todas estas tecnologías, a pesar de que todavía hay las que no las prevén todas (Glazer, Jenkins y Schaper, 2005):

- **Gestión del conocimiento**, es decir, tener la capacidad de aplicar toda la información disponible para resolver cualquier problema o necesidad que pueda surgir en el entorno de la organización. Esta información tendrá que estar muy clasificada, indexada y organizada.
- **Busca y recuperación de la información**, es decir, tendría que proporcionar herramientas para la recuperación de la información que contienen los documentos de la organización, así como también integrar herramientas de *clustering* (ideal para el análisis de datos estadísticos, reconocimiento de patrones, etc.), taxonomías (vocabularios controlados necesarios para la indexación de los documentos), minería de datos y textos (extracción de información de un conjunto de datos y transformarla en una estructura comprensible para su uso posterior), etc.
- **Gestión documental**, es decir, tendría que dar soporte a la creación de nuevos documentos, así como también facilitar la recepción de otros documentos provenientes de fuentes externas. Asimismo, también debería facilitar el control de versiones y niveles de seguridad de acceso, así como

también gestionar flujos de trabajo y asignar documentos a tareas y procesos.

- **Gestión del ciclo de vida y del archivo de documentos**, es decir, definición y control de todas las fases en que se encuentra inmerso un documento, así como también establecimiento de programas de archivo, preservación digital y destrucción de documentos digitales.
- **Gestión de contenidos web**, es decir, ofrecer prestaciones que permitan la publicación rápida y controlada de los contenidos en la web (intranet, extranet, otros portales de la organización...).
- **Portales**, es decir, proporcionar a través de un cliente web el acceso a todas las aplicaciones y recursos de información que tenga la organización. Este portal tendrá que ofrecer servicios de agregación, integración y personalización.
- **Gestión de archivos digitales y documentos multimedia**, tanto los que producen como los que consumen las organizaciones. Se tendrán que tener muy en cuenta, no obstante, los derechos legales y las restricciones a que están sometidos.
- **Trabajo colaborativo**, es decir, entorno de trabajo que apoye proyectos y equipos virtuales a partir de herramientas de comunicación síncrona y asíncrona.

2. Sistemas de gestión de contenidos y sistemas de gestión documental (*DMS, Document Management Systems*)

La gestión de contenidos y la gestión de la información y la documentación son dos conceptos que no pueden entenderse de manera separada. Ahora bien, hay que decir que, hasta hace poco, tanto las empresas informáticas que han desarrollado aplicaciones para la gestión de contenidos, los conocidos CMS, como los proyectos y los equipos de gestión de contenidos de las empresas productoras no han tenido demasiado en cuenta los métodos y las técnicas de gestión documental, así como tampoco algunas tecnologías citadas en el apartado anterior. En estos últimos años, sin embargo, la situación ha ido cambiando y ya podemos decir que algunos de estos gestores de contenidos prevén muchas tareas documentales, tan necesarias para la buena difusión y recuperación de la información. Desde el año 2000, se han ido diseñando herramientas y aplicaciones que integran varias tecnologías (gestión de contenidos, gestión documental, gestión del conocimiento, preservación, busca y recuperación, etc.), y esto se debe a la necesidad que tienen muchas empresas y organizaciones de gestionar los ciclos de vida de sus documentos, para conseguir sus objetivos y llevar a cabo sus actividades de una manera más eficiente. Esta integración de funcionalidades y tecnologías incluso va más allá, y en algunos casos se produce también en plataformas de *ERP (Enterprise Resource Planning, es decir, sistemas de planificación de recursos empresariales)*, *CRM (Customer Relationship Management, es decir, gestión de la relación con los clientes)*, sistemas de aula virtual, etc. Poco a poco, pues, los sistemas de gestión documental se han ido integrando en entornos más amplios, como por ejemplo la gestión de la información en las empresas, sin olvidarse de su capacidad para gestionar y controlar todos aquellos procesos que tienen que ver con la creación, el almacenamiento, el tratamiento y la presentación de la información.

