

Els sistemes de gestió de documents multimèdia

Noemí Mases Blanch

PID_00202574



Els textos i imatges publicats en aquesta obra estan subjectes –llevat que s'indiqui el contrari– a una llicència de Reconeixement-NoComercial-SenseObraDerivada (BY-NC-ND) v.3.0 Espanya de Creative Commons. Podeu copiar-los, distribuir-los i transmetre'ls públicament sempre que en citeu l'autor i la font (FUOC. Fundació per a la Universitat Oberta de Catalunya), no en feu un ús comercial i no en feu obra derivada. La llicència completa es pot consultar a <http://creativecommons.org/licenses/by-nc-nd/3.0/es/legalcode.ca>

Índex

Introducció	5
Objectius	6
1. Origen, funcions i arquitectura dels sistemes de gestió de continguts (CMS, Content Management Systems)	7
2. Sistemes de gestió de continguts i sistemes de gestió documental (DMS, Document Management Systems)	10
3. Flux de treball, control i accés d'un sistema de gestió de documents multimèdia (MDMS, Multimedia Document Management Systems)	12
4. La gestió de la qualitat en un sistema de gestió documental.	14
5. Exemples de sistemes de gestió de documents multimèdia	15
5.1. Alfresco	15
5.2. Nuxeo	16
5.3. KnowledgeTree	16
5.4. OpenKM	16
6. El concepte de Web semàntic	18
6.1. Què és el Web semàntic?	18
6.2. Arquitectura i estàndards del Web semàntic	19
6.3. El Web semàntic i els sistemes de gestió de continguts	20
6.4. Projectes i iniciatives internacionals entorn del Web semàntic	21
6.5. El futur del Web semàntic	22
Bibliografia	23

Introducció

Actualment, les empreses i institucions tenen la necessitat de gestionar els cicles de vida dels seus documents, independentment dels suports i formats, per aconseguir els seus objectius i dur a terme les seves activitats i processos d'una manera més eficient. Moltes d'aquestes organitzacions ja disposen de sistemes de gestió de continguts, coneguts com a CMS (*Content Management Systems*) que creen, editen, gestionen i publiquen continguts digitals. Ara bé, no són tantes les que preveuen altres tecnologies, mètodes i tècniques que tenen a veure amb la gestió documental, com ara la gestió del coneixement, la gestió del cicle de vida dels documents o la preservació digital, tan necessaris per a la bona difusió i recuperació de la informació, i també per al funcionament òptim dels fluxos de treball dins de l'entorn organitzatiu.

Aquest mòdul pretén incidir en el fet que la gestió de continguts i la gestió de la informació i la documentació en les organitzacions són dos conceptes crucials i no es poden entendre de manera separada. En el dos primers apartats s'expliquen les funcionalitats i l'arquitectura dels sistemes de gestió de continguts tradicionals i s'enumeren les tecnologies que haurien de tenir per a funcionar, també, com a sistemes de gestió documental. En el tercer apartat s'analitzen a fons els fluxos de treball d'aquests sistemes de gestió documental, que evidentment suporten i gestionen documents multimèdia i, en darrer lloc, es fa una breu introducció de com es du a terme la gestió de la qualitat en aquests entorns, i es mostren alguns exemples d'aplicacions que podem trobar actualment en el mercat.

Finalment, en el mòdul s'introdueix el concepte de Web semàntic, tant en expansió avui dia. D'una banda, s'expliquen el funcionament i els estàndards creats entorn d'aquesta tecnologia, basada a oferir en el web dades ben definides, estructurades i enllaçades i, de l'altra, com els sistemes de gestió de continguts actuals l'estan aplicant de manera progressiva. Finalment, s'exposen algunes de les iniciatives que s'estan duent a terme a escala internacional i es descriuen alguns dels objectius previstos a curt termini.

Objectius

Amb l'estudi d'aquest mòdul, assolireu els objectius següents:

- 1.** Entendre la gestió documental com un benefici a curt i llarg termini de les empreses i institucions que produeixen, difonen i comercialitzen continguts i documents.
- 2.** Veure com les eines actuals de gestió documental van més enllà del tractament de la informació, i s'estenen des del procés de creació fins al procés de publicació i personalització de continguts digitals.
- 3.** Prendre consciència que no tots els sistemes de gestió de continguts actuals ofereixen serveis documentals, ni tots els sistemes de gestió documental ofereixen serveis de continguts.
- 4.** Tenir present que tant la gestió de continguts com la gestió documental s'han de dur a terme en entorns de treball col·laboratiu, per a aprofitar al màxim totes les potencialitats de tots dos sistemes.
- 5.** Conèixer el flux de treball i les tasques de control i accés d'un sistema de gestió de documents multimèdia.
- 6.** Conèixer alguns exemples de sistemes de gestió documental amb versions de codi obert i de pagament.
- 7.** Entendre el concepte de Web semàntic com un salt qualitatiu pel que fa al web convencional.
- 8.** Comprendre tant la seva arquitectura com els seus estàndards, i també conèixer les principals iniciatives dutes a terme a escala internacional.
- 9.** Prendre consciència dels enormes beneficis d'aquesta nova manera d'entendre el Web, que impliquen la interoperabilitat, l'optimització de recursos i el desenvolupament de tecnologies obertes.

