# RDF i Web Semàntica

**Jordi Pascual Aran**
ETIG / ETIS

**Oscar Celma Herrada**

09/01/2006

Agraeixo a Oscar Celma Herrada els consells facilitats
per a la redacció i estructuració d'aquest treball.

Dedico especialment el present document a la meva dona,
per el seu suport, la seva confiança i sobretot
per la seva paciència i els seus comentaris crítics,
sempre assenyats i carregats de raó.

# Resum

La Web actual està evolucionant cap a un nou concepte de xarxa per a compartir informació: la Web Semàntica. En aquesta Web, els documents en línia apareix estructurada, de manera que facilita l'intercanvi d'informació entre sistemes amb estructures de dades diferents i permet al mateix temps que aquesta informació sigui "comprensible" per a les màquines gràcies a l'ús de metadades. En aquest sentit, les bases de dades juguen un paper important, ja que constitueixen el suport necessari per a emmagatzemar i gestionar aquests nous tipus de documents.

En aquest treball s'explica el concepte de Web Semàntica, junt amb la seva estructura i els diferents termes relacionats amb aquesta idea. A més, es fa especial atenció al paper dels sistemes gestors de bases de dades en aquest camp, tenint en compte sobretot el nivell de compatibilitat que ofereixen aquests per a tractar dades en notació RDF, basada en el llenguatge XML.

# Índex de continguts

Índex de figures ..... 6
1. Introducció ..... 7
1.1 Justificació del TFC i context en el qual es desenvolupa. ..... 7
1.2 Aportació del TFC. ..... 7
1.3 Objectius del TFC. ..... 7
1.4 Enfocament i mètode seguit. ..... 7
1.5 Organització del projecte. ..... 7
2. Què és la Web Semàntica? ..... 8
2.1 Definició ..... 8
2.2 Per a què serveix? ..... 9
2.2.1 La Web avui dia ..... 9
2.2.2 La Web Semàntica ..... 10
2.3 Com funciona? ..... 11
2.3.1 Metadades ..... 11
2.3.2 Ontologies ..... 12
2.3.4 Agents ..... 13
2.4 La Web Semàntica actualment. ..... 14
3. Llenguatges entorn de la Web Semàntica ..... 17
3.1 XML ..... 19
3.2 RDF ..... 22
3.3 OWL ..... 24
3.4 Serveis Web ..... 26
3.5 Comunicació entre Agents ..... 27
3.6 Altres protocols ..... 27
3.6.1. SKOS-Core ..... 27
3.6.2 PICS ..... 27
3.6.2 Consultes ..... 28
4. Aplicacions de la Web Semàntica ..... 29
4.1 Sindicació i RSS ..... 29
4.2 FOAF ..... 30
4.3 Aplicacions d'ontologies ..... 31
4.3.1 Portals Web ..... 31

4.3.2 Col·leccions multimèdia ..... 31
4.3.3 Gestió de web's corporatives ..... 32
4.3.4 Agents i Serveis Web ..... 32
4.3.5 Informàtica omnipresent ..... 33
4.3.6 Buscadors semàntics ..... 33
4.3.7 Altres aplicacions ..... 34
5. Bases de Dades i RDF ..... 35
5.1 Definició ..... 35
5.2 Per a què serveixen? ..... 36
5.3 SGBD existents ..... 36
5.3.1 Lliures ..... 37
5.3.2 Comercials ..... 38
5.4 Un problema d'estructures ..... 41
6. Implementant Web Semàntica ..... 43
6.1 Generar dades RDF ..... 44
6.2 Diferents solucions de Web Semàntica ..... 46
6.2.1 RDF Interface ..... 46
6.2.2 Sesame ..... 48
6.2.3 JENA ..... 51
6.2.4 RDF Suite ..... 52
6.2.5 rdfDB ..... 52
6.2.6 Oracle i RDF ..... 53
6.3 Publicació en la Web ..... 55
7. Valoració econòmica ..... 56
8. Conclusió ..... 57
Glossari ..... 59
Bibliografia ..... 60
Llocs Web ..... 61
ANNEX 1: El problema de la Web actualment ..... 62
ANNEX 2: Projectes i eines relacionats amb la SW ..... 63

# Índex de figures

Figura 2.1 La mida d'Internet en Terabytes ..... 9
Figura 2.2 Representació UML d'una ontologia ..... 13
Figura 3.1 Estructura de la Web Semàntica ..... 17
Figura 3.2 Exemple d'informació a la Web ..... 19
Figura 3.3 Exemple de "tripleta" RDF ..... 23
Figura 3.4 Relacions de classes amb RDF-Schema ..... 24
Figura 4.1 El "botó taronja" de la sindicació ..... 30
Figura 5.1 Bases de dades suportant XML ..... 40
Figura 5.2 Problemàtica XML - SGBDR ..... 41
Figura 6.1 Exemple d'ontologia OWL ..... 45
Figura 6.2 Comprovar la sintaxis d'una ontologia ..... 45
Figura 6.3 Exemple de fitxer RDF ..... 45
Figura 6.4 Diagrama de classes dins l'ontologia ..... 46
Figura 6.5 Administració de RDF Interface ..... 47
Figura 6.6 Bases de dades amb RDF Interface ..... 47
Figura 6.7 Fitxers RDF amb InferEd ..... 47
Figura 6.8 Consultes amb RDF Query Analyzer ..... 48
Figura 6.9 Instal·lació de Java 2 ..... 49
Figura 6.10 Instal·lació d'un servidor Apache ..... 49
Figura 6.11 Instal·lació de MySQL ..... 50
Figura 6.12 Configuració de Sesame ..... 50
Figura 6.13 Utilització de Sesame ..... 50
Figura 6.14 Estructura de Sesame ..... 51
Figura 6.15 Administració de PostgreSQL ..... 52
Figura 6.16 Instal·lació d'Oracle 9i ..... 53
Figura 6.17 Administració de Oracle 10g ..... 53
Figura 6.18 Exemple d'inserció de dades RDF per Oracle 10g ..... 54

# 1. Introducció

## 1.1 Justificació del TFC i context en el qual es desenvolupa.

Un nou terme està prenen forma al voltant d'Internet: la Web Semàntica. Tot i que aquesta expressió aparegué per primer cop l'any 2001, era necessari establir abans els fonaments d'aquest concepte, i és des de fa poc que s'estan obtenint resultats tangibles, a una velocitat exponencial – com correspon a qualsevol àmbit de les Noves Tecnologies. La Web Semàntica permet donar un significat al contingut de la web actual, de manera que sigui "comprensible" per a una màquina. En l'àmbit de la Web Semàntica, la informació es presenta estructurada. El llenguatge RDF (*Resource Description Framework*) és el format establert per a estructurar les dades de la Web i fer-les entenedores per als ordinadors. A més, RDF proveeix interoperabilitat entre les aplicacions que intercanvien informació a través de la Web. Aquest projecte explica el concepte de Web Semàntica i presenta el funcionament dels sistemes gestors de bases de dades per desar-hi informació en notació RDF.

## 1.2 Aportació del TFC.

Les principals aportacions d'aquest treball consisteixen en donar una explicació clara, però per això menys tècnica, del que representa la Web Semàntica i presentar l'important paper que jugaran en aquest sentit els sistemes gestors de bases de dades, fent especial atenció al tractament de les dades RDF.

## 1.3 Objectius del TFC.

Aquest treball pretén comprendre i donar a conèixer el concepte de "Web Semàntica", presentar els diferents llenguatges, protocols i notacions que hi intervenen, demostrar la utilitat de la Web Semàntica mitjançant exemples i aplicacions al món de la Web actual, estudiar el paper que juguen els sistemes gestors de bases de dades en aquest camp.

## 1.4 Enfocament i mètode seguit.

Per a aconseguir aquests objectius, s'introdueix el concepte de Web Semàntica, se'n presenta l'estructura i els conceptes relacionats, es mostren algunes aplicacions de la Web Semàntica en la Web actual, es recorden els sistemes gestors de bases de dades més importants i es relacionen amb la Web Semàntica i sobretot amb el concepte de RDF, per acabar presentant diferents maneres d'implementar una aplicació web amb semàntica.

## 1.5 Organització del projecte.

Aquest Treball de Final de Carrera s'ha organitzat seguint els objectius. Primerament es presenta el concepte de Web Semàntica, en contraposició a la Web tal i com la coneixem actualment. En el següent capítol, més tècnic, es presenta l'estructura de la Web Semàntica i els diferents conceptes i llenguatges que hi intervenen. El tema 4 introdueix algunes de les aplicacions de la Web Semàntica ja existents. A continuació es presenten els sistemes gestors de bases de dades (SGBD) i es relacionen amb el sistema de notació RDF. Finalment, en el capítol 6, es presenten diferents alternatives per a implementar la Web Semàntica mitjançant els SGBD. Els capítols 7 i 8 corresponen a la valoració econòmica i la conclusió final.

# 2. Què és la Web Semàntica?

En aquest primer capítol es dóna una definició de Web Semàntica, es presenta l'entorn de treball, s'identifiquen els propòsits i objectius d'aquest concepte i s'expliquen els components que faran possible assolir aquests objectius.

## *2.1 Definició*

*"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."*[1]

La noció de Web Semàntica apareix a finals dels anys 90 [Berners-Lee 2001]. Aquesta definició fou establerta pel propi inventor de la Web[2] i president del consorci W3C[3], amb la finalitat d'aconseguir que les màquines puguin entendre i, per tant utilitzar, els continguts de la Web. Aquesta nova web es serviria d'agents capaços de navegar i realitzar operacions, estalviant feina a les persones i optimitzant els resultats de cerca.

Per aconseguir aquest objectiu la Web Semàntica (*Semantic Web*, d'ara endavant SW) proposa descriure els recursos de la web mitjançant representacions comprensibles, no només per a les persones, sinó també per a programes que puguin ajudar, o fins i tot substituir, a les persones en aquelles tasques que siguin excessivament rutinàries o inabordables per a l'ésser humà.

Donant més significat a la Web, és a dir, més semàntica, es poden obtenir solucions a problemes habituals en la recerca d'informació gràcies a l'ús d'una infrastructura comuna, mitjançant la qual és possible compartir, processar i transmetre informació de forma senzilla.

Aquesta web estesa i basada en el significat, es serveix de llenguatges universals que resolen els problemes ocasionats per una web sense semàntica en la que de vegades l'accés a la informació es converteix en una tasca difícil i frustrant.

---

[1] Tim Berners-Lee, James Hendler, Ora Lassila, "The Semantic Web", *Scientific American*, Maig 2001.

"La Web Semàntica és una extensió de la Web, dotada d'un significat més gran, en la que qualsevol usuari d'Internet podrà trobar respostes a les seves preguntes de forma més ràpida i senzilla gràcies a una informació millor definida."

[2] L'aparició de la WWW o Web se situa a l'any 1989 [Abrams 1998, Connolly 2000], quan Tim Berners-Lee presentà el seu projecte de "World Wide Web" [Berners-Lee 1989] al CERN (a Suïssa), amb les característiques essencials que perduren actualment. El propi Berners-Lee completà l'any 1990 el primer servidor web i el primer client, i un any més tard publicà el primer esborrany de les especificacions de HTML i HTTP.