Queda todavía camino por recorrer. Ni todos los gestores de contenidos ofrecen servicios documentales, ni todos los gestores documentales ofrecen servicios de contenidos. Ahora bien, las organizaciones han de tener en cuenta que la información que contienen sus documentos es crucial. Es por eso por lo que no tendrían que perder de vista estas consideraciones (Asprey y Middleton, 2003):

- Siempre habrá la necesidad de gestionar documentos y sus contenidos.
- Si no se gestionan los documentos y contenidos web, las organizaciones pueden sufrir riesgos elevados.
- La calidad también puede verse afectada si hay una falta de gestión de estos documentos y contenidos web.
- Para poder llevar a cabo estrategias de gestión del conocimiento, la gestión de estos documentos y contenidos web es imprescindible.

- Integrar sistemas de gestión de contenidos con sistemas de flujo de trabajo permite rediseñar las estrategias y los procesos de negocio.
- El riesgo de amenazas en cuanto a la continuidad de las organizaciones y a sus actividades está siempre presente.
- El fracaso de los proyectos tiene un coste social elevado.
- La colaboración entre la gestión de documentos y la gestión de contenidos web es básica y esencial.

En definitiva, las herramientas actuales de gestión documental tratan los contenidos digitales como documentos. Además, van más allá del tratamiento de la información. Ahora se extienden desde el proceso de creación hasta el proceso de publicación y personalización. Por otro lado, tanto la gestión de contenidos como la gestión documental se tienen que llevar a cabo en entornos de trabajo colaborativos, para aprovechar al máximo todas las potencialidades de ambos sistemas.

3. Flujo de trabajo, control y acceso de un sistema de gestión de documentos multimedia (MDMS, *Multimedia Document Management Systems*)

Un sistema de gestión de documentos multimedia, conocido por MDMS (*Multimedia Document Management System*) y, según en qué entornos, también por MAM (*Media Asset Management*), gestiona los contenidos de los documentos multimedia generados por organizaciones, empresas e instituciones que se dediquen a difundirlos. Es importante remarcar que esta gestión se realiza en cualquier momento del ciclo de vida de estos documentos, y la manera en que se lleve a cabo dependerá de las características de cada organización. Veamos los tipos (Caridad y otros, 2011):

- Según la procedencia de sus contenidos: empresas que producen sus propios contenidos multimedia, que trabajan con contenidos ajenos, o las dos cosas.
- Según su **función**: empresas que producen, difunden, comercializan contenidos, etc.
- Según el **ámbito geográfico** en que trabaja: empresas de ámbito local, regional, estatal o internacional.
- Según la **relación** que tenga con las otras empresas: es decir, si forma parte de una multinacional, si tiene acuerdos de intercambio, si trabaja de manera autónoma, etc.

El ciclo de vida de un sistema de estas características abarca cinco fases: planificación, diseño conceptual y lógico, implantación, mantenimiento y evaluación. Con el tiempo, habría que añadir también la fase de decaída por pérdida de efectividad, y la fase de sustitución por un nuevo sistema.

A continuación, se explican las cinco acciones del flujo de trabajo de un MDMS o MAM:

1) Introducción de los contenidos en el sistema (ingesta), ya sean propios de la organización o ajenos. En cualquier caso, se tendrán que tener muy en cuenta los derechos de uso de estos contenidos.

2) Edición de los contenidos en tiempo real, es decir, directamente sobre estos, y de manera no lineal, sin tener que visualizar todo el contenido de un documento para seleccionar los fragmentos que se deseen editar. A menudo, las organizaciones convierten los contenidos multimedia en dos versiones (alta y baja calidad) y dependiendo del uso que se quiera hacer de ellos, se utiliza-

Lectura complementaria

Leed la guía que presenta la Norma ISO 15489 sobre el diseño e implementación de un sistema de gestión de documentos en una organización:

AENOR (2006). *UNE-ISO/TR 15489-2. Información y documentación-Gestión de documentos-Parte 2: Directrices*. Madrid: AENOR.

rá una u otra. Hay que decir que cada vez más estos sistemas almacenarán el contenido en la máxima calidad posible y generarán, bajo demanda, versiones de las calidades que convengan en cada momento, en función de los condicionantes tecnológicos y de uso.

3) Difusión de los contenidos de manera simultánea y en diferentes formatos. Las imágenes de un acontecimiento, por ejemplo, se pueden enviar en baja calidad a teléfonos móviles; en alta calidad, con sonido estereofónico y texto añadido a través de redes ópticas para que puedan ser visualizadas en una pantalla de ordenador; y en alta calidad y con posibilidad de seleccionar el idioma para que puedan ser recibidas en un televisor a través de TDT.