1. Origen, funcions i arquitectura dels sistemes de gestió de continguts (CMS, *Content Management Systems*)

L'aparició dels sistemes de gestió de continguts, coneguts com a CMS, de l'anglès *Content Management Systems*, té lloc a partir de l'any 2000, en un moment en què hi ha una important evolució tecnològica i una voluntat d'exploració de l'entorn Internet, tant dels usuaris com de les empreses. És en aquest moment quan les empreses veuen a la Xarxa un recurs tecnològic idoni per a millorar la seva estructura i fer més eficients els seus processos interns. Simultàniament, també és el moment en què les tecnologies web que permeten aquests usos s'estan tornant accessibles per a empreses i institucions. Les pàgines web estàtiques i poc articulades es comencen a veure com a insuficients. És per això que sorgeix la necessitat d'utilitzar entorns més dinàmics i escalables, és a dir, adaptables als canvis. Ara bé, a banda d'una necessitat tecnològica, també hi ha una necessitat econòmica. Les empreses volen aplicar aquestes noves eines de manera ràpida, senzilla i amb un baix cost.

És en aquest context, doncs, que apareixen els sistemes de gestió de continguts. Un context en el qual cal afegir la gran varietat de documents digitals que generen i contenen les empreses i institucions. Documents que a la vegada contenen un gran volum de dades textuais, numèriques i gràfiques, d'imatges, de sons, d'interaccions, etc., i que es presenten en diversos formats. Davant aquesta diversitat, els productors moltes vegades han d'optar per gestionar-ho de manera segmentada, utilitzant diversos sistemes de gestió de continguts.

Però què és exactament un sistema de gestió de continguts o CMS?

Un CMS és una aplicació web que es fa servir per a crear, editar, gestionar i publicar continguts digitals (Lundgren, 2012).

Val a dir que un CMS és un conjunt d'aplicacions informàtiques ben articulades que integren documents en diversos formats convertits en un de sol, com per exemple l'XML, i també en creen de nous en aquest mateix format. Aquests documents o recursos digitals resultants reben el nom genèric de *continguts*. Un dels avantatges d'aquests sistemes és que no són necessaris coneixements informàtics avançats per a gestionar els continguts del lloc web. D'altra banda, mantenen separats el contingut de la seva presentació final, fet que permet obtenir importants beneficis en l'exploració. Vegem, però, quines són les seves funcions principals (Tramullas, 2006):

a) Funcions d'edició de continguts

- Creació de continguts
- Gestió de continguts
- Publicació de continguts
- Presentació de continguts

b) Funcions d'explotació del web

- Entorns col·laboratius
- Perfils d'usuaris
- Sindicació de continguts
- Articulació de funcionalitats
- Integració d'aplicacions

D'altra banda, un CMS ha d'oferir, com a mínim, una arquitectura que suporti els processos de publicació, de flux de treball i de gestió de repositoris d'informació, i també un repositori d'aquest tipus, eines d'integració d'informació externa, i models i plantilles per als productes finals. Seria desitjable, a més, que els sistemes de gestió de continguts incorporessin totes aquestes tecnologies, tot i que encara n'hi ha que no les preveuen totes (Glazer, Jenkins i Schaper, 2005):

- **Gestió del coneixement**, és a dir, tenir la capacitat d'aplicar tota la informació disponible per a resoldre qualsevol problema o necessitat que pugui sorgir en l'entorn de l'organització. Aquesta informació ha d'estar ben classificada, indexada i organitzada.
- **Cerca i recuperació de la informació**, és a dir, hauria de proporcionar eines per a la recuperació de la informació que contenen els documents de l'organització, i també integrar eines de clusterització (*clustering*) (ideal per a l'anàlisi de dades estadístiques, reconeixement de patrons, etc.), taxonomies (vocabularis controlats necessaris per a la indexació dels documents), mineria de dades i textos (extreure informació d'un conjunt de dades i transformar-la en una estructura comprensible per a usar-la posteriorment), etc.
- **Gestió documental**, és a dir, hauria de donar suport a la creació de nous documents, i també hauria de facilitar la recepció d'altres documents provinents de fonts externes. D'altra banda, també hauria de facilitar el control de versions i nivells de seguretat d'accés, i també gestionar fluxos de treball i assignar documents a tasques i processos.
- **Gestió del cicle de vida i de l'arxiu de documents**, és a dir, definició i control de totes les fases en què es troba immers un document, i també

establiment de programes d'arxiu, preservació digital i destrucció de documents digitals.

- **Gestió de continguts web**, és a dir, oferir prestacions que permetin la publicació ràpida i controlada dels continguts en el web (intranet, extranet, altres portals de l'organització...).
- **Portals**, és a dir, proporcionar per mitjà d'un client web l'accés a totes les aplicacions i recursos d'informació que tingui l'organització. Aquest portal haurà d'oferir serveis d'agregació, integració i personalització.
- **Gestió d'arxius digitals i documents multimèdia**, tant els que produeixen com els que consumeixen les organitzacions. S'hauran de tenir molt en compte, però, els drets legals i les restriccions a què sovint estan sotmesos.
- **Treball col·laboratiu**, és a dir, entorn de treball que doni suport a projectes i equips virtuals a partir d'eines de comunicació síncrona i asíncrona.