[3] El Consorci World Wide Web (W3C) és una associació internacional formada per organitzacions membres del consorci, personal i el públic en general, que treballen conjuntament per a desenvolupar estàndards Web. La missió del W3C és: "Guiar la Web cap al seu màxim potencial a través del desenvolupament de protocols i pautes que asseguren el creixement futur de la Web".

## *2.2 Per a què serveix?*

### 2.2.1 La Web avui dia

Resulta prou complicat d'avaluar la mida de la Web, però s'estima que hi ha uns 600 milions d'usuaris, i que s'hi han publicat de l'ordre de 167 Terabytes (TB), un volum d'informació que equival a entre 14 i 28 milions de llibres [Lyman 2003] [4]. Aquestes xifres només es refereixen al que s'anomena la "web superficial", formada pels documents estàtics accessibles en la web (és a dir, la informació que podem trobar amb l'ajuda d'un motor de cerca). Però si considerem també la "web profunda"[5], estem parlant d'una quantitat d'informació 500 vegades més gran que la web superficial, i que creix a un ritme encara més gran [O'Neill 2003]. Es creu que la mida de la web profunda ha superat ja el volum total d'informació impresa existent en tot el planeta.

| Mitjà | Mida en TB |
|---|---|
| Web Superficial | 167 |
| Web Profunda | 91.850 |
| Correu electrònic (originals) | 440.606 |
| missatgeria Instantània | 274 |
| TOTAL | 532.897 |

Font: "How much information? 2003", [Lyman 2003].

**Figura 2.1 La mida d'Internet en Terabytes**

Avui en dia pràcticament tot està representat en la web i, amb l'ajuda d'un bon cercador, podem trobar informació sobre qualsevol cosa que necessitem. La web està en camí de convertir-se en una mena d'enciclopèdia universal del coneixement humà. D'altra banda, la web ens permet realitzar diferents activitats del dia a dia amb una comoditat i economia inèdits; sense moure'ns de casa podem adquirir tot tipus de productes i serveis: fer les compres de la setmana al supermercat, adquirir llibres, consultar l'ultima novetat al cinema, comprar entrades pel teatre, gestionar el compte bancari, matricular-se a la universitat, etc.

Però els buscadors actuals tenen els seus límits. Si per exemple volem saber més coses sobre el grup musical "Ketchup", introduirem aquest mot en un buscador (per exemple Google) i trobarem efectivament informació relacionada amb el grup de música, però també amb d'altres enllaços a receptes de cuina, al club de fans de la salsa de tomàquet, restaurants (veure Annex 1).

---

[4] Varian & Lyman calculen que un llibre d'unes 300 pàgines equival a 1 MB d'espai disc. Per al càlcul de la mida d'Internet, assumeixen que la mida mitja d'una pàgina web és entre 10 i 20 KB [Lyman 2003].

[5] La 'Web Profunda' està formada per tots els documents als que no es pot accedir a través dels directoris i enllaços que ofereixen els resultats dels buscadors convencionals.

Aquest exemple ens serveix per certificar que la mida descomunal de la web fa que algunes tasques ens demanin un temps excessiu, un temps del qual moltes vegades no disposem. Ens caldria doncs utilitzar algun tipus d'eina que fos capaç de realitzar una cerca i, a més, amb capacitat de comprendre i decidir si el resultat obtingut s'ajusta a les nostres necessitats. Però desenvolupar aquest tipus de programes és molt difícil i comporta un cost molt elevat degut a la codificació utilitzada actualment pels continguts de la web. Un programa (un motor de cerca) pot trobar la informació i portar-la a l'abast de les persones, però no pot saber-ne el significat, la qual cosa limita enormement la seva autonomia. El problema doncs, és la falta de capacitat per expressar significats (semàntica) de les representacions en les quals es basa la web; els continguts i serveis de la web es presenten en formats (com HTML) i interfícies (com formularis) comprensibles per a les persones, però no per a les màquines. L'origen del problema el trobem en la limitació dels buscadors a l'hora d'utilitzar expressions complexes a nivell semàntic: només comprenen paraules clau, però inconnexes, que no permet reconèixer significats més elaborats.

### 2.2.2 La Web Semàntica

Aquí és on entra en joc el concepte de SW. Si recuperem la definició de SW veurem que això significa, ni més ni menys, afegir a la web actual la capacitat de comprendre i expressar significats, obtenint així una web més eficaç i fàcil a utilitzar. La SW recupera la noció d'"ontologia"[6] del camp de la Intel·ligència Artificial (AI) com a vehicle per aconseguir aquest objectiu, així com la noció de "metadades"[7]. Això implica que, a més de comprendre el significat d'una expressió de recerca, un buscador semàntic tindrà en compte el perfil de l'usuari i establirà patrons de comportament en funció d'un històric de cerques realitzades pel mateix usuari. En aquest sentit, podríem introduir una cerca del tipus "anar amb avió a Roma la setmana que ve": el buscador seria capaç de comprendre quina és la ciutat de sortida (p.e. Barcelona) i a quines dates concretes ens estem referint (en funció del dia en que s'executa la petició i la data del sistema); el resultat seria una llista d'horaris i preus que s'ajusten a les nostres necessitats.

Però aquest concepte és més potent encara, ja que la SW també comprèn els Serveis Web. Podem definir els Serveis Web[8] (*Web Services*, a partir d'ara WS) com un conjunt d'aplicacions o de tecnologies amb capacitat per a interactuar en la Web; aquestes aplicacions o tecnologies intercanvien dades entre elles amb l'objectiu d'oferir un servei. Els WS no són un invent de la Web Semàntica, però prenen aquí un nou significat. Per a entendre aquest punt, recuperem l'exemple dels vols a Roma: una vegada feta la cerca, el buscador ens ofereix una llista de possibles vols, aleshores en triarem un. Però el procés real no s'acaba aquí ja que si anem a Roma, a més del bitllet d'avió, ens caldrà un hotel, potser un cotxe de lloguer i els desplaçaments de l'aeroport al centre ciutat. I a més ens cal pagar tot això. Gràcies a la SW, un agent[9] serà capaç de realitzar totes aquestes operacions requerint una interacció mínima amb l'usuari.

---

[6] Veure capítol 2.3.3.

[7] Veure capítol 2.3.1.

[8] Veure capítol 3.4.

[9] Veure capítol 2.3.4.

## *2.3 Com funciona?*

La SW no pretén que els ordinadors puguin comprendre el llenguatge humà de forma natural; no es tracta d'implementar una IA que permeti a la Web de pensar. Es tracta simplement de reagrupar la informació de manera útil, a la manera d'una enorme base de dades, on tot està escrit en un llenguatge estructurat.

Així, de la mateixa manera que la web actual s'ha construït principalment al voltant d'identificadors URI[10], de protocol HTTP i de llenguatge HTML, la Web Semàntica es basa en URI, HTTP i el llenguatge RDF (*Resource Description Framework*). El format RDF és la gran novetat dels treballs d'investigació al voltant de la Web Semàntica. Gràcies a RDF s'utilitza una infrastructura basada en metadades.

El llenguatge RDF s'explica més endavant. Abans veurem què son les metadades, introduirem el concepte d'ontologia i veurem quin paper juguen els agents en la Web Semàntica.

### 2.3.1 Metadades

Etimològicament, metadada (meta+dada)[11] significa "dades a propòsit de dades".

Les metadades son doncs dades estructurades que descriuen informació; ens diuen el qui, què, quan, com, on i perquè, sobre les dades que es documenten.

Degut a la gran quantitat de dades que conté la web, cal, necessàriament, posar un ordre a tota aquesta informació, i la manera de fer-ho és generant metadades que expliquin el contingut d'una pàgina. Aquestes metadades no són directament visibles per l'usuari, però sí comprensibles per als agents, de manera que permeten identificar correctament el tipus de dada en funció de la cerca a realitzar.

Existeixen diferents estàndards de metadades. Els més rellevants actualment són:

- DC (*Dublin Core*): defineix un conjunt de propietats recomanades per a descripcions bibliogràfiques electròniques.
- LOM (*Learning Object Metadata*): es centra en el conjunt mínim de propietats que permeten que els objectes educacionals[12] siguin gestionats, ubicats i avaluats.
- EML (*Educational Modelling Language*): EML és un llenguatge per a modelar objectes educatius. EML utilitza metadades LOM. [Cámara 2002].

---

[10] En el Glossari es trova el significat de totes aquestes sigles.

[11] Meta-: del grec *meta*, prefix que expressa "que encercla, que engloba" ; p.e.: "metallenguatje". I, en conseqüència : " Punt a què cal arribar ; objectiu a abastar." *Diccionari de la Llengua Catalana*.

[12] Un "objecte educacional" es defineix com una entitat reutilitzada o referenciada durant qualsevol activitat d'aprenentatge basada en la tecnologia [Cámara 2002].

Així mateix, hi ha diferents organismes que proposen i gestionen cadascun d'aquests estàndards:

- DCMI (*Dublin Core Metadata Initiative*) per a Dublin Core.
- IEEE (*Institute of Electrical and Electronics Engineers*) i LTSC (*Learning Technology Standards Committee*) per a LOM.
- CEN/ISSS LTWS (*Centre Européen de Normalisation* / Information Society *Standardization System - Learning Technologies WorkShop*) per a EML.

### 2.3.2 Ontologies

La definició més estesa d'ontologia és la donada per Gruber [Gruber 1993]:

*"An ontology is a formal, explicit specification of a shared conceptualization."*

En aquesta definició, "conceptualització" es refereix a un model abstracte d'algun fenomen del món del que s'identifiquen els conceptes importants; "explícit" fa referència a la necessitat d'especificar els diferents conceptes que formen una ontologia; "formal" indica que l'especificació ha de representar-se per via d'un llenguatge degudament definit; "compartit" comporta que una ontologia ha de ser acceptada, com a mínim, pel grup de persones que han d'utilitzar-la.

Una definició potser més aclaridora és la donada per Weigand [Weigand 1997]:

*"An ontology is a database describing the concepts in the world or some domain, some of their properties and how the concepts relate to each other."*

Una ontologia és doncs una base de dades que descriu el conjunt de conceptes que formen part d'un domini determinat, les seves característiques i les relacions entre aquests conceptes.

El desenvolupament d'una ontologia inclou les etapes següents:

- Definir classes en la ontologia.
- Col·locar las classes en un jerarquia de taxonomies (subclasse-superclasse).
- Definir els atributs (*slots*) de les classes i descriure els valors permesos per a aquests atributs.
- Establir els valors dels atributs amb exemples.

I quina relació hi ha entre les ontologies i les metadades? Doncs bé, una ontologia estableix les relacions entre els diferents termes que apareixen en la descripció dels documents. Els ordinadors "comprendran", podran interpretar, el significat de les dades semàntiques d'una pàgina web seguint els vincles amb ontologies prèviament especificades. Una ontologia ha de ser prou acurada per a diferenciar entre un banc per seure i un lloc on es guarden els estalvis.