4) Archivo, donde se decide qué contenidos se tienen que almacenar. Esta acción se realizará tanto para garantizar la futura reutilización de estos contenidos como para preservarlos a corto o largo plazo. Estos contenidos serán almacenados en todas las fases, desde la ingesta, pasando por la edición y acabando por la difusión. Otra cuestión muy importante será el control de acceso que se establezca, ya que no todos los usuarios del sistema podrán acceder a estos contenidos del mismo modo y bajo las mismas condiciones.

5) Análisis¹ y tratamiento documental en todo el ciclo de vida de los contenidos. Esto significa que todos los contenidos serán identificados, descritos y etiquetados a través de metadatos que proporcionen información adicional.

⁽¹⁾ Los detalles de este análisis se explican en el módulo 5 de esta asignatura.

4. La gestión de la calidad en un sistema de gestión documental

En el año 2001, la ISO (International Organization for Standardization) publicó la Norma **ISO 15489 Records Management** sobre gestión de documentos, que contiene dos partes: la Norma **ISO 15489-1 General** y el informe técnico **ISO/TR 15489-2 Guidelines**. Asimismo, en el 2006, la AENOR (Asociación Española de Normalización y Certificación) publicó la traducción al español de las dos partes de esta norma. La Norma ISO 15489-1 establece cómo tiene que ser un sistema de gestión documental para garantizar la fiabilidad, la autenticidad, la integridad y la disponibilidad de los documentos. Así pues, explica qué etapas, procesos y controles se tienen que seguir para el diseño e implementación de un sistema de estas características. Las pautas son aplicables a cualquier organización pública o privada, y también sirven para cualquier persona responsable de crear y mantener documentos, afectando a toda actividad, proceso o tarea que se documente, y a cualquier documento creado o recibido sea cual sea su soporte y formato. Por otro lado, la norma se dirige tanto a los profesionales de la gestión de la información y la documentación como a los directivos y responsables de las organizaciones, con el objetivo de concienciarlos de las ventajas que ofrece la gestión de documentos y mostrarles las características que ha de tener una adecuada política de este tipo. Por otro lado, el informe técnico ISO/TR 15489-2 contiene una descripción más detallada de todo el proceso de implementación y de las funcionalidades de un sistema de gestión documental, y se dirige, sobre todo, al personal técnico encargado de emprender y dirigir este tipo de sistemas.

En el año 2001, la Comisión Europea divulgó una especificación complementaria, el **MoReq**, que describe los requisitos funcionales que tiene que tener una aplicación informática destinada a la gestión de documentos electrónicos. En definitiva, son unas pautas para el diseño de la aplicación y también una evaluación de aplicaciones que serán útiles para la gestión de estos tipos de documentos.

Lectura complementaria

Si queréis conocer más a fondo la Norma ISO 15489 y saber qué otras normas se han publicado sobre la gestión documental a lo largo de estos últimos años, leed el siguiente artículo:

J. A. Alonso (2007, septiembre-diciembre). "La Norma ISO 15489: un marco sistemático de buenas prácticas de gestión documental en las organizaciones". *Ítem* (núm. 47) (Revisado en julio del 2008).