2. Sistemes de gestió de continguts i sistemes de gestió documental (DMS, *Document Management Systems*)

La gestió de continguts i la gestió de la informació i la documentació són dos conceptes que no es poden entendre de manera separada. Ara bé, val a dir que, fins fa poc, tant les empreses informàtiques que han desenvolupat aplicacions per a la gestió de continguts, els coneguts CMS, com els projectes i els equips de gestió de continguts de les empreses productores, no han tingut massa en compte els mètodes i les tècniques de gestió documental, i també algunes tecnologies esmentades en l'apartat anterior. En aquests darrers anys, però, la situació ha anat canviant i ja podem dir que alguns d'aquests gestors de continguts preveuen moltes tasques documentals, tan necessàries per a la bona difusió i recuperació de la informació. Des de l'any 2000, s'han anat dissenyant eines i aplicacions que integren diverses tecnologies (gestió de continguts, gestió documental, gestió del coneixement, preservació, cerca i recuperació, etc.), i això es deu a la necessitat que tenen moltes empreses i organitzacions de gestionar els cicles de vida dels seus documents, per a aconseguir els seus objectius i dur a terme les seves activitats d'una manera més eficient. Aquesta integració de funcionalitats i tecnologies fins i tot va més enllà, i en alguns casos es produeix també en plataformes d'ERP (*Enterprise Resource Planning*, és a dir, sistemes de planificació de recursos empresarials), CRM (*Customer Relationship Management*, és a dir, gestió de la relació amb els clients), sistemes d'aula virtual, etc. A poc a poc, doncs, els sistemes de gestió documental s'han anat integrant en entorns més amplis, com ara la gestió de la informació en les empreses, sense oblidar-se de la seva capacitat per a gestionar i controlar tots aquells processos que tenen a veure amb la creació, l'emmagatzematge, el tractament i la presentació de la informació.

Queda encara camí per recórrer. Ni tots els gestors de continguts ofereixen serveis documentals, ni tots els gestors documentals ofereixen serveis de continguts. Ara bé, les organitzacions han de tenir en compte que la informació que contenen els seus documents és crucial. És per això que no haurien de perdre de vista aquestes consideracions (Asprey i Middleton, 2003):

- Sempre hi haurà la necessitat de gestionar documents i els seus continguts.
- Si no es gestionen els documents i continguts web, les organitzacions poden patir riscos elevats.
- La qualitat també es pot veure afectada si hi ha una manca de gestió d'aquests documents i continguts web.
- Per a poder dur a terme estratègies de gestió del coneixement, la gestió d'aquests documents i continguts web és imprescindible.
- Integrar sistemes de gestió de continguts amb sistemes de flux de treball permet redissenyar les estratègies i els processos de negoci.

- El risc d'amenaques pel que fa a la continuïtat de les organitzacions i a les seves activitats hi és sempre present.
- El fracàs dels projectes té un cost social elevat.
- La col·laboració entre la gestió de documents i la gestió de continguts web és bàsica i essencial.

En definitiva, les eines actuals de gestió documental tracten els continguts digitals com a documents. A més, van més enllà del tractament de la informació. Ara s'estenen des del procés de creació fins al procés de publicació i personalització. D'altra banda, tant la gestió de continguts com la gestió documental s'han de dur a terme en entorns de treball col·laboratiu, per a aprofitar al màxim totes les potencialitats de tots dos sistemes.

3. Flux de treball, control i accés d'un sistema de gestió de documents multimèdia (MDMS, *Multimedia Document Management Systems*)

Un sistema de gestió de documents multimèdia, conegut com a MDMS (*Multimedia Document Management System*) i, segons en quins entorns, també com a MAM (*Media Asset Management*), gestiona els continguts dels documents multimèdia generats per organitzacions, empreses i institucions que es dediquin a difondre'ls. És important remarcar que aquesta gestió es fa en qualsevol moment del cicle de vida d'aquests documents, i la manera com es dugui a terme dependrà de les característiques de cada organització. Vegem-ne els tipus (Caridad i altres, 2011):

- Segons la **procedència** dels seus continguts: empreses que produeixen els seus propis continguts multimèdia, que treballen amb continguts aliens, o totes dues coses.
- Segons la **funció**: empreses que produeixen, difonen, comercialitzen continguts, etc.
- Segons l'**àmbit geogràfic** en què treballa: empreses d'àmbit local, regional, estatal o internacional.
- Segons la **relació** que tingui amb les altres empreses: és a dir, si forma part d'una multinacional, si té acords d'intercanvi, si treballa de manera autònoma, etc.

El cicle de vida d'un sistema d'aquestes característiques comprèn cinc fases: planificació, disseny conceptual i lògic, implantació, manteniment i avaluació. Amb el temps, caldria afegir també la fase de decaiguda per pèrdua d'efectivitat, i la fase de substitució per un nou sistema.

A continuació, s'expliquen les cinc accions del flux de treball d'un MDMS o MAM:

1) Introducció dels continguts en el sistema (ingesta), siguin propis de l'organització o aliens. En qualsevol cas, s'hauran de tenir molt en compte els drets d'ús d'aquests continguts.