Estudiem un exemple d'ontologia per entendre millor aquest concepte. Primerament hem de definir un domini: UOC. Dins aquest domini trobem diferents conceptes: Consultor, Estudiant, TFC, Assignatura. Aquests conceptes els definim com a classes dotades de certes propietats: Consultor(Nom), TFC(AnyPublicacio), Assignatura(NomAssignatura). A continuació cal definir les relacions entre classes, d'herència o altres: Consultor i Estudiant "són subclasses" de Persona, Consultor "és responsable" d'una Assignatura, Estudiant "realitza un" TFC. La representació UML (*Unified Modelling Language*) seria[13]:

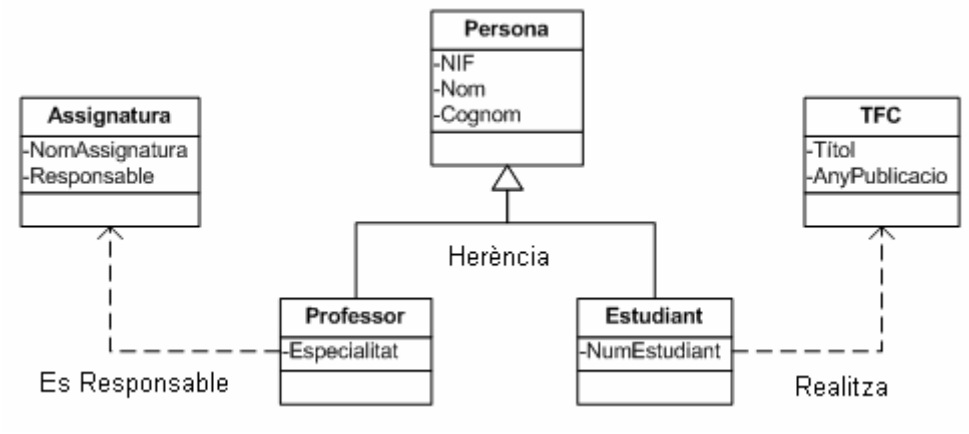



**Figura 2.2 Representació UML d'una ontologia**

Ens quedaria un últim pas per a definir completament la ontologia: establir els valors dels atributs amb exemples. La manera de completar aquest punt variarà segons el llenguatge utilitzat[14]. Existeixen diferents estàndards per a definir ontologies, però el més estès és OWL, definit per W3C i basat en el llenguatge RDF. Tots els estàndards i llenguatges s'estudien en el capítol 3 d'aquest document.

### 2.3.4 Agents

Idealment, els usuaris de la Web Semàntica no seran les persones, sinó els agents. Un agent és una entitat de software autònoma (sense necessitat d'intervenció humana) que realitza accions complexes per aconseguir un objectiu concret. En un mateix medi, sovint hi ha diferents agent que interactuen: tenen un caràcter social. Els agents poden modificar el seu comportament en funció dels seus objectius, són reactius a l'entorn que els envolta, i es comuniquen (i col·laboren) amb altres agents per a realitzar una tasca determinada.

Per a que un agent pugui realitzar una tasca, aquest ha de tenir la capacitat de processar el llenguatge natural. A més necessita planificar les tasques a realitzar; això implica que l'agent coneix l'entorn de treball. En la web actual això és impossible, però en la web semàntica l'agent podrà llegir els continguts semàntics d'una pàgina en un llenguatge formal (OWL), utilitzant coneixements compartits (ontologies).

[13] Un model de classe UML pot transformar-se en una ontologia de Web Semàntica, [Provost 2002].

[14] Es pot consultar un exemple d'ontologia OWL al punt 3.3.

Els sistemes multi-agents s'estableixen com una eina per a l'aprenentatge automàtic en Internet. Els agents poden prendre decisions en funció del perfil d'usuari. Mitjançant aquesta eina s'aconsegueixen recomanacions del tipus "accedeix a aquesta pàgina perquè em sembla que t'interessa", "aquesta persona té interessos semblants als teus, podries posar-t'hi en contacte". Així doncs, podrem delegar certes tasques repetitives o complicades a un agent, que serà capaç de prendre la decisió correcte tenint en compte les meves preferències: en l'exemple dels vols a Roma, nosaltres ens limitem a expressar una necessitat: "anar amb avió a Roma la setmana que ve", i és l'agent qui farà la resta (cerca de vols, reserves, hotels, etc).

Dit d'una altra manera, l'agent és un assistent personal que està dins la computadora i que juga diferents papers en la representació d'una funció específica o d'un usuari. En la Web Semàntica són els encarregats de les cerques de serveis. En aquest sentit la semàntica facultarà als agents per a descriure's els uns als altres la funció exacta que pertoca a cadascun i quines dades han de rebre per treballar correctament.

L'organisme FIPA (*Foundation for Intelligent Physical Agents*)[15] s'encarrega d'estandarditzar aspectes relacionats amb la tecnologia d'agents i sistemes multi-agents. Un dels estàndards desenvolupats és el *FIPA Agent Management Specification*[16], en el que es descriu el model de referència de la plataforma d'agents i la funcionalitat de cadascun dels seus components: sistemes d'administració d'agents, directoris d'agents i canals de comunicació d'agents.

## *2.4 La Web Semàntica actualment*

Com potser ja s'ha intuït, i com veurem en el capítol 3, RDF i OWL formen el nucli de la Web Semàntica. En aquest sentit podem dir que la Web Semàntica ha deixat de ser una idea i que avui en dia ja s'està aplicant en diferents camps. La confirmació l'obtenim directament de W3C, quan publica en el seu *W3C Semantic Web News and Events Archive* la noticia següent:

> *RDF and OWL Working Groups Complete Deliverables, Close 2004-07-01, W3C is pleased to announce that the RDF Core and Web Ontology Working Groups have successfully completed all deliverables. Together these W3C Working Groups developed twelve W3C Recommendations specifying the Resource Description Framework (RDF) and the OWL Web Ontology Language.*

Aleshores, un cop fixats i ben definits els principis i estàndards de la SW, ja estem preparats per a la seva implantació i posterior utilització. En aquest sentit, tot i que evidentment no s'han assolit encara els objectius definits per Tim Berners-Lee, trobem actualment que els fonaments de la Web Semàntica s'utilitzen en diferents aspectes del món Web:

[15] http://www.fipa.org

[16] http://www.fipa.org/specs/fipa00023/

- Descripció de recursos: s'estan definint esquemes de metadades basats en RDF, aplicant estàndards com el de DCMI[17].
- Definició d'ontologies i de tesaures[18] especialitzats per a la organització i recuperació temàtica de la informació basats en RDF o OWL. Per a la definició de tesaures s'ha creat l'estàndard SKOS (*Simple Knowledge Organization System*)[19].
- Descripció d'imatges: hi ha diferents serveis que utilitzen les imatges com a mitjà per expressar informació, com museus, informació geogràfica (com Google Earth), sistemes de fotografia digital de premsa, etc. El W3C ha desenvolupat un sistema per a inserir descripcions d'imatges en format RDF en la imatge mateixa, el RDFPic[20].
- Agregació o sindicació de continguts: L'àmbit de treball de la sindicació és l'aplicació de RDF a portals web i bitàcores. Curiosament l'estàndard que més s'utilitza (RSS)[21] no fou creat per el W3C.

Però el procés d'implantació de la Web Semàntica no és ni fàcil ni evident. Constantment sorgeixen nous esquemes que tracten de concretar algun dels aspectes de la SW mitjançant vocabularis específics RDF[22]. Aquest creixement incontrolat de pseudo-estàndards dificulta la tasca de W3C per a aconseguir d'unificar idees i establir protocols específics comuns.

Un altre problema que es planteja és el predomini de "*middlewares*"[23]; són molt poques les solucions existents per a "crear" o "consumir" WS directament[24].

Una dificultat important per a portar a la pràctica la Web Semàntica es la de acordar una única ontologia per a una mateixa comunitat. Convergir en una representació comuna és una tasca complexa ja que típicament cada domini comporta unes peculiaritats i un punt de vista propi, que caldrà expressar en la propia ontologia. Recordem, per exemple, que no obtenim la mateixa definició del mot "banc"[25] segons si ens trobem en el camp financer o en l'àmbit immobiliari.

---

[17] Vist en el punt 2.3.1.

[18] El mot "tesaure" prové del llatí *thesaurus* (tresor). En el camp de les Ciències de la Informació, un tesaure es defineix com un sistema de documentació que consta d'una llista de termes (que poden estar definits per més d'una paraula) relatius a un domini determinat i d'un conjunt de relacions semàntiques entre els termes que el composen. Les relacions poden ser de equivalència, associació i jerarquia. Cada terme porta associada la seva definició.

[19] http://www.w3.org/2001/sw/Europe/events/200406-esp/trabajo-final-extratesauros/node6.html

[20] http://www.w3.org/TR/photo-rdf/

[21] Veure capítol 4.1.

[22] Trobem alguns d'aquests esquemes al lloc: http://www.schemaweb.info/schema/BrowseSchema.aspx

[23] *Middleware:* Literalment, "elements del mig". Programa que s'utilitza per a connectar dues o més aplicacions.

[24] Algunes d'aquestes eines estan recollides a: http://www.w3.org/RDF/#developers.

[25] Banc: "Seient estret i llarg, on caben algunes persones"; "establiment públic que fa comerç amb el diner". *Diccionari de la Llengua Catalana*.

La SW està en marxa i són molts els projectes i sistemes d'informació que usen RDF: bloggers, xarxes socials, agregadors i sindicadors de continguts, buscadors semàntics. Parlarem de totes aquestes aplicacions de la SW al capítol 4.

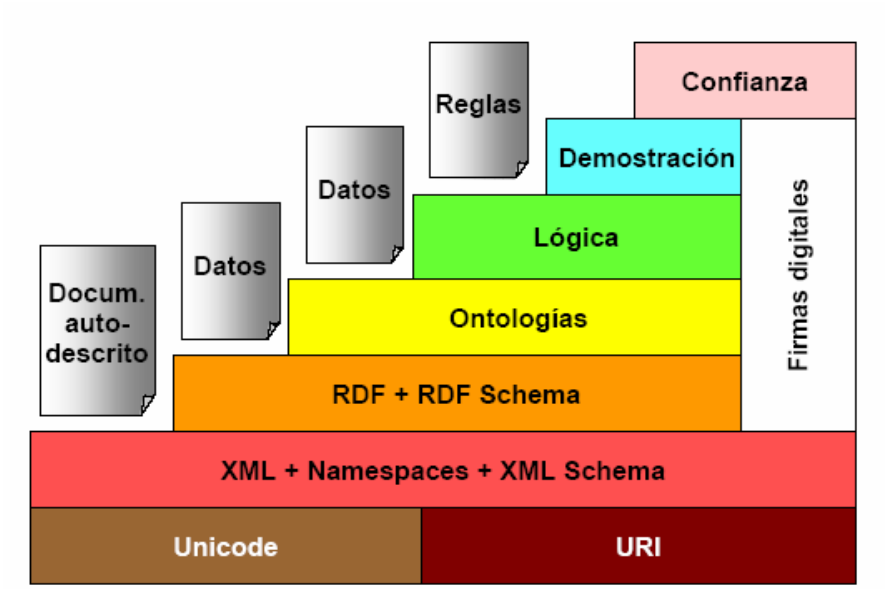
Com es tracta d'un tema en constant evolució, cal estar en contacte amb els canals oficials i consultar activament les pàgines d'organismes com W3C. Concretament, es pot accedir a *Semantic Web Activity Statement*[26] on es publiquen les especificacions de les tecnologies que sorgeixen entorn de la Web Semàntica. També es poden trobar les novetats publicades per W3C en la versió castellana: "Archivo de noticias 2005"[27].