http://eprints.rclis.org/12263/1/alonso_garcia_lloveras_-_la_norma_iso_15489.pdf

5. Ejemplos de sistemas de gestión de documentos multimedia

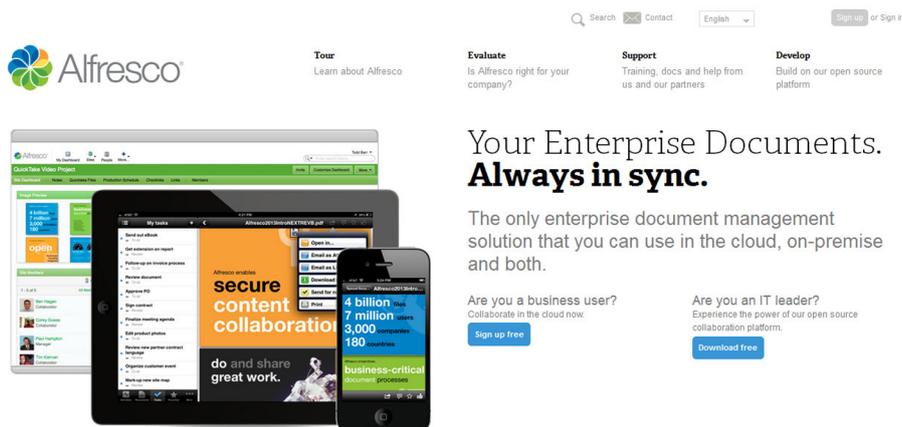
Los sistemas de gestión de documentos multimedia que se encuentran en el mercado suelen ofrecer las prestaciones clásicas, como por ejemplo definición de esquemas de metadatos, herramientas de control terminológico, indexación y recuperación del texto completo de los documentos. La alternativa más fácil para una organización, que no siempre resulta ser la más rentable, es la adquisición de una herramienta integrada, es decir, un paquete informático completo que incluya el software y el hardware. Estos paquetes simplifican *a priori* las decisiones técnicas, pero muchas veces no se adaptan al funcionamiento concreto de la organización que los compra. Es por este motivo que muchas veces se opta por el diseño de una aplicación hecha a medida que abrace todas las fases de un sistema de gestión documental. Este diseño se puede llevar a cabo por parte del personal técnico de la propia organización, por una empresa externa, o bien por la suma de profesionales de ambos lugares. No obstante, los principales inconvenientes de esta opción son la rentabilidad a medio plazo y el elevado coste económico que supone, solamente asequible para las grandes organizaciones. Veamos, a continuación, algunas de las aplicaciones que podemos encontrar en el mercado.

5.1. Alfresco

Alfresco (<http://www.alfresco.com>) es una plataforma de gestión documental de código abierto que funciona con Tomcat y utiliza MySQL. Se puede usar desde cualquier ubicación, es decir, tanto en las propias instalaciones de la organización como en dispositivos tipos tabletas o teléfonos inteligentes. Permite almacenar y compartir documentos que genera y custodia una empresa. Además, cuenta con herramientas sociales que permiten a sus usuarios compartir y descubrir nuevos contenidos. Por otro lado, se integra con MS Office, Apple iWork, Adobe Creative Suite, Google Docs, etc. En cuanto al esquema de precios, cabe decir que ofrece varias opciones:

- Versión gratuita que incluye 10 GB de almacenamiento.
- Versión Alfresco Community, descargable libremente y utilizada por los usuarios que quieran gestionar sus propios documentos.
- Versión Alfresco Standard Network, con soporte técnico, funciones administrativas y más almacenamiento.
- Versión Alfresco Enterprise Network, cuando se desea implementar el sistema en toda la empresa. Los precios parten de 10 dólares por usuario al mes.

Figura 1. Portada de Alfresco



5.2. Nuxeo

Nuxeo (<http://www.nuxeo.com>) es una plataforma de gestión de contenidos de código abierto. Ofrece gestión documental para contenidos multimedia, colaboración social (espacios sociales, perfiles con las actividades de los usuarios, mensajes y *microblogging*, etc.) y gestión de casos específicos (con buzones de distribución, ejemplos de metadatos, árboles de clasificación, etc.). También ofrece una sección con *plugins* y una herramienta de configuración y personalización de la aplicación.

5.3. KnowledgeTree

KnowledgeTree (<http://www.knowledgetree.com>) es un gestor documental de código abierto para la pequeña y mediana empresa. Está basado en una “pila” PHP, MySQL y Apache. Se presenta en dos versiones:

- Versión gratuita, funcional y muy completa.
- Versión comercial, que añade soporte técnico y un *plugin* para subir documentos trabajando directamente desde Microsoft Office.

Aparte de la configuración básica, permite realizar personalizaciones, como por ejemplo añadir el logotipo de la empresa, cambiar las opciones que se quieren visualizar en la página de inicio, etc. Ofrece, además, un repositorio central de documentos con control de versiones, la posibilidad de gestionar autorías, metadatos, flujos de trabajo de los documentos y herramientas de auditoría.

5.4. OpenKM

OpenKM (<http://www.openkm.com>) también es un gestor documental de código abierto. De entre los servicios que ofrece destacan el control de versiones, metadatos, tesauros, escaneo, seguridad, comentarios, foros sobre los documentos, flujo de trabajo, motor de busca, etc. Se trata de una aplicación Ja-

va J2EE que se ejecuta sobre Apache Tomcat. Se puede instalar y ejecutar en diferentes plataformas y, además de la versión gratuita, también ofrece una versión profesional de pago.