2) Edició dels continguts en temps real, és a dir, directament sobre aquests, i de manera no lineal, és a dir, sense haver de visualitzar tot el contingut d'un document per a seleccionar els fragments que es vulguin editar. Sovint, les organitzacions converteixen els continguts multimèdia a dues versions (alta i baixa qualitat), i depenent de l'ús que se'n vulgui fer s'utilitzarà una o l'altra. Val a dir que cada vegada més aquests sistemes emmagatzemaran el contin-

Lectura complementària

Llegiu la guia que presenta la norma ISO 15489 sobre el disseny i implementació d'un sistema de gestió de documents en una organització:

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

gut en la màxima qualitat possible i generaran, sota demanda, versions de les qualitats que convinguin en cada moment, en funció dels condicionants tecnològics i d'ús.

3) Difusió dels continguts de manera simultània i en diferents formats. Les imatges d'un esdeveniment, per exemple, es poden enviar en baixa qualitat a telèfons mòbils; en alta qualitat, amb so estereofònic i text afegit a través de xarxes òptiques perquè puguin ser visualitzades en una pantalla d'ordinador; i en alta qualitat i amb possibilitat de seleccionar l'idioma perquè puguin ser rebudes en un televisor per mitjà de TDT.

4) Arxiu, en què es decideix quins continguts s'han d'emmagatzemar. Aquesta acció es fa tant per garantir la reutilització futura d'aquests continguts, com per preservar-los a curt o llarg termini. Val a dir que aquests continguts han de ser emmagatzemats en totes les fases, des de la ingesta, passant per l'edició i acabant per la difusió. Una altra qüestió molt important serà el control d'accés que s'estableixi, ja que no tots els usuaris del sistema podran accedir a aquests continguts de la mateixa manera i sota les mateixes condicions.

5) Anàlisi¹ i tractament documental en tot el cicle de vida dels continguts. Això significa que tots els continguts han de ser identificats, descrits i etiquetats per mitjà de metadades que proporcionin informació addicional.

⁽¹⁾Els detalls d'aquesta anàlisi s'expliquen en el mòdul 5 d'aquesta assignatura.

4. La gestió de la qualitat en un sistema de gestió documental

L'any 2001, la ISO (International Organization for Standardization) va publicar la norma **ISO 15489 Records Management** sobre gestió de documents, que conté dues parts: la norma **ISO 15489-1 General** i l'informe tècnic **ISO/TR 15489-2 Guidelines**. Així mateix, l'any 2006, AENOR (Asociación Española de Normalización y Certificación) va publicar la traducció al castellà de les dues parts d'aquesta norma. La norma ISO 15489-1 estableix com ha de ser un sistema de gestió documental per a garantir la fiabilitat, l'autenticitat, la integritat i la disponibilitat dels documents. Així doncs, explica quines etapes, processos i controls s'han de seguir per al disseny i implementació d'un sistema d'aquestes característiques. Les pautes són aplicables a qualsevol organització pública o privada, i també serveixen per a qualsevol persona responsable de crear i mantenir documents, i afecta tota activitat, procés o tasca que es documenti i qualsevol document creat o rebut, sigui quin sigui el suport i format. D'altra banda, la norma s'adreça tant als professionals de la gestió de la informació i la documentació com als directius i responsables de les organitzacions, amb l'objectiu de conscienciar-los dels avantatges que ofereix la gestió de documents i mostrar-los les característiques que ha de tenir una política adequada d'aquest tipus. D'altra banda, l'informe tècnic ISO/TR 15489-2 conté una descripció més detallada de tot el procés d'implementació i de les funcionalitats d'un sistema de gestió documental, i s'adreça, sobretot, al personal tècnic encarregat d'endegar i dirigir aquests tipus de sistemes.

L'any 2001, la Comissió Europea va publicar una especificació complementària, el **MoReq**, que descriu els requisits funcionals que ha de tenir una aplicació informàtica destinada a la gestió de documents electrònics. En definitiva, són unes pautes per al disseny de l'aplicació i també una avaluació d'aplicacions que siguin útils per a la gestió d'aquests tipus de documents.

Lectura complementària

Si voleu conèixer més a fons la norma ISO 15489 i saber quines altres normes s'han publicat sobre la gestió documental al llarg d'aquests últims anys, llegiu l'article següent:

J. A. Alonso (2007, setembre-desembre). "La norma ISO 15489: un marco sistemático de buenas prácticas de gestión documental en las organizaciones". *Ítem* (núm. 47) (revisat el juliol de 2008).