[26] http://www.w3.org/2001/sw/Activity

[27] http://www.w3c.es/Noticias/2005/

# 3. Llenguatges entorn de la Web Semàntica

El primer llenguatge per a la construcció de la Web Semàntica fou SHOE[28], creat per Jim Hendler en 1997. D'aleshores ençà, s'han definit altres llenguatges i estàndards amb finalitats similars, com XML, RDF, DAML+OIL[29], i més tard OWL. La tecnologia que s'ha generat per a fer possible la Web Semàntica inclou diferents nivells de llenguatges: per a la representació de metadades, llenguatges de consulta, entorns de desenvolupament, mòduls de gestió d'ontologies.... Els estàndards oficials[30] de la Web Semàntica són els publicats pel W3C. L'estructura de SW és la següent:




Font: "Tecnologías XML y Web Semántica", [Labra 2005].

**Figura 3.1 Estructura de la Web Semàntica**

Com queda reflectit en la Figura 3.1, un aspecte central de la infrastructura SW és la capacitat d'identificació i de localització dels recursos. Aquesta infrastructura es basa en la noció d'URI, que permet d'assignar un identificador únic a un conjunt de recursos. Com ja hem vist en l'apartat 2.3, aquesta noció és la base dels llenguatges del W3C.

Una altra característica de tots aquests llenguatges és que es poden expressar en una mateixa notació, XML (*eXtensive Markup Language*)[31], aprovat el 1998. XML és un llenguatge que permet estructurar documents segons un vocabulari definit per l'usuari i permet l'intercanvi d'informació estructurada entre diferents plataformes.

---

[28] http://www.cs.umd.edu/projects/plus/SHOE/

[29] http://www.w3.org/TR/daml+oil-reference/

[30] De fet el que publica el W3C són recomanacions; però aquestes recomanacions són directament enteses per la indústria i per la comunitat Web en general com estàndards.

[31] Veure apartat 3.1.

Abans de l'aparició del concepte de Web Semàntica, tota la informació publicada en l'àmbit d'Internet s'expressava en HTML. Gràcies a XML, es comencen a estructurar els documents, que és un dels primers requeriments per a l'adopció de la SW. El problema és que amb XML cada usuari pot definir les marques de text que consideri necessàries, donant-li el nom que li convingui. Evidentment això fa que, encara que la informació estigui estructurada, existeix un "llenguatge" o metodologia diferent i pràcticament pròpia per a cada document. Aleshores la compartició o comprensió d'aquesta informació per part de diferents sistemes o agents es fa bastant difícil. Es fa necessari doncs proporcionar un model únic, un conjunt de regles per a proporcionar informació descriptiva senzilla. Amb RDF s'aconsegueix aquest propòsit. Per entendre el vincle XML-RDF podríem dir que: XML és a la sintaxis, com RDF és a la semàntica.

En un nivell superior, es necessita un sistema per a definir vocabularis que corresponguin a un tema (o un domini) específic. És el paper de les ontologies. Una ontologia defineix els termes utilitzats per a descriure i representar una àrea de coneixement. Les ontologies son utilitzades per persones, bases de dades i aplicacions que comparteixen informació específica. Així, s'obtenen ontologies sobre medicina, sobre immobiliària o sobre reparació de vehicles. Les ontologies inclouen definicions sobre conceptes bàsics i les relacions existents entre aquests. D'aquesta manera generen coneixement codificat, estructurat i reutilitzable. El primer llenguatge formal que aparegué per a definir ontologies fou SHOE. Amb l'aparició de XML vingué XOL (*XML Ontology exchange Language*)[32]. Seguidament aparegueren dos llenguatges molt similars: DAML (*DARPA Agent Markup Language*)[33], creat per l'Agencia de Defensa dels Estats Units (DARPA)[34]; i OIL (*Ontology Inference Language*)[35], desenvolupat per el programa IST (*Information Society Technologies*)[36] de la Comissió Europea. Més endavant aquests dos llenguatges es varen fusionar per a constituir una nova especificació; DAML+OIL. OWL és el llenguatge d'ontologies desenvolupat per W3C, i és una extensió de RDF, afegint la possibilitat d'utilitzar expressions lògiques (del tipus AND i OR). Tot i que RDF i OWL són avui dia els més consolidats, existeixen altres llenguatges interessants, menys divulgats, com *TopicMaps*[37], *OCML*[38] o *WebODE*[39].

Més enllà de les ontologies, existeixen d'altres nivells encara en procés d'investigació. El nivell Lògic: permetrà descriure coneixement declaratiu a nivell d'aplicació. El nivell de Demostració: s'encarregarà dels processos d'inferència, intentarà demostrar com s'ha obtingut una informació (o resposta) determinada. El nivell de Confiança: es basa en la utilització de firmes digitals i altres recursos per a garantir la credibilitat de les operacions realitzades (com per exemple en l'àmbit del comerç electrònic).

---

[32] http://www.ai.sri.com/pkarp/xol/

[33] http://www.daml.org/

[34] *Defense Advanced Research Projects Agency,* www.darpa.mil/

[35] http://www.ontoknowledge.org/oil/

[36] www.cordis.lu/ist/

[37] http://www.topicmaps.org/

[38] http://kmi.open.ac.uk/projects/ocml/

[39] http://delicias.dia.fi.upm.es/webODE/

## *3.1 XML*

Per què la Web Semàntica es basa en XML? Ja hem apuntat que el problema bàsic de la Web actual és que la informació publicada és comprensible per les persones, però no per les màquines. Les pàgines escrites en HTML contenen marques que determinen multitud de paràmetres relacionats amb el format del text, però no inclou la possibilitat permet introduir dades a propòsit del contingut d'aquestes pàgines.

XML permet definir etiquetes (i les relacions entre elles) que tindran un significat relacionat amb l'estructura del document. XML incorpora algunes etiquetes predeterminades, però permet a més que l'usuari defineixi noves etiquetes i ampliï les existents. Es tracta doncs d'un llenguatge formal, estructurat, ampliable i fàcil d'utilitzar ja que fa servir la mateixa nomenclatura d'etiquetes que HTML.

Mitjançant un exemple es pot comprendre millor la funció de XML i la manera de tractar un contingut respecte a HTML:

a) Volem comprar un llibre que tracti XML. Fem una cerca a la pàgina web del FNAC i obtenim, entre d'altres, el títol següent:

| | |
|---|---|
| **XML 2005** Óscar González , Anaya Multimedia | Precio Editor: 12,30 €<br>**Precio Fnac: 11,69 €** |

Font: FNAC, www.fnac.es

**Figura 3.2 Exemple d'informació a la Web**

b) La representació en HTML d'aquesta informació és:

```
<td width="650" class="arial11">
<a
href="http://www.fnac.es/dsp/?servlet=extended.HomeExtendedServlet&Code1=109002295&Code2=
211&prodID=526430" class="arial11bold">XML 2005</a> 
<a
href='http://www.fnac.es/dsp/?servlet=actor.HomeActorServlet&Code1=3614941518&Code2=210&act
orID=113632&prodID=526430' class='arial11'>Óscar González</a>, 
Anaya Multimedia<span class="arial11normallibros"></span>

<td width="200" align="center">
  <table align="center" border="0" cellpadding="0" cellspacing="0">
    <tr>
      <td NOWRAP align="center">
        <span class="arial11">Precio Editor: 12,30 &euro; </span>
      </td>
    </tr>
    <tr>
      <td align="center" NOWRAP>
      <span class="arial11bold">Precio Fnac: 11,69 &euro;</span>
      </td>
    </tr>
  </table>
</td>
```

c) Una possible descripció XML podria ser:

```
<?xml version="1.0"?>

<llibre>
  <titol> XML 2005 </titulo>
  <llocLlibre
      href="http://www.fnac.es/dsp/?servlet=extended.HomeExtendedServlet&Code1=109002295&Code2=211&prodID=526430" />
  <autor> Óscar González </autor>
  <llocAutor
      href="'http://www.fnac.es/dsp/?servlet=actor.HomeActorServlet&Code1=3614941518&Code2=210&actorID=113632&prodID=526430" />
  <editorial> Anaya Multimedia </editorial>
  <preuEditor> 12.30 </preuEditor>
  <preuFNAC> 11.69 </preuFNAC>
</llibre>
```

XML s'utilitza també com a llenguatge per a l'intercanvi d'informació entre aplicacions de sistemes diferents. Suposem una organització que utilitza diferents sistemes d'informació: una aplicació de gestió comptable (com Contaplus), un sistema de stocks i ventes (de tipus SAP) i un gestor de bases de dades de propòsit general (per exemple SQL Server). Si volem realitzar una operació que combini informacions que es trobem en cadascun d'aquests sistemes, ens enfrontem possiblement a un problema que suposarà varies hores de feina, ja que cada sistema té les seves particularitats per a realitzar-ne consultes, i després caldrà creuar manualment els diferents resultats. Seria clarament convenient l'existència d'un únic llenguatge estàndard –com XML– amb el qual poder accedir tots i cadascun dels sistemes.

A partir de diferents necessitats d'usuaris i programadors, s'han concebut el que s'anomenen "tecnologies XML", un conjunt de mòduls que ofereixen determinats serveis. En destaquen els més importants:

- XSL (*Extensible Stylesheet Language*)[40]: Llenguatge extensible de fulles d'estil CSS[41]. El seu objectiu principal és mostrar com hauria d'estar estructurat el contingut, de quina forma cal fer el disseny de continguts i com s'hauria de presentar la paginació del document (ja sigui en un entorn web o altre).

- XPath (*XML Path Language*)[42]: Llenguatge de camins XML. Especifica com accedir a parts d'un document XML per a l'ús posterior via XPointer.

- XLink (*XML Linking Language*)[43]: Llenguatge d'enllaç XML. Descriu una manera estàndard d'afegir hiper-enllaços a un fitxer XML.

[40] http://www.w3.org/TR/xsl/

[41] http://www.w3.org/TR/REC-CSS2/

[42] http://www.w3.org/TR/xpath

[43] http://www.w3.org/TR/xlink/

- XPointer (*XML Pointer Language*)[44]: Llenguatge de direccionament XML. Permet l'accés a l'estructura interna d'un document XML, és a dir, als seus elements, atributs i continguts. Un XPointer funciona com una URL, però en comptes d'apuntar a documents en la Web apunta a porcions de dades dins d'un fitxer XML.