Figura 2. Portada de OpenKM



Lectura complementaria

Leed este artículo publicado en *La Vanguardia* el 19 de julio del 2011 por Eva Domínguez, en el que se describen cuatro editores y gestores de contenidos multimedia especializados en diferentes ámbitos de aplicación: Documenta (ámbito educativo), Klynt y Zeega (para documentales interactivos) y 3WDOC (para documentales web):

<http://blogs.lavanguardia.com/elcuartobit/editores-de-contenidos-multimedia>

Lectura complementaria

Si queréis saber más sobre la plataforma Documenta, leed la entrevista realizada en junio del 2011 a Cristina Casanova y Andrea Contino y publicada en la revista *Mosaic*, editada por la UOC:

<http://mosaic.uoc.edu/2011/06/30/cristina-casanova-y-andrea-contino>

6. El concepto de Web semántica

6.1. ¿Qué es la Web semántica?

Del mismo modo que la Web 2.0, hablamos de Web semántica para referirnos a un uso concreto de un conjunto de herramientas y tecnologías. En el año 2002, James Hendler, Tim Berners-Lee y Eric Miller establecieron la siguiente definición:

“La Web semántica es una extensión de la actual web en que a la información disponible se le otorga un significado muy definido que permita a los ordenadores y las personas trabajar en colaboración. Se basa en la idea de ofrecer en la web datos definidos y enlazados, permitiendo que aplicaciones heterogéneas localicen, integren, razonen y reutilicen la información presente en la web”.

El W3C (World Wide Web Consortium) la define de la manera siguiente:

“La Web semántica es la representación de datos en la web. Es un esfuerzo colaborativo liderado por W3C con la participación de un gran número de investigadores y socios industriales. Se basa en el uso de RDF, que integra una gran variedad de aplicaciones mediante el uso de XML para la sintaxis y el uso de URI para su identificación”.

Partiendo de esta última definición, el *RDF* (*Resource Description Framework*) es un modelo de datos que ofrece una especificación para la descripción de metadatos en la web. Organiza la información en forma de “sujeto-predicado-objeto” y permite expresarla sintácticamente mediante *XML* (*Extensible Markup Language*). Además, utiliza la expresión *URI* (*Uniform Resource Identifier*) para identificar, de manera universal, un espacio de nombres de recursos de información. En definitiva, la Web semántica va mucho más allá que la web convencional. Utiliza metadatos para describir los recursos de información, parte de vocabularios RDF para representar estos metadatos, desarrolla esquemas RDF u ontologías para describir las relaciones entre los recursos que se han descrito y las propiedades utilizadas para caracterizarlos, localiza, interconecta y reutiliza fuentes de datos RDF e infiere nueva información a partir de las relaciones lógicas que se establecen entre los datos.

6.2. Arquitectura y estándares de la Web semántica

Como sabemos, las tecnologías básicas de una web convencional son las siguientes:

- Localización de objetos y recursos a través de URL.
- Utilización del protocolo HTTP para establecer la comunicación entre clientes y servidores.
- Marcaje de los documentos con el lenguaje HTML.

La Web semántica, en cambio, localiza los objetos y recursos a través de URI y extiende el HTML con tecnologías como RDF para almacenar y describir los datos. El protocolo HTTP sigue siendo, a bajo nivel dentro de esta arquitectura, el encargado de la transmisión de datos entre clientes y servidores.

A continuación, analizamos más a fondo los estándares de la Web semántica (Pastor Sánchez, 2011):

- **XML²**: metalenguaje para la definición de vocabularios orientados al intercambio de datos estructurados, independientemente de la plataforma con la que se ejecuten las aplicaciones que procesan estos datos. Un documento XML contiene datos cuya estructura se define a partir de dos métodos: el *DTD (Document Type Definition)* o el XML Schema, considerado el sucesor del anterior porque ofrece definiciones más expresivas y exactas. Los diferentes elementos se estructuran en torno a un elemento raíz y los valores de los atributos se marcan entre comillas (dobles o simples). Por otro lado, este lenguaje también permite añadir datos literales y comentarios. En definitiva, con XML se consigue un salto cualitativo, puesto que permite separar estructura, contenido y formato visual.
- **RDF³**: modelo para la descripción estructurada de recursos de información en la Red. Como ya hemos comentado anteriormente, las descripciones se realizan del tipo “sujeto-predicado-objeto” y se denominan sentencias. El “sujeto” es el recurso que se tiene que describir, el “predicado” es una propiedad o relación del recurso y el “objeto” es el valor que se le asigna a esta propiedad o el recurso a partir del cual se establece la relación. Este modelo trata de dar respuesta a ciertas limitaciones planteadas por la estructura jerárquica del lenguaje XML.
- **RDFa⁴ (Resource Description Framework -in- Attributes)**: incluir información semántica en un sitio web también se puede hacer de manera más sencilla que a partir del modelo RDF. Esto es posible gracias a RDFa, que permite añadir declaraciones RDF en forma de atributos en etiquetas con código (X)HTML. De este modo, coexisten dos versiones en una misma URL: una