http://eprints.rclis.org/12263/1/Alonso_Garcia_Lloveras_-_La_norma_ISO_15489.pdf

5. Exemples de sistemes de gestió de documents multimèdia

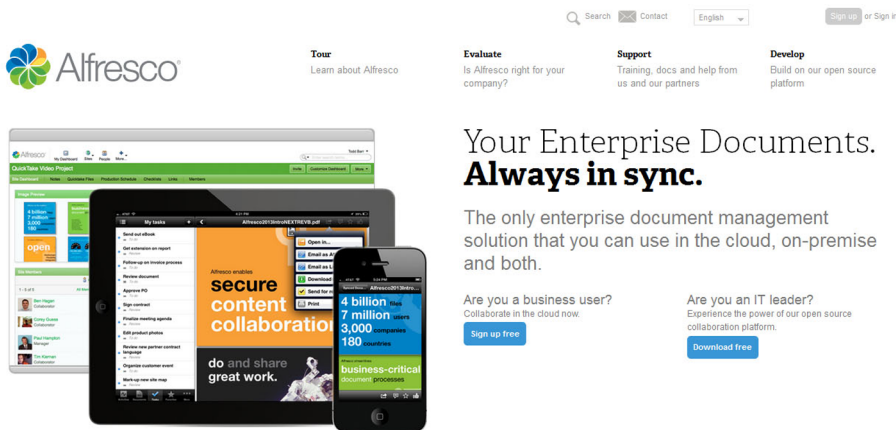
Els sistemes de gestió de documents multimèdia que es troben en el mercat solen oferir les prestacions clàssiques, com ara definició d'esquemes de metadades, eines de control terminològic, indexació i recuperació del text complet dels documents. L'alternativa més fàcil per a una organització, que no sempre resulta la més rendible, és l'adquisició d'una eina integrada, és a dir, un paquet informàtic complet que inclogui el programari i el maquinari. Aquests paquets simplifiquen *a priori* les decisions tècniques, però moltes vegades no s'adapten al funcionament concret de l'organització que els compra. És per aquest motiu que moltes vegades s'opta pel disseny d'una aplicació feta a mida que abracci totes les fases d'un sistema de gestió documental. Aquest disseny el pot dur a terme el personal tècnic de l'organització mateixa, una empresa externa, o bé fer-lo els professionals de tots dos llocs. No obstant això, els principals inconvenients d'aquesta opció són la rendibilitat a mitjà termini i l'elevat cost econòmic que representa, solament assequible per a les grans organitzacions. Vegem, a continuació, algunes de les aplicacions que podem trobar en el mercat.

5.1. Alfresco

Alfresco (<http://www.alfresco.com>) és una plataforma de gestió documental de codi obert que funciona amb Tomcat i utilitza MySQL. Es pot fer servir des de qualsevol ubicació, és a dir, tant en les instal·lacions mateixes de l'organització com en dispositius com tauletes o telèfons intel·ligents. Permet emmagatzemar i compartir documents que genera i custodia una empresa. A més, disposa d'eines socials que permeten als seus usuaris compartir i descobrir nous continguts. D'altra banda s'integra amb MS Office, Apple iWork, Adobe Creative Suite, Google Docs, etc. Pel que fa a l'esquema de preus, val a dir que ofereix diverses opcions:

- Versió gratuïta, que inclou 10 GB d'emmagatzematge.
- Versió Alfresco Community, descarregable lliurement i utilitzat pels usuaris que vulguin gestionar els seus propis documents.
- Versió Alfresco Standard Network, amb suport tècnic, funcions administratives i més emmagatzematge.
- Versió Alfresco Enterprise Network, quan es vol implementar el sistema a tota l'empresa. Els preus parteixen de 10 dòlars per usuari al mes.

Figura 1. Portada d'Alfresco



5.2. Nuxeo

Nuxeo (<http://www.nuxeo.com>) és una plataforma de gestió de continguts de codi obert. Ofereix gestió documental per a continguts multimèdia, col·laboració social (espais socials, perfils amb les activitats dels usuaris, misatges i microblogs, etc.) i gestió de casos específics (amb bústies de distribució, exemples de metadades, arbres de classificació, etc.). També ofereix una secció amb connectors (*plug-ins*) i una eina de configuració i personalització de l'aplicació.

5.3. KnowledgeTree

KnowledgeTree (<http://www.knowledgetree.com>) és un gestor documental de codi obert per a la petita i mitjana empresa. Està basat en una "pila" PHP, MySQL i Apache. Es presenta en dues versions:

- Versió gratuïta, funcional i molt completa.
- Versió comercial, que afegeix suport tècnic i un connector per a pujar documents treballant directament des del Microsoft Office.

A part de la configuració bàsica, permet fer personalitzacions, com ara afegir el logotip de l'empresa, canviar les opcions que es volen visualitzar a la pàgina d'inici, etc. Ofereix, a més, un repositori central de documents amb control de versions, la possibilitat de gestionar autories, metadades, fluxos de treball dels documents i eines d'auditoria.

5.4. OpenKM

OpenKM (<http://www.openkm.com>) també és un gestor documental de codi obert. Entre els serveis que ofereix destaquen el control de versions, metadades, tesaurus, escaneig, seguretat, comentaris, fòrums sobre els documents, flux de treball, motor de cerca, etc. Es tracta d'una aplicació Java J2EE que

s'executa sobre l'Apache Tomcat. Es pot instal·lar i executar en diferents plataformes i, a banda de la versió gratuïta, també ofereix una versió professional de pagament.