- XQL (*XML Query*)[45]: Llenguatge de Consulta XML. Facilita l'extracció de dades des d'un document XML. Permet realitzar consultes flexibles per a extreure dades d'un document XML de la Web.

- XQuery[46], és la proposta de W3C com a llenguatge que permet manipular i generar documents XML a partir de fonts XML, bases de dades relacionals o objectes. XQuery és a XML com SQL és als SGBDR[47].

La creació d'aplicacions que utilitzin tecnologies XML comporta avantatges com:

- Estandardització de la Informació.
- Integració d'aplicacions.
- Portabilitat de la informació.
- Compatibilitat entre sistemes.
- Millora de l'accés a la informació.

L'any 2001 apareix XML-Schema[48]. Es tracta d'una ampliació de XML 1.0 orientat a dades. Consta de tres punts importants: definició de tipus de dades, de manera que reconeix per exemple si una xifra és de fet un enter o una data; mètodes per estructurar i validar documents; regles d'utilització de XML-Schema. Proporciona un esquema abstracte de dades, de manera que ofereix una possibilitat real d'interoperabilitat entre aplicacions de diferents sistemes.

Es mostra a continuació la definició d'un tipus de dades amb XML-Schema:

```
<xs :schema xmlns :xs="http://www.w3.org/2001XMLSchema"
  targetNamespace="http://www.uoc.edu/alumnes"
  xmlns="http://www.UOC.edu/alumnes"
<xs :element name="alumnes">
 <xs :complexType>
  <xs :sequence>
      <xs :element name="alumne" minOccurs="1" maxOccurs="2000" type="TipusAlumne">
  </xs :sequence>
 </xs :complexType>
</xs :element>
```

[44] http://www.w3.org/TR/xptr/

[45] http://www.w3.org/TR/xquery-requirements/

[46] http://www.w3.org/XML/Query/

[47] Veurem aquests sistemes en el capítol 5.

[48] http://www.w3.org/XML/Schema

```
<xs :complexType name="TipusAlumne">
  <xs :sequence>
    <xs :element name="nom" type="xs :string" />
    <xs :element name="cognom" type="xs :string" />
  </xs :sequence>
  <xs :attribute name="dni" type="xs :string" />
  </xs :complexType>
<xs :/schema>
```

## *3.2 RDF*

Com hem vist en el punt anterior, XML permet estructurar documents, però té un gran inconvenient: la possibilitat que ofereix a l'usuari de crear etiquetes personalitzades per a estructurar un document, fa que cada usuari generi pràcticament un nou llenguatge, un vocabulari d'etiquetes personal. Això implica a més, un problema a nivell de portabilitat, ja que un document estructurat en sistema A no serà comprensible per un sistema B que utilitzi un conjunt d'etiquetes diferent. Apareix doncs la necessitat de definir un estàndard capaç d'estructurar documents i que atorgui un significat comú i únic per a cada etiqueta.

RDF[49] es defineix l'any 1999 (tot i que no s'estableix com estàndard fins el 2004, junt amb OWL) com un "marc de descripció de recursos" (*Resource Description Framework*). Efectivament es tracta d'un llenguatge de descripció de recursos de la xarxa. RDF ofereix una estructura semàntica inambigua (gràcies a l'ús d'URI) que permet codificar, intercanviar i processar metadades. RDF proporciona també regles per a descriure conceptes de manera que siguin comprensibles pels ordinadors i proporciona un mitjà per a definir vocabularis amb propietats ben definides per a la descripció dels recursos d'una comunitat.

Es basa en una estructura de "tripletes" (o enunciats), és a dir conjunts de tres elements, que són:

- Recurs: un recurs és qualsevol element que es pugui identificar inequívocament (amb URI).
- Enunciat: Atribut o propietat d'un recurs. També s'associa una URI a un enunciat. Un enunciat es pot vincular amb d'altres atributs.
- Propietat: Relaciona el valor d'una propietat amb un recurs.

El model relacional RDF es presenta com un graf dirigit on els recursos i enunciats són nodes, i les propietats són arcs dirigits en el sentit recurs-enunciat:

[49] http://www.w3.org/RDF/

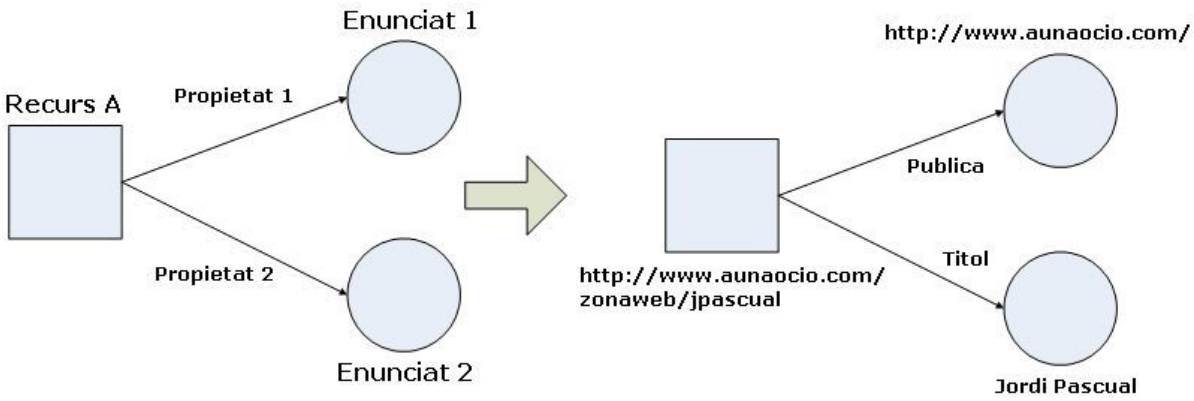



**Figura 3.3 Exemple de "tripleta" RDF**

RDF utilitza la sintaxis XML. El graf anterior quedaria expressat de la forma:

```
<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://www.aunaocio.com/zonaweb/jpascual">
  <dc:title>Jordi Pascual</dc:title>
  <dc:publisher>aunaocio</dc:publisher>
</rdf:Description>
</rdf:RDF>
```

RDF permet representar parelles del tipus recurs-propietat i relacions entre recursos, però no inclou mecanismes per a descriure propietats, ni les relacions entre aquestes propietats i els recursos corresponents; no aporta cap ajut per a descriure semàntiques. RDF-Schema, també conegut com *RDF vocabulary description language*[50], defineix classes i propietats necessàries per a descriure altres classes, propietats i recursos. RDF-Schema recupera totes les característiques de RDF i li afegeix semàntica; és, doncs, una ontologia.

El sistema de classes i propietats és molt similar al que s'empra en la programació orientada a objectes. Podem definir, per tant, una relació de jerarquia o d'herència entre classes.

[50] http://www.w3.org/TR/rdf-schema

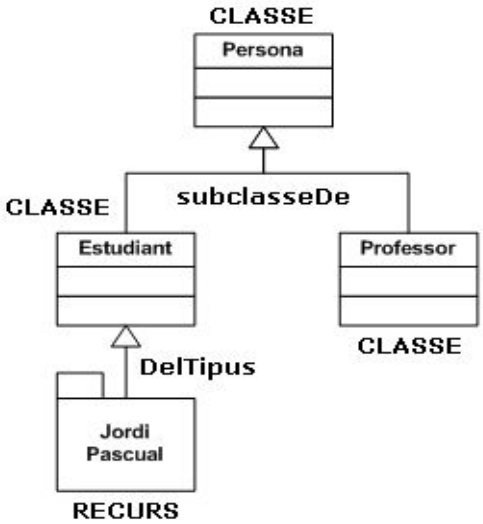



**Figura 3.4 Relacions de classes amb RDF-Schema**

L'esquema de la figura 3.4 s'expressa com segueix:

```
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
<rdf:Description rdf:ID="Persona">
<rdf:Description rdf:ID="Estudiant"><rdf:subClassOf rdf:resource="#Persona"/></rdf:Desciption>
<rdf:Description rdf:ID="Professor"><rdf:subClassOf rdf:resource="#Persona"/></rdf:Desciption>
<rdf:Description rdf:ID="Jordi Pascual"><rdf:type rdf:resource="#Estudiant"/></rdf:Desciption>
</rdf:RDF>
```

## *3.3 OWL*

*"OWL is to knowledge management systems what SQL is to database management systems."*[51]

OWL (*Web Ontology Language*)[52] proporciona un llenguatge per a la definició d'ontologies estructurades basades en la Web. Els llenguatges anteriors es van idear pensant en ontologies per a comunitats d'usuaris específiques (sobretot en el camp de les ciències i en aplicacions de comerç electrònic), però no es van definir per a ser compatibles amb l'arquitectura de la *World Wide Web* en general, i de la Web Semàntica en particular.

OWL estén RDF-Schema per a poder expressar relacions complexes entre diferents classes i ofereix una precisió més gran a l'hora d'establir restriccions a classes i propietats específiques. OWL aprofita les característiques de RDF per afegir les següents capacitats a les ontologies:

- Capacitat de ser distribuïdes a través de diferents sistemes
- Escalabilitat segons les necessitats de la Web
- Compatible amb els estàndards Web d'accessibilitat i internacionalització
- Ontologies obertes i extensibles

[51] Ian Horrocks, "Laying the foundation for the next-generation Web", *IST Results*, Març 2005.

[52] http://www.w3.org/TR/2004/REC-owl-ref-20040210/

OWL permet definir ontologies que poden ser utilitzades a través de diferents sistemes. Usuaris, bases de dades i aplicacions necessiten compartir informació específica d'una àrea de coneixement donada (com les finances, la medicina, l'esport, etc). Les ontologies s'encarreguen de definir els conceptes necessaris per a descriure i representar d'un determinat domini i les relacions existents entre aquests conceptes. En aquest sentit, mitjançant OWL es poden definir i aplicar restriccions sobre: classes (conceptes), propietats de classes i relacions entre classes.

OWL consta de tres nivells classificats segons la seva capacitat d'expressió: OWL-Lite, és la versió més senzilla d'ontologia i, per tant, la més restringida; OWL-DL afegeix a la versió anterior el que s'anomena "descripcions lògiques"[53], essent més potent que la versió anterior però encara amb certes restriccions; OWL-Full, que inclou tota la potència de OWL i RDF, sense restriccions, i sense garantia d'eficàcia.

Proposem a continuació un petit exemple d'ontologia descrita amb OWL.