⁽²⁾Si queréis saber más sobre XML, consultad el enlace <http://www.w3.org/xml>

⁽³⁾Si queréis saber más sobre RDF, consultad el enlace <http://www.w3.org/rdf>

⁽⁴⁾Si queréis saber más sobre RDFa, consultad los enlaces <http://www.w3.org/tr/xhtml1-rdfa-primer> y <http://www.w3.org/tr/rdfa-syntax>

adaptada para ser consultada por personas, y la otra adaptada para su procesamiento a través de máquinas.

- **RDF Schema**⁵: el modelo RDF permite describir y representar, de manera explícita, características de todo tipo de recursos. Esto facilita la comprensión precisa de conjuntos de datos por parte de las máquinas, que posteriormente pueden ser procesados para extraer información o ejecutar procesos. El modelo RDFS (RDF Schema) es la ampliación del modelo RDF, enriqueciéndolo con un vocabulario que tiene cierta semántica. Con RDFS se pueden describir esquemas sencillos utilizando clases y subclasses, así como también definir propiedades, su dominio de aplicación y su rango de valores posibles.
- **OWL**⁶ (*Web Ontology Language*): lenguaje que permite publicar y compartir, en la web, ontologías que definen los términos y las relaciones de un vocabulario. Se trata de una extensión de los modelos RDF y RDFS, no es un lenguaje diferente. Las ontologías son muy importantes en la Web semántica. Son mecanismos muy potentes, sin los cuales los agentes de software no serían capaces de procesar los datos para la obtención de información. Por otro lado, definen las clases, las propiedades y los individuos (instancias), junto con las subclasses, las subpropiedades, el dominio, el rango y otros aspectos, como la relación inversa entre propiedades y un conjunto de axiomas adicionales para definir restricciones. Incorporan, además, enumeraciones de instancias de clase y combinaciones de clase.
- **SPARQL**⁷ (*SPARQL Protocol and RDF Query Language*): lenguaje de interrogación que desde el 2008 se ha convertido en estándar oficial del W3C para la recuperación de sentencias a partir de fuentes de datos RDF. Permite expresar patrones del tipo “sujeto-predicado-objeto” que se utilizan para encontrar correspondencias en el conjunto de sentencias RDF.

⁽⁵⁾ Si queréis saber más sobre RDF Schema, consultad el enlace <http://www.w3.org/tr/rdf-schema>

⁽⁶⁾ Si queréis saber más sobre OWL, consultad el enlace <http://www.w3.org/2004/owl>

⁽⁷⁾ Si queréis saber más sobre SPARQL, consultad el enlace <http://www.w3.org/tr/rdf-sparql-query>

6.3. La Web semántica y los sistemas de gestión de contenidos

Los sistemas de gestión de contenidos son fundamentales en el despliegue de la Web semántica, ya que se basan en procesos de organización y estructuración de la información (la mayoría utilizan bases de datos para almacenar los contenidos, separando estructura, información y presentación). Este elevado nivel de estructuración de la información permite que esta se pueda reutilizar para añadir atributos RDF al código (X)HTML de las páginas. Algunos sistemas como Drupal⁸, por ejemplo, introducen de manera automática estos atributos justo durante el proceso de renderizado del código (X)HTML. Este mismo y otros, como Liferay, permiten añadir módulos que ofrecen contenidos en formato RDF.

⁽⁸⁾ Si queréis saber cómo Drupal añade atributos RDF al código (X)HTML, consultad el siguiente enlace: <http://drupal.org/project/rdfx>

Sea como fuere, es necesario mapear los elementos de los contenidos que gestiona un CMS en sus correspondientes elementos de vocabularios RDF. Se trata de un proceso similar a cuando se definen los canales RSS desde un sitio web, pero mucho más complicado, puesto que se tiene que tener muy claro qué vocabulario es el más adecuado para representar un determinado campo de un tipo de contenido, además de qué nivel de detalle y agregación. Por otro lado, los CMS permiten incorporar datos externos RDF, e incluso la recuperación selectiva de datos a través del lenguaje SPARQL (Pastor Sánchez, 2011).