Figura 2. Portada d'OpenKM

OpenKM
Knowledge Management

Inicio Producto Modulos Capturas Videos Socios Recursos Información Contacto

DEMO ONLINE! DESCARGAR OPENKM

Comunidad OPENKM GRATIS Profesional VERSION

Cloud TODO INCLUIDO Universidad Y FORMACION

Videos destacados

OpenKM - Multiempresa
OpenKM - Knowledge Base
OpenKM - Móviles

OpenKM
OpenKM es una aplicación web de gestión documental que utiliza estándares y tecnologías Open Source.
OpenKM proporciona capacidades completas de gestión de documentos incluyendo el control de versiones, metadatos, escaneo, comentarios, foros sobre el documento, workflow, etc. Esto permite que las actividades sociales en torno al contenido se utilicen para conectar a las personas a otras personas, la información a la información, y las personas a la información, ayudando a gestionar, de forma más eficiente, la inteligencia colectiva que reside en los recursos humanos de la compañía.

Lectura complementària

Llegiu aquest article publicat a *La Vanguardia* el 19 de juliol de 2011 per Eva Domínguez, en el qual es descriuen quatre editors i gestors de continguts multimèdia especialitzats en diferents àmbits d'aplicació: Documenta (àmbit educatiu), Klynt i Zeega (per a documentals interactius) i 3WDOC (per a documentals web):

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

Lectura complementària

Si voleu saber més sobre la plataforma Documenta, llegiu l'entrevista feta el juny de 2011 a Cristina Casanova i Andrea Contino i publicada a la revista *Mosaic*, editada per la UOC:

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

6. El concepte de Web semàntic

6.1. Què és el Web semàntic?

De la mateixa manera que el Web 2.0, parlem de Web semàntic per a referir-nos a un ús concret d'un conjunt d'eines i tecnologies. L'any 2002, James Hendler, Tim Berners-Lee i Eric Miller van establir la definició següent:

“El Web semàntic és una extensió del Web actual en el qual a la informació disponible se li atorga un significat ben definit que permeti als ordinadors i les persones treballar en col·laboració. Es basa en la idea d'oferir en el Web dades definides i enllaçades, i permetre que aplicacions heterogènies localitzin, integrin, raonin i reutilitzin la informació present en el Web”.

El W3C (World Wide Web Consortium) el defineix de la manera següent:

“El Web semàntic és la representació de dades en el Web. És un esforç col·laboratiu liderat per W3C amb la participació d'un gran nombre d'investigadors i socis industrials. Es basa en l'ús d'RDF, que integra una gran varietat d'aplicacions mitjançant l'ús d'XML per a la sintaxi i l'ús d'URI per a la identificació”.

Partint d'aquesta última definició, l'RDF (*Resource Description Framework*) és un model de dades que ofereix una especificació per a la descripció de metadades en el Web. Organitza la informació en forma de subjecte-predicat-objecte i permet expressar-la sintàcticament mitjançant XML (*Extensible Markup Language*). A més, utilitza l'expressió URI (*Uniform Resource Identifier*) per a identificar de manera universal un espai de noms de recursos d'informació. En definitiva, el Web semàntic va molt més enllà que el Web convencional. Utilitza metadades per a descriure els recursos d'informació, parteix de vocabularis RDF per a representar aquestes metadades, desenvolupa esquemes RDF o ontologies per a descriure les relacions entre els recursos que s'han descrit i les propietats utilitzades per a caracteritzar-los, localitza, interconnecta i reutilitza fonts de dades RDF i infereix nova informació a partir de les relacions lògiques que s'estableixen entre les dades.

6.2. Arquitectura i estàndards del Web semàntic

Com sabem, les tecnologies bàsiques d'un web convencional són les següents:

- Localització d'objectes i recursos per mitjà d'URL.
- Utilització del protocol HTTP per a establir la comunicació entre clients i servidors.
- Marcatge dels documents amb el llenguatge HTML.

El Web semàntic, en canvi, localitza els objectes i recursos per mitjà d'URI i estén l'HTML amb tecnologies com RDF per a emmagatzemar i descriure les dades. El protocol HTTP continua essent, a baix nivell dins d'aquesta arquitectura, l'encarregat de la transmissió de dades entre clients i servidors.

A continuació, analitzem més a fons els estàndards del Web semàntic (Pastor Sánchez, 2011):

- **XML²**: metallenguatge per a la definició de vocabularis orientats a l'intercanvi de dades estructurades, independentment de la plataforma amb la qual s'executin les aplicacions que processen aquestes dades. Un document XML conté dades l'estructura de les quals es defineix a partir de dos mètodes: el DTD (*Document Type Definition*) o l'XML Schema, considerat el successor de l'anterior perquè ofereix definicions més expressives i exactes. Els diferents elements s'estructuren entorn d'un element arrel i els valors dels atributs es marquen entre cometes (dobles o simples). D'altra banda, aquest llenguatge també permet afegir dades literals i comentaris. En definitiva, amb XML s'aconsegueix un salt qualitatiu, ja que permet separar estructura, contingut i format visual.
- **RDF³**: model per a la descripció estructurada de recursos d'informació a la Xarxa. Com ja hem comentat anteriorment, les descripcions són del tipus subjecte-predicat-objecte i s'anomenen *sentències*. El subjecte és el recurs que s'ha de descriure, el predicat és una propietat o relació del recurs i l'objecte és el valor que s'assigna a aquesta propietat o el recurs a partir del qual s'estableix la relació. Aquest model tracta de donar resposta a certes limitacions plantejades per l'estructura jeràrquica del llenguatge XML.
- **RDFa⁴** (*Resource Description Framework -in- Attributes*): inclou informació semàntica en un lloc web també es pot fer de manera més senzilla que a partir del model RDF. Això és possible gràcies a RDFa, que permet afegir declaracions RDF en forma d'atributs en etiquetes amb codi (X)HTML. D'aquesta manera, coexisteixen dues versions en una mateixa URL: una adaptada per a ser consultada per persones, i l'altra adaptada per al processament per mitjà de màquines.