El fragment de fitxer OWL següent es correspon amb l'exemple del punt 2.3.2. Mostra com les classes «Estudiant» i «Professor» són subclasses de «Persona», i que la classe «Persona» té les propietats «NIF» i «edat». Es mostra també com es defineix la representació de la classe «Estudiant» en diferents idiomes i algunes relacions entre classes com «esResponsableDe».

```
owl:Class rdf:ID="Persona">

<owl:DatatypeProperty rdf:ID="edat">
  <rdfs:range
      rdf:resource="http://www.w3.org/2001/XLMSchema#nonNegativeInteger"/>
</owl:DatatypeProperty>

<owl:DatatypeProperty rdf:ID="NIF"
  <rdf:type="http://www.w3.org/2002/07/owl#FunctionalProperty">
  <rdfs:domain rdf:resource="#Persona"/>
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#int"/>
</owl:DatatypeProperty>

<owl:Class rdf:ID="Estudiant">
 <rdfs:subClassOf rdf:resource="#Persona"/>
 <rdfs:label xml:lang="fr">estudiant</rdfs:label>
 <rdfs:label xml:lang="en">student</rdfs:label>
 <rdfs:label xml:lang="fr">étudiant</rdfs:label>
</owl:Class>

<owl:Class rdf:ID="Professor">
  <rdfs:subClassOf rdf:resource="#Persona"/>
</owl:Class>
```

[53] La lògica de descripcions o DL (*Description logics*) fa referència: a descripcions de conceptes usades per a descriure un domini; a la semàntica basada en la lògica que ofereix una traducció en lògica de predicats de primer ordre.

```
<owl:Class rdf:ID="personesUniversitat">
  <owl:unionOf rdf:parseType="Collection">
    <owl:Class rdf:about="#Professor"/>
    <owl:Class rdf:about="#Estudiant"/>
  </owl:unionOf>
</owl:Class>

<owl:Class rdf:about="#Professor">
  <rdfs:subClassOf>
    <owl:Restriction>
    <owl:onProperty rdf:resource="#esResponsableDe"/>
    <owl:someValuesFrom rdf:resource="#Asignatura"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
```

## *3.4 Serveis Web*

Pel que fa als Serveis Web, trobem diferents estàndards basats en XML ja assentats: WSDL (*Web Service Description Language*)[54], per a la descripció de serveis; SOAP (*Simple Object Access Protocol*)[55], per a l'intercanvi de missatges entre serveis; UDDI (*Universal Description, Discovery and Integration*)[56], sistema de directori per a localitzar serveis dinàmicament.

BPEL4WS (*Business Process Execution Language for Web Services*), desenvolupat per IBM, afegeix a WSDL descripcions per a coordinar el flux d'operacions en processos de negocis i transaccions. Tant la Java 2 Enterprise Edition (J2EE) de Sun, com la plataforma .NET44 de Microsoft inclouen eines per a desenvolupar i executar Serveis Web[57]. A més, empreses com IBM, SAP o Microsoft comencen a oferir registres UDDI[58].

Últimament s'han descrit llenguatges i plataformes addicionals per a la creació de Serveis Web basats en ontologies. Destaquem OWL-S[59], definit a partir de DAML l'any 2004; WSMO (*Web Service Modeling Ontology*)[60] creat el 2005; SWSL (*Semantic Web Services Language*)[61]. Podem trobar d'altres informacions sobre les activitats entorn dels Serveis Web a l'adreça: http://www.w3.org/2002/ws/.

---

[54] http://www.w3.org/TR/wsdl

[55] http://www.w3.org/TR/soap/

[56] http://www.uddi.org/

[57] http://java.sun.com/webservices/jwsdp/

[58] https://uddi.microsoft.com/register.aspx

[59] www.daml.org/services/owl-s/

[60] http://www.wsmo.org/

[61] http://www.daml.org/services/swsl/

## 3.5 Comunicació entre Agents

La comunicació és un aspecte important dins una comunitat d'agents, ja que permet compartir i enviar missatges entre agents. Els mètodes tradicionals permeten una comunicació, però no són suficients per aconseguir un comportament social entre els agents. Per a que això sigui possible es necessari que els missatges tinguin un significat i un contingut semàntic.

S'estan desenvolupant diversos llenguatges per a la transmissió de missatges. Els més importants són: ACL-FIPA (*Agent Communication Language FIPA*)[62], és un estàndard per a la comunicació d'agents desenvolupat per l'organització internacional FIPA (*Foundation for Intelligent Physical Agents*)[63]; KQML (*Knowledge Query and Manipulation Language*)[64].

## 3.6 Altres protocols

### 3.6.1. SKOS-Core

SKOS-Core (Simple Knowledge Organization Systems) [65] és una proposta de W3C per a la utilització de sistemes d'organització de coneixements en l'entorn de la Web Semàntica. De fet, és considerat com un complement de OWL.

Els sistemes d'organització de coneixements o KOS (*Knowledge Organization Systems*) comprenen tots els tipus d'esquemes que permeten organitzar la informació i gestionar el coneixement. Els KOS són mecanismes que es troben en el cor de cada biblioteca, museu o arxiu. Els tesaures són un tipus concret de KOS.

### 3.6.2 PICS

L'especificació PICS (*Platform Independent Content Rating Scheme*) [66] permet l'associació entre etiquetes (metadades) i els continguts a Internet. PICS es va idear en un principi per a gestionar el control parental, per a gestionar les connexions de menors a Internet. Però el seu ús s'ha estès a d'altres utilitats com: privacitat, llicències, firma de codis, protecció de la propietat intel·lectual, etc. Ajuda a determinar el nivell de conveniència de determinats fitxers de dades segons l'entorn on es troba l'usuari. PICS constitueix una plataforma sobre la qual s'han construït diferents serveis i aplicacions de filtratge de continguts. Es presenta a continuació un exemple extret de l'especificació de HTML 4.01:

*<META http-equiv="PICS-Label" content='(PICS-1.1 "http://www.gcf.org/v2.5" labels on "1994.11.05T08:15-0500" until "1995.12.31T23:59-0000" for "http://w3.org/PICS/Overview.html" ratings (suds 0.5 density 0 color/hue 1))'>*

---

[62] http://www.limsi.fr/Individu/jps/enseignement/examsma/2004/DURO_HOARAU/

[63] http://www.fipa.org

[64] www.cs.umbc.edu/kqml/

[65] http://www.w3.org/2004/02/skos/

[66] http://www.w3.org/PICS/

Es pot assignar una etiqueta PICS a qualsevol recurs que tingui URI, i existeixen dos procediments d'assignació. El primer consisteix en encarregar a un organisme independent la monitorització del recurs o lloc web; les etiquetes s'enregistren en el servidor d'aquest organisme. Per al segon mètode, cal contactar un servei de classificació per a obtenir la metadada corresponent que es colorarà en la pròpia pàgina. Un d'aquests serveis gratuïts és el generador d'etiquetes PICS de *Vancouver-Webpages*[67].

### 3.6.2 Consultes

Una de les tasques pendents en relació amb RDF és la creació d'un llenguatge de consulta, similar al SQL de las bases de dades, que permeti expressar cerques complexes mitjançant una sintaxis declarativa senzilla. En l'espera de trobar un estàndard vàlid que convingui a tothom, s'han consolidat diferents iniciatives particulars com: RDQL (*RDF Query Language*)[68], creat per *Hewlett Packard* l'any 2003; RQL (*RDF Schema Query Language*)[69], de l'institut ICS-FORTH[70]; SeRQL (*Sesame RDF Query Language*)[71], de l'empresa holandesa *Aidministrator Nederland*; SPARQL (*SPARQL Protocol and RDF Query Language*)[72], és finalment la proposta publicada per W3C el juliol de 2005.

De moment ja s'ha posat la primera pedra: XQuery[73], és un llenguatge que permet manipular i generar documents XML a partir de fonts XML, bases de dades relacionals o objectes. XQuery és a XML com SQL és als SGBDR.

---

[67] http://vancouver-webpages.com/PICS/HOWTO.html

[68] http://www.w3.org/Submission/2004/SUBM-RDQL-20040109/

[69] http://zeus.ics.forth.gr/forth/ics/isl/publications/paperlink/dql-rdf.pdf

[70] http://www.ics.forth.gr/

[71] http://www.openrdf.org/doc/sesame/users/ch06.html

[72] http://www.w3.org/TR/rdf-sparql-query/

[73] http://www.w3.org/XML/Query/

# 4. Aplicacions de la Web Semàntica

RDF s'integra en diverses aplicacions incloent:

- catàlegs de biblioteques
- directoris mundials
- sindicació i agregació de noticies, software i continguts
- col·leccions personals de música, fotos i esdeveniments

Tractarem en aquest punt aquelles tecnologies web que utilitzen, de forma més o menys activa, les particularitats de la Web Semàntica. Veurem noves funcionalitats que apareixen a partir de la idea de SW, i d'altres que ja existien abans i que prenen una nova significació amb el suport de RDF i les ontologies.

## *4.1 Sindicació i RSS*

L'acrònim RSS s'aplica als protocols següents:

- *Rich Site Summary* (RSS 0.91)
- *RDF Site Summary* (RSS 0.9 i 1.0)
- *Really Simple Syndication* (RSS 2.0)

RSS és un vocabulari RDF que permet la catalogació d'aquella informació que s'actualitza molt sovint (noticies i esdeveniments) de manera que sigui possible compartir-la de forma eficient amb d'altres usuaris, aplicacions o llocs web. A aquest procés se'l coneix amb el nom de "sindicació".

Els arxius RSS contenen metadades sobre fonts d'informació especificades pels usuaris. La funció principal és notificar als usuaris, de forma automàtica, qualsevol canvi que es produeixi en els recursos d'interès seleccionats.

Els programes que llegeixen i presenten informació RSS de diferents fonts s'anomenen "agregadors" o *Newsfeed*[74], en anglès. Un agregador de notícies és doncs, un sistema que recull les capçaleres de les noves informacions aparegudes en els llocs d'Internet definits per l'usuari, de manera que no calgui accedir successivament a cadascun dels recursos per a veure si hi ha nous continguts. S'utilitzen principalment per a consultar Blogs, tot i que alguns portals de noticies comencen a implementar la sindicació (per exemple, www.elmundo.es). S'anomenen *Feeds* (de l'anglès "alimentadors") als elements que proporcionen la informació als agregadors. Es tracta de pàgines en format RSS (o Atom[75]) que permeten visualitzar el títol de la noticia, la data de publicació, un petit resum sobre el seu contingut i la seva URL.

[74] Cada dia apareixen nous agregadors, com els recentment publicats per *Google* i *Yahoo*. Fins i tot alguns navegadors incorporen un *NewsFeed*, com Opera 7.5.

[75] Atom és un format semblant a RSS. Més informació a: http://atomenabled.org/

L'usuari pot saber quins llocs o portals web implementen la possibilitat de sindicació de continguts per que hi trobarà un logotip de color taronja amb les sigles XML o RDF o RSS, com el de la figura 4.1.

**Figura 4.1 El "botó taronja" de la sindicació**

Un exemple de l'aplicació de RSS es pot trobar en el lloc de l'Oficina Espanyola de Noticies del W3C: http://www.w3c.es/noticias.rss.

## *4.2 FOAF*

*FOAF* (*Friend Of A Friend*)[76] és un projecte dins l'entorn de la Web Semàntica per a descriure persones, vincles entre elles, i coses que fan o creen. Es tracta d'un vocabulari RDF que permet disposar d'informació personal de forma senzilla i simplificada per a ser processada, compartida i reutilitzada. Es mostra a continuació un exemple de fitxer FOAF:

```
<?xml version="1.0" standalone="yes"?>
<rdf:RDF
      xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
      xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
      xmlns:foaf="http://xmlns.com/foaf/0.1/">

<foaf:Person>
   <foaf:name>Jordi Pascual</foaf:name>
</foaf:Person>
</rdf:RDF>
```

Una conseqüència de FOAF és FOAF-a-Matic[77], una aplicació Javascript que permet crear una descripció FOAF d'una persona. Amb aquesta descripció, les dades personals passen a integrar un motor de cerca on serà possible obtenir informació d'una persona en concret i de les comunitats a les que pertany, d'una forma ràpida i senzilla.