6.4. Proyectos e iniciativas internacionales en torno a la Web semántica

A escala internacional, se han desarrollado varias aplicaciones e iniciativas en torno a la Web semántica (Pastor Sánchez, 2011):

- **LOD⁹ (Linked Open Data)**: conocida como “datos abiertos entrelazados”, es una técnica que permite publicar información estructurada en la web. Este proceso se realiza a través de la tecnología web, pero con un cambio: los datos están pensados para que los programas informáticos puedan leerlos, interpretarlos y procesarlos automáticamente. Cada vez más, las empresas privadas y las administraciones públicas utilizan esta técnica para gestionar sus datos.
- **SKOS¹⁰ (Simple Knowledge Organization System)**: modelo para representar varios tipos de vocabularios controlados, como por ejemplo tesauros, clasificaciones, taxonomías o listas de encabezamientos de materia. Es una ontología OWL Full simple, basada en RDF, que permite representar diferentes tipos de lenguajes documentales a través de clases y propiedades.
- **Europeana¹¹**: desde el año 2008, esta iniciativa pretende ofrecer una única puerta de entrada a una gran cantidad de recursos, resultado de la digitalización de varios objetos culturales por parte de más de 1.500 instituciones de la Unión Europea (bibliotecas, archivos, museos, organizaciones culturales...). Estos objetos culturales se almacenan en un catálogo centralizado, elaborado a partir de la recopilación de metadatos de repositorios externos. Estos metadatos son mapeados en un único esquema de descripción denominado *ESE (Europeana Semantic Elements)*. Ahora bien, para no perder parte de los metadatos originales, se utiliza el modelo *EDM (Europeana Data Model)*. Por otro lado, el uso de RDF en los procesos de representación de la información y en la definición de vocabularios ha hecho que Europeana se incorpore a iniciativas como Linked Open Data, antes comentada, y por lo tanto, a la Web semántica.

⁽⁹⁾Si queréis saber más sobre LOD, consultad el enlace <http://www.w3.org/wiki/linkedata>

Enlaces recomendados

Encontraréis ejemplos de LOD en el registro de fuentes de datos (*data-sets*) de la Open Knowledge Foundation <http://thedatahub.org> y en el catálogo de la Fundación CTIC Centro Tecnológico <http://datos.fundacionctic.org/sandbox/catalog/faceted>

⁽¹⁰⁾Si queréis saber más sobre LOD, consultad el enlace <http://www.w3.org/tr/skos-reference>

⁽¹¹⁾Si queréis saber más sobre Europeana, consultad el enlace <http://www.europeana.eu>

Enlace recomendado

La versión estadounidense de Europeana es la Digital Public Library of America: <http://dp.la>.

- **DBpedia**¹²: iniciativa emprendida en el 2007 que permite la obtención de datos estructurados a partir de los contenidos de la Wikipedia. Para referirse a personas, lugares, trabajos, especies u organizaciones, DBpedia utiliza diferentes subclases de recursos. Cada clase tiene asociada una serie de propiedades que permiten la descripción de objetos. Los datos de DBpedia se pueden consultar en forma de páginas, donde se muestran de manera estructurada las diferentes sentencias sobre un concepto, fruto de aplicar la ontología de mapeado. Por otro lado, es la iniciativa con mayor presencia en Linked Open Data.
- **Calais**¹³: iniciativa de Reuters que convierte información de blogs, CMS, webs o aplicaciones en información semántica (formato RDF). Se basa en ClearForest, una empresa de software que desarrolla análisis de textos y soluciones de minería de textos.

⁽¹²⁾Si queréis saber más sobre DBpedia, consultad el enlace <http://wiki.dbpedia.org>

Enlace recomendado

Si queréis descargaros el *dataset* de DBpedia, consultad el enlace <http://wiki.dbpedia.org/downloads37>.