⁽²⁾Si voleu saber més sobre XML, consulteu l'enllaç <http://www.w3.org/XML>

⁽³⁾Si voleu saber més sobre RDF, consulteu l'enllaç <http://www.w3.org/RDF>

⁽⁴⁾Si voleu saber més sobre RDFa, consulteu els enllaços <http://www.w3.org/TR/xhtml-rdfa-primer> i <http://www.w3.org/TR/rdfa-syntax>

- **RDF Schema**⁵: el model RDF permet descriure i representar de manera explícita característiques de tot tipus de recursos. Això facilita la comprensió precisa de conjunts de dades de les màquines, que posteriorment poden ser processades per extreure informació o executar processos. El model RDFS (RDF Schema) és l'ampliació del model RDF, tot enriquint-lo amb un vocabulari que té certa semàntica. Amb RDFS es poden descriure esquemes senzills utilitzant classes i subclasses, i també definir propietats, el seu domini d'aplicació i el seu rang de valors possibles.
- **OWL**⁶ (*Web Ontology Language*): llenguatge que permet publicar i compartir en el Web ontologies que defineixen els termes i les relacions d'un vocabulari. Es tracta d'una extensió dels models RDF i RDFS, no és pas un llenguatge diferent. Les ontologies són molt importants en el Web semàntic. Són mecanismes molt potents, sense els quals els agents de programari no serien capaços de processar les dades per a l'obtenció d'informació. D'altra banda, defineixen les classes, les propietats i els individus (instàncies), juntament amb les subclasses, les subpropietats, el domini, el rang i altres aspectes com la relació inversa entre propietats i un conjunt d'axiomes addicionals per a definir restriccions. Incorporen, a més, enumeracions d'instàncies de classe i combinacions de classe.
- **SPARQL**⁷ (*SPARQL Protocol and RDF Query Language*): llenguatge d'interrogació que des del 2008 s'ha convertit en estàndard oficial del W3C per a la recuperació de sentències a partir de fonts de dades RDF. Permet expressar patrons del tipus subjecte-predicat-objecte que s'utilitzen per a trobar correspondències en el conjunt de sentències RDF.

⁽⁵⁾Si voleu saber més sobre RDF Schema, consulteu l'enllaç <http://www.w3.org/TR/rdf-schema>

⁽⁶⁾Si voleu saber més sobre OWL, consulteu l'enllaç <http://www.w3.org/2004/OWL>

⁽⁷⁾Si voleu saber més sobre SPARQL, consulteu l'enllaç <http://www.w3.org/TR/rdf-sparql-query>

6.3. El Web semàntic i els sistemes de gestió de continguts

Els sistemes de gestió de continguts són fonamentals en el desplegament del Web semàntic, ja que es basen en processos d'organització i estructuració de la informació (la majoria utilitzen bases de dades per a emmagatzemar els continguts, i separen estructura, informació i presentació). Aquest elevat nivell d'estructuració de la informació permet que aquesta es pugui reutilitzar per a afegir atributs RDF al codi (X)HTML de les pàgines. Alguns sistemes com Drupal⁸, per exemple, introdueixen de manera automàtica aquests atributs just durant el procés de renderització del codi (X)HTML. Aquest mateix i d'altres, com Liferay, permeten afegir mòduls que ofereixen continguts en format RDF.

⁽⁸⁾Si voleu saber com Drupal afegeix atributs RDF al codi (X)HTML, consulteu aquest enllaç: <http://drupal.org/project/rdfx>

Sigui com sigui, és necessari mapar els elements dels continguts que gestiona un CMS als seus elements corresponents de vocabularis RDF. Es tracta d'un procés similar a quan es defineixen els canals RSS des d'un lloc web, però molt més complicat, ja que s'ha de tenir molt clar quin vocabulari és el més adequat per a representar un determinat camp d'un tipus de contingut, a més de quin

nivell de detall i agregació. D'altra banda, val a dir que els CMS permeten incorporar dades externes RDF, i fins i tot la recuperació selectiva de dades per mitjà del llenguatge SPARQL (Pastor Sánchez, 2011).