DOAC (*Description Of A Career*)[78] és una extensió del vocabulari *FOAF* que permet compartir un curriculum de forma que pugui ésser processat per qualsevol aplicació. A l'estil d'un *Curriculum Vitae* clàssic, inclou informació sobre educació, experiència laboral, publicacions, idiomes i altres habilitats.

FOAFNAUT[79] i FOAF Explorer[80] són navegadors que mostren relacions entre estructures FOAF.

---

[76] http://www.foaf-project.org/

[77] http://www.ldodds.com/foaf/foaf-a-matic.es.html

[78] http://ramonantonio.net/doac/

[79] http://www.foafnaut.org/

[80] http://xml.mfd-consult.dk/foaf/explorer/

## *4.3 Aplicacions d'ontologies*

Veurem en aquest punt algunes de les funcionalitats de la web que poden aprofitar-se de la tecnologia de les ontologies.

### 4.3.1 Portals Web

Un portal Web és un lloc d'Internet que proporciona informació relacionada amb un domini concret (per exemple marmiton.org, un portal de receptes de cuina). Un portal Web ofereix als seus usuaris una sèrie de serveis com: rebre noticies, comunicació entre usuaris (*blogs*)[81], construcció de comunitats, trobar enllaços cap a d'altres recursos web d'interès.

L'èxit d'un portal web recaurà bàsicament en la capacitat d' indexació dels continguts i en la sindicació[82]. Una gran part de la informació d'un portal és enviada pels membres de la comunitat, que queda enregistrada en una de les subcategories del portal (per exemple, receptes de cuina vegetarianes). Però un índex de continguts no és una eina prou potent per a la cerca de continguts; es fa necessària l'existència d'una ontologia relacionada amb la temàtica del portal, que defineixi els conceptes relacionats amb el domini i els possibles vincles entre conceptes. Exemples de portals basats en ontologies són *OntoWeb*[83] i *The Open Directory Project*[84].

### 4.3.2 Col·leccions multimèdia

Gràcies a les ontologies es poden assignar anotacions semàntiques a les col·leccions d'imatges, àudio o vídeo. Ja hem comentat les dificultats que tenen les màquines actualment per a comprendre un text en llenguatge natural; és obvi que extreure el significat semàntic d'un fitxer multimèdia posa encara més problemes. Per això, aquests tipus de recursos s'indexen amb capçaleres i metadades. El problema és que diferents usuaris poden descriure aquests objectes de maneres diferents. Idealment, les ontologies capturen informacions addicionals sobre el domini de cerca i milloren la recopilació d'imatges.

Hi ha dos tipus d'ontologies multimèdia: relatives al medi, contenen descripcions de propietats i taxonomies de diferents tipus de media (per exemple, la durada d'un vídeo); relatives al contingut, descriuen el domini al qual correspon un recurs. Existeixen diferents protocols per a entorns multimèdia: SMIL (*Synchronised Multimedia Integration*)[85], SVG (*Scalable Vector Graphics*)[86], VoiceXML[87].

---

[81] El mot bitàcora prové de l'antic "Quadern de Bitàcora", redactat pel capità d'un vaixell: registre cronològic de les maniobres i decisions preses al llarg de la jornada. Així doncs una bitàcora, *Weblog* o *Blog* és un lloc web que conté anotacions ordenades cronològicament.

[82] Veure capítol 4.1.

[83] http://www.ontoweb.org/

[84] http://dmoz.org/

[85] www.w3.org/AudioVideo/

Un exemple seria una col·lecció d'imatges publicades en el lloc web d'un museu. La Universitat de Hèlsinki porta a terme un projecte de portal semàntic dels museus finlandesos[88].

### 4.3.3 Gestió de web's corporatives

Les grans companyies (per exemple Atos Origin)[89] acostumen a tenir una pila de pàgines web relacionades amb temes diferents: catàlegs de productes, retalls de premsa, ofertes de treball, e-bussiness, e-learning, recerca i tecnologies, etc. Les ontologies poden ajudar a indexar tots aquests documents i proporcionar una extracció d'informació més precisa. Hi ha empreses que disposen de la seva pròpia taxonomia per organitzar la seva informació, però una sola ontologia acostuma a ser insuficient ja que només mostren un punt de vista del domini (el de l'empresa en qüestió, sense tenir en compte altres entitats del mateix camp). Es fa necessària la possibilitat de treballar amb múltiples ontologies, ja que no tindran el mateix punt de vista d'un mateix concepte un enginyer de sistemes de la companyia A i un comercial de zona de l'empresa B.

### 4.3.4 Agents i Serveis Web

La Web Semàntica proporciona als agents la possibilitat de comprendre i integrar diverses fonts d'informació. Un exemple d'agent semàntica utilitzant serveis web seria un "planificador d'activitats", capaç d'inferir les preferències d'un usuari (com tipus de pel·lícules, plats preferits, tipus de lectura) i organitzar un cap de setmana en funció de tota aquesta informació. L'efectivitat d'aquesta tasca dependrà de la quantitat i qualitat de l'oferta existent en l'entorn i en la capacitat de definició del perfil d'usuari (que emmagatzema les seves preferències). Durant el procés de cerca d'aquestes activitats s'accedirà a serveis específics, que poden incloure sistemes de puntuació (per exemple un portal dedicat al cine com http://www.allocine.fr, amb una valoració de les pel·lícules més votades pel públic i per la premsa especialitzada).

*Agentcities*[90] és un exemple d'iniciativa que estudia l'ús d'agents en un entorn de serveis web distribuïts. Pretén crear una xarxa d'agents que representin una ciutat i els serveis que pot oferir aquesta ciutat. En una primera etapa es defineixen serveis orientats a l'oci i el consum, però més endavant es pot aplicar a serveis de comerç electrònic (*business to business* i *business to enterprise*).

---

[86] www.w3.org/Graphics/SVG/

[87] www.voicexml.org/

[88] http://www.cs.helsinki.fi/group/seco/

[89] http://www.atosorigin.com

[90] http://www.agentcities.net/

### 4.3.5 Informàtica omnipresent

El conjunt de petits dispositius electrònics sense fils i portables formen el paradigma de la informàtica omnipresent. Constantment apareixen nous aparells com telèfons portables multimèdia amb connexió a Internet, reproductors mp3, discs durs externs, aparells de fotos digitals i en general tot tipus de dispositius relacionats d'alguna manera amb els ordinadors. La tendència (l'exemple més clar son els telèfons d'última generació) és aconseguir utilitzar tota la potència dels serveis que ofereix Internet des de qualsevol punt on es trobi l'usuari. Això implica la necessitat d'intercomunicar diferents dispositius, de fabricants diferents, amb propòsits diferents i utilitzant tecnologies diferents. Caldrà doncs una ontologia que descrigui les característiques de cada dispositiu, el tipus d'accés que utilitza, les regles d'ús definides pel propietari i altres requeriments tècnics relacionats amb la connexió d'un element determinat a la xarxa ad hoc.

En aquest sentit s'està treballant en esquemes RDF per a dispositius com CC/PP (*Composite Capability/Preference Profile*)[91] i UAProf (*WAP User Agent Profiling Specification*)[92], junt amb protocols de serveis web semàntics (*DAML-S*)[93]. W3C ha creat un grup de treball anomenat *Mobile Web Initiative*[94] encarregat d'establir un estàndard per a crear continguts i llocs Web per a dispositius portables.

### 4.3.6 Buscadors semàntics

Els buscadors semàntics són un exemple més d'aplicacions basades en Web Semàntica. L'objectiu en aquest cas és satisfer les expectatives de cerca d'informació per part d'un usuari que necessita respostes precises.

Exemples de buscadors semàntics són: *Swoogle*[95], busca documents emprant RDF i OWL, fou dissenyat per la UMBC (*University of Maryland Baltimore County*); *Kartoo*[96]: motor de cerca més orientat a gràfics que a text creat per dos cosins francesos (*Laurent* i *Nicholas Baleydier*), presenta la informació en gràfics semàntics; el mateix equip de Kartoo va presentar en 2004 el buscador *Ujiko*[97], que genera un perfil d'usuari a mesura que es realitzen cerques permetent definir llocs preferits i llocs indesitjables basant-se en un sistema de punts.

---

[91] http://www.w3.org/Mobile/CCPP/

[92] http://w3development.de/rdf/uaprof_repository/

[93] www.daml.org/services/

[94] http://www.w3.org/Mobile/

[95] http://swoogle.umbc.edu

[96] http://www.kartoo.com

[97] http://www.ujiko.com

### 4.3.7 Altres aplicacions

Es poden trobar altres exemples d'aplicacions basades en la Web Semàntica al lloc de W3C: *SWAD-Europe: Semantic web applications - analysis and selection*[98].

També es recomana consultar el *WSIndex* (*Web Services Directory*)[99]: un portal d'informació sobre els nous serveis web que inclou múltiples enllaços relacionats amb *XML*, Web Semàntica, *Weblogs*, *Web Services*, etc.

En l'Annex 2 es refereix una llista de projectes existents entorn de la Web Semàntica, i d'aplicacions específiques nascudes a ran d'aquests projectes.

[98] http://www.w3.org/2001/sw/Europe/reports/open_demonstrators/hp-applications-survey

[99] http://www.wsindex.org/

# 5. Bases de Dades i RDF

En aquest apartat s'explica què són i per a què serveixen els SGBD. A continuació es mostren els SGBD més importants del mercat descrivint breument cadascun d'ells. Finalment s'indiquen quins d'aquests sistemes suporten XML (la base de la Web Semàntica).

## *5.1 Definició*

Els SGBD (Sistemes Gestors de Bases de Dades)[100] són un programari específic, dedicats a servir de interfície entre les bases de dades i les aplicacions que les utilitzen. Els SGBD incorporen una sèrie de serveis per a la gestió de bases de dades: permeten l'accés a les dades de manera senzilla, gestionen els permisos d'accés a dades per a múltiples usuaris, manipulen els continguts de les bases de dades (inserció, supressió, modificació).

L'arquitectura ANSI/SPARC[101], va definir l'any 1975 els tres nivells d'abstracció per a u sistema de gestió de bases de dades:

- Nivell intern (o físic): estableix com s'han d'enregistrar les dades i els mètodes per accedir-hi.
- Nivell conceptual: anomenat també MCD (model conceptual de dades) o MLD (model lògic de dades). Defineix l'ordenació de la informació dins la base de dades.
- Nivell extern: determina com es mostra la informació (vistes d'usuari).

Existeixen cinc models de SGBD diferenciats segons el tipus de representació de dades utilitzat:

- Model jeràrquic: Fou el primer model de SGBD. Les dades es classifiquen en forma d'arbre descendent i s'insereixen punters per a les relacions.
- Model de xarxa: Treballa amb punters com el model jeràrquic, però sense respectar l'estructura d'arbre.
- Model relacional: Les dades s'enregistren en taules de dues dimensions (files i columnes). Aquests tipus de sistemes reben el nom de SGBDR (SGBD Relacionals) i són els més utilitzats. S'aplica la teoria de relacions per manipular les dades.

[100] En alguns texts apareix l'acrònim anglès, DBMS (*DataBase Management System*).

[101] *ANSI Standards Planning And Requirements Committee.*

- Model deductiu: Utilitza taules per emmagatzemar els dades, però es manipulen gràcies a la teoria de predicats.
- Model objecte: les dades s'organitzen en classes, utilitzant la teoria d'objectes de la Programació Orientada a Objectes. Aquests sistemes s'anomenen SGBDO.

## 5.2 Per a què serveixen?

En general un SGBD inclou les característiques següents:

- Independència: Les dades han de ser independents, de manera que el nivell físic i lògic puguin modificar-se sense que això tingui conseqüències sobre les aplicacions finals; els canvis han de ser transparents per als usuaris.
- Temps de resposta: el sistema ha de donar respostes ràpides a les consultes, això implica l'ús d'algoritmes de cerca eficients.
- Redundància mínima: cal evitar en la mesura possible l'existència d' informacions redundants, per evitar possibles errors i aprofitar l'espai de memòria.
- Integritat: cal garantir la validesa de les dades i aquestes han de guardar una coherència entre elles. Cal gestionar correctament les dades que fan referència a altres dades.
- Administració centralitzada: l'administrador del SGBD ha de poder gestionar les diferents bases de dades de manera centralitzada.
- Control de concurrència: cal permetre l'accés simultani a la base de dades de múltiples usuaris, evitant que es produeixin inconsistències en la informació.
- Seguretat: calen mecanismes de gestió del drets d'accés dels usuaris a les dades. Són necessaris també sistemes de recuperació de dades en cas d'atacs malintencionats o errors involuntaris.

## 5.3 SGBD existents

Es presenten a continuació els Sistemes de Bases de Dades més importants, classificant-los segons si són "*open source*"[102] o no.

[102] El codi obert (*open source* en anglès) és el programari el codi font del qual està públicament disponible.

### 5.3.1 Lliures

- PostgreSQL[103]:

  És un SGBDR pensat principalment per a sistemes d'explotació de tipus Linux i Unix (*AIX, BSD, HP-UX, SGI IRIX, Mac OS X, Solaris, SunOS, Tru64)*, encara que també hi ha una versió per Windows (la v8.0.4).

  Inclou la majoria de tipus de dades SQL92 i SQL99 (*integer, numeric, boolean, char, varchar, date, interval, timestamp*). També suporta l'enregistrament de fitxers binaris de gran capacitat (fins a 1Gb) com imatges, so i vídeo. Inclou també utilitats per programar en C/C++, Java, Perl, Python, Ruby, Tcl, ODBC.

  El codi font PostgreSQL està sota la llicència BSD (*Berkeley Software Distribution*). Aquesta llicència dóna llibertat a l'usuari per utilitzar, modificar i distribuir PostgreSQL; qualsevol canvi realitzat per l'usuari és propietat de l'usuari. Mida màxima d'una base de dades: il·limitat. Mida màxima d'una taula: 32TB.

- MySQL[104]:

  MySQL s'ha convertit en el SGBD més popular actualment. Des de la versió 3.23.19, MySQL està sota la llicència GPL (*General Public License*), (tant per Linux com per Windows). Això significa que pot utilitzar-se de forma gratuïta, pero tenint en compte que obliga a distribuir qualsevol producte derivat sota la mateixa llicència. Si un desenvolupador desitja incorporar MySQL en la seva aplicació però no vol distribuir-la sota llicència GPL, pot adquirir la llicència comercial.

  MySQL col·labora sovint amb PHP (és el cas de molts CMS que hem vist en el capítol 5) per a la publicació de dades a Internet.

- Firebird[105]:

  Firebird fou creat el 1981 i treballa sobre Linux, Windows, i la majoria de variants Unix. Firebird fou programat en C i $C^{++}$. Es basa en la versió 6 d'Interbase[106] i és de llicència pública. Es tracta d'un SGBD relativament petit però potent, amb requeriments de maquinari baixos.

- SQLite[107]:

  Es tracta d'un petit conjunt de llibreries C capaces de gestionar bases de dades i SQL. Tota una base de dades s'enregistra en un sol fitxer. És un gestor de llicència pública del tipus "*zero-configuration*": no necessita ser administrat ni configurat. Pot emmagatzemar fitxers binaris grans (imatges, vídeo, música). És més ràpid que altres gestors més populars. Capacitat màxima de base de dades: 2TB.

---

[103] http://www.postgresql.org

[104] http://www.mysql.com

[105] http://firebird.sourceforge.net

[106] Veure apartat 5.3.2.

[107] http://www.sqlite.org

- Ingres[108]:

  SGB *open source*. Ofereix escalabilitat, integració i flexibilitat multiplataforma (Linux, Unix, Windows). Suporta XML.

### 5.3.2 Comercials[109]

- Oracle[110]:

  És considerat un dels SGBD més complerts, i també el més costos (tan pel preu de la llicència com pels requeriments de maquinari). Les primeres versions es trobaven per a Unix i Windows. A partir de la versió 8.0.5 també és disponible per Linux. Durant molt de temps a tingut el monopoli dels servidors empresarials, tot i que els darrers anys comença a notar la competència d'altres gestors com SQL Server de Microsoft i els SGBD *open source* en general.

  Inclou funcionalitats com: servidor d'aplicacions, encriptació, desenvolupament d'aplicacions, creació d'Intranets, Data Warehousing, e-business. Incorpora eines per treballar amb XML.

- Microsoft SQL Server[111]:

  És l'alternativa de Microsoft en el camp dels SGBD, i per tant només treballa en plataformes Windows. És un sistema costos i potent (semblant a Oracle en aquest sentit), que incorpora múltiples possibilitats com: suport XML, creació de llocs Intranet, e-Business, autoconfigurable, monitorització d'eseveniments, escalabilitat, Data Warehousing.

- Paradox[112]:

  Base de dades relacional bàsicament per entorns Windows, desenvolupada actualment per Corel i inclosa en el paquet ofimàtica WordPerfect Office. En els temps del MS-DOS era una base de dades molt popular. Més tard aparegué la versió per Windows, que no pogué competir amb Microsoft Access.

- Interbase[113]:

  Borland InterBase està pensat per entorns *zero-configuration* (com SQLite). És prou compacte per executar-se en sistemes maquinaris senzills.

---

[108] http://www3.ca.com/Solutions/Product.asp?id=1013

[109] Degut a que es tracta de sistemes de pagament, l'accés a les seves característiques tècniques és limitat.

[110] http://www.oracle.com

[111] http://www.microsoft.com/sql/default.mspx

[112] http://www.corel.fr

[113] http://www.borland.com/interbase/

- Informix[114]:

  SGBD creat per Informix i actualment propietat d'IBM. Abans d'esdevenir un gestor de caràcter general per a tot tipus de plataformes, estava especialitzat en aplicacions de tipus GIS (dades geogràfiques), Datawarehouse i Datamining. Incorpora eines per la publicació de dades en pàgines web.

- DB2[115]:

  També propietat d'IBM, és un SGBD clarament de caràcter general, per tot tipus de plataformes. És flexible, escalable, fàcil d'utilitzar i amb un preu ajustat. Acostuma a treballar en equip amb WebSphere per a l'implementació de portals de continguts.

- Sybase SQL Server[116]:

  Multiplataforma. Data Mining, Warehousing, desenvolupament d'aplicacions. Suporta XML i Serveis Web.

- Fox Pro[117]:

  Producte de Microsoft. L'última versió proporciona components XML i Web Services.

- MS Access[118]:

  SGBD de Microsoft (per a Windows) de caràcter personal; és a dir, per a gestionar bases de dades petites. No és molt adequat per accessos simultanis ni per a gestionar la seguretat d'usuaris.

- mSQL[119]:

  mSQL fou creat l'any 1994 per Hughes Technologies. El seu nom ve de "mini" SQL. És un gestor lleuger amb un ús molt eficaç de la memòria. Té certes limitacions pel que fa a l'ús de SQL.

- dBase[120]:

  Fou el primer SGBD en un entorn MS-DOS i fou el sistema més venut fins a l'aparició de Windows. Després de ser venut per Borland a dataBased Intelligence Inc., ha sortit el 2005 una versió per XP que inclou fitxers XML i entorns gràfics "Drag & Drop".

---

[114] http://www.informix.com

[115] http://www.db2.com

[116] http://www.sybase.com

[117] http://msdn.microsoft.com/vfoxpro/

[118] http://msdn.microsoft.com/vfoxpro/

[119] http://www.hughes.com.au/products/msql/

[120] http://www.dbase.com

- Filemaker[121]:

  Solució per Macintosh i Windows. L'última versió permet la publicació de dades en la Web i suporta Serveis Web. Permet l'intercanvi de dades amb XML. Permet enregistrar fitxers multimèdia.

- Pervasive SQL[122]:

  Motor de bases de dades que ofereix: integritat de les dades, alt rendiment, flexibilitat, escalabilitat i baix cost.

Els gestors de bases de dades més utilitzats actualment en el mercat són els produïts per Oracle, IBM, Informix i Sybase, per aquest ordre. Tots ells, junt amb d'altres, suporten la generació de documents XML a partir de consultes SQL:

- Oracle 9i: Mòdul *Cartridge XML*. Correspondència entre documents XML i taules.
- SQL Server 2000: Correspondència entre documents XML i taules. Enregistra metadades XML.
- Sybase: Inserció de documents XML en taules.
- Informix: Tipus de dades específics per enregistrar XML.
- DB2: *XML Extender*: cerques en documents XML. DAD: èina per descomposar documents XML en diferents taules.

En la figura 5.1 es mostren els SGBD que suporten XML.

| Producte | Propietari |
|---|---|
| Oracle | Oracle |
| SQL Server 2000 | Microsoft |
| Sybase | Sybase |
| Informix | IBM |
| DB2 | IBM |
| MySQL | MySQL AB |
| SQLite | Richard Hipp |
| PostgreSQL | Berkeley Computer Science Department |
| Access 2002 | Microsoft |
| FileMaker | FileMaker |
| FoxPro | Microsoft |

Font: "*Consulting, writing, and research in XML and databases*", Ronald Bourret.[123]

**Figura 5.1 Bases de dades suportant XML**

[121] http://www.filemaker.com

[122] http://www.pervasive.com

[123] http://www.rpbourret.com

## 5.4 Un problema d'estructures

Hem vist en el capítol 5 la implicació dels CMS en la Web (i per tant també en la Web Semàntica). Recordem que aquests sistemes separen els continguts de la seva representació, i que la majoria enregistren aquests continguts en bases de dades. Per a la gestió d'un conjunt de bases de dades necessitarem un sistema gestor. D'aquí la importància dels SGBD en la Web Semàntica. D'altra banda, hem vist també que la Web Semàntica està formada per documents estructurats