⁽¹³⁾Si queréis saber más sobre Calais, consultad el enlace <http://www.opencalais.com>

6.5. El futuro de la Web semántica

La Web semántica apuesta por la interoperabilidad, la optimización de recursos y el desarrollo de tecnologías abiertas que benefician a todos los usuarios de la Red. Sus posibilidades son enormes y de esto se han dado cuenta tanto los fabricantes de software como los diseñadores de sistemas, los investigadores e incluso los políticos. La tecnología Linked Open Data, por ejemplo, permite que los contenidos estén interconectados, así como también que los datos se adapten a procesos inteligentes y dispositivos de todo tipo. Actualmente, los usuarios pueden interconectar los ordenadores, las tabletas y los teléfonos inteligentes a través de servicios con datos y programas ubicados en la “nube” y que usan estándares de la Web semántica para almacenar la información. Los CMS, además, también incorporan tecnologías de la Web semántica y cada día se desarrollan nuevas ontologías que formalizan objetos de información para ser publicados en la Red.

Los objetivos a corto plazo, posiblemente, giran en torno a aumentar la cantidad de fuentes de datos (*datasets*) abiertas y de enlaces relacionados, y de mejorar los resultados de los motores de busca basados en la Web semántica, así como también de integrar las redes sociales en esta tecnología. Incluso se habla de una evolución a sistemas más complejos, próximos a la inteligencia artificial.

Bibliografía

Bibliografía básica

Asprey, L.; Middleton, M. (2003). *Integrative document and content management: strategies for exploiting enterprise knowledge*. Hershey: Idea Group.

Caridad, M; Hernández, T.; Rodríguez, D.; Pérez, B. (2011). *Documentación audiovisual: nuevas tendencias en el entorno digital*. Madrid: Síntesis.

Glazer, D.; Jenkins, T.; Schaper, H. (2005). *Enterprise content management technology: what you need to know*. Waterloo: Open Text Corporation.

Lundgren, D. J. (2012). *CMS introducció*. Barcelona: UOC.

Pastor Sánchez, J. A. (2011). *Tecnologías de la Web semántica*. Barcelona: UOC.

Tramullas, J. (coord.) (2006). *Tendencias en documentación digital*. Gijón: Trea.

Bibliografía complementaria

AENOR (2006). *UNE-ISO 15489-1. Información y documentación -Gestión de documentos- Parte 1: Generalidades*. Madrid: AENOR.

AENOR (2006). *UNE-ISO/TR 15489-2. Información y documentación -Gestión de documentos- Parte 2: Directrices*. Madrid: AENOR.

Alonso, J.A. (2007, septiembre-diciembre). "La norma ISO 15489: un marco sistemático de buenas prácticas de gestión documental en las organizaciones". *Ítem* (núm. 47). (Revisado en julio del 2008). <http://eprints.rclis.org/12263/1/Alonso_Garcia_Lloveras_-_La_norma_ISO_15489.pdf>

Casanova, C.; Contino, A. (2011, 30 de junio). "Entrevista a Cristina Casanova y Andrea Contino, creadores de la Plataforma Documenta". *Mosaic*. <<http://mosaic.uoc.edu/2011/06/30/cristina-casanova-y-andrea-contino>>

Domínguez, E. (2011, 19 de julio). "Editores de contenidos multimedia". *El cuarto bit*. <<http://blogs.lavanguardia.com/elcuartobit/editores-de-contenidos-multimedia>>

Enlaces

Alfresco: <http://www.alfresco.com>

Calais: <http://www.opencalais.com>

DBpedia: <http://wiki.dbpedia.org> y <http://wiki.dbpedia.org/Downloads37>

Digital Public Library of America: <http://dp.la>

Drupal: <http://drupal.org/project/rdfx>

Europeana: <http://www.europeana.eu>

Fundación CTIC Centro Tecnológico (*datasets*): <http://datos.fundacionctic.org/sandbox/catalog/faceted>

KnowledgeTree: <http://www.knowledgetree.com>

Linked Open Data: <http://linkeddata.org> y <http://www.w3.org/wiki/LinkedData>

Nuxeo: <http://www.nuxeo.com>

Open Knowledge Foundation (*datasets*): <http://thedatahub.org>

OpenKM: <http://www.openkm.com>

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

RDF Schema: <http://www.w3.org/TR/rdf-schema>

RDFa: <http://www.w3.org/TR/xhtml-rdfa-primer>

<http://www.w3.org/TR/rdfa-syntax>

SKOS: <http://www.w3.org/TR/skos-reference>

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