6.4. Projectes i iniciatives internacionals entorn del Web semàntic

A escala internacional, s'han desenvolupat diverses aplicacions i iniciatives entorn del Web semàntic (Pastor Sánchez, 2011):

- **LOD⁹** (*Linked Open Data*): coneguda com a “dades obertes entrelaçades”, és una tècnica que permet publicar informació estructurada en el Web. Aquest procés es fa per mitjà de la tecnologia web, però amb un canvi: les dades estan pensades perquè els programes informàtics les puguin llegir, interpretar i processar automàticament. Cada vegada més, les empreses privades i les administracions públiques utilitzen aquesta tècnica per a gestionar les seves dades.
- **SKOS¹⁰** (*Simple Knowledge Organization System*): model per a representar diversos tipus de vocabularis controlats com ara tesaurus, classificacions, taxonomies o llistes d'encapçalaments de matèria. És una ontologia OWL Full simple, basada en RDF, que permet representar diferents tipus de llenguatges documentals per mitjà de classes i propietats.
- **Europeana¹¹**: des de l'any 2008, aquesta iniciativa pretén oferir una única porta d'entrada a una gran quantitat de recursos, resultat de la digitalització de diversos objectes culturals de més de 1.500 institucions de la Unió Europea (biblioteques, arxius, museus, organitzacions culturals...). Aquests objectes culturals s'emmagatzemen en un catàleg centralitzat, elaborat a partir de la recopilació de metadades de repositoris externs. Aquestes metadades són mapades a un únic esquema de descripció anomenat *ESE* (*Europeana Semantic Elements*). Ara bé, per no perdre part de les metadades originals, s'utilitza el model EDM (*Europeana Data Model*). D'altra banda, l'ús d'RDF en els processos de representació de la informació i en la definició de vocabularis ha fet que Europeana s'incorpori a iniciatives com el *Linked Open Data*, comentada abans, i per tant, al Web semàntic.
- **DBpedia¹²**: iniciativa engegada l'any 2007 que permet l'obtenció de dades estructurades a partir dels continguts de la Wikipedia. Per a referir-se a persones, llocs, treballs, espècies o organitzacions, DBpedia utilitza diferents subclasses de recursos. Cada classe té associada una sèrie de propie-

⁽⁹⁾Si voleu saber més sobre LOD, consulteu l'enllaç <http://www.w3.org/wiki/LinkedData>

Enllaços recomanats

Trobareu exemples de LOD en el registre de fonts de dades (*datasets*) de l'Open Knowledge Foundation <http://thedatahub.org> i en el catàleg de la Fundació CTIC Centro Tecnológico <http://datos.fundacionctic.org/sandbox/catalog/faceted>

⁽¹⁰⁾Si voleu saber més sobre SKOS, consulteu l'enllaç <http://www.w3.org/TR/skos-reference>

⁽¹¹⁾Si voleu saber més sobre Europeana, consulteu l'enllaç <http://www.europeana.eu>

Enllaç recomanat

La versió nord-americana d'Europeana és la Digital Public Library of America: <http://dp.la>.

⁽¹²⁾Si voleu saber més sobre DBpedia, consulteu l'enllaç <http://wiki.dbpedia.org>

tats que permeten la descripció d'objectes. Les dades de DBpedia es poden consultar en forma de pàgines, on es mostren de manera estructurada les diferents sentències sobre un concepte, fruit d'aplicar l'ontologia de mapatge. D'altra banda, val a dir que és la iniciativa amb més presència a *Linked Open Data*.

- **Calais**¹³: iniciativa de Reuters que converteix informació de blogs, CMS, webs o aplicacions en informació semàntica (format RDF). Es basa en ClearForest, una empresa de programari que desenvolupa anàlisis de textos i solucions de mineria de textos.

Enllaç recomanat

Si us voleu descarregar la font de dades de DBpedia, consulteu l'enllaç <http://wiki.dbpedia.org/Downloads37>.

⁽¹³⁾Per saber més de Calais vegeu: <http://www.opencalais.com>

6.5. El futur del Web semàntic

El Web semàntic aposta per la interoperabilitat, l'optimització de recursos i el desenvolupament de tecnologies obertes que beneficien tots els usuaris de la Xarxa. Les seves possibilitats són enormes i d'això s'han adonat tant els fabricants de programari com els dissenyadors de sistemes, els investigadors i fins i tot els polítics. La tecnologia *Linked Open Data*, per exemple, permet que els continguts estiguin interconnectats, i també que les dades s'adaptin a processos intel·ligents i dispositius de tot tipus. Actualment, els usuaris poden interconnectar els ordinadors, les tauletes i els telèfons intel·ligents per mitjà de serveis amb dades i programes ubicats en el "núvol" i que fan servir estàndards del Web semàntic per a emmagatzemar la informació. Els CMS, a més, també incorporen tecnologies del Web semàntic i cada dia es desenvolupen noves ontologies que formalitzen objectes d'informació per a ser publicats a la Xarxa.

Els objectius a curt termini possiblement giren entorn d'augmentar la quantitat de fonts de dades obertes i d'enllaços relacionats, i de millorar els resultats dels motors de cerca basats en el Web semàntic, i també d'integrar les xarxes socials en aquesta tecnologia. Fins i tot es parla d'una evolució a sistemes més complexos, pròxims a la intel·ligència artificial.

Bibliografia

Bibliografia bàsica

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.

Bibliografia complementària

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, setembre-desembre). "La norma ISO 15489: un marco sistemático de buenas prácticas de gestión documental en las organizaciones". *Ítem* (núm. 47; revisat el juliol de 2008). http://eprints.rclis.org/12263/1/Alonso_Garcia_Lloveras_-_La_norma_ISO_15489.pdf

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

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

Enllaços

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

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

DBpedia: <http://wiki.dbpedia.org> i <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> i <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> i <http://www.w3.org/TR/rdfa-syntax>

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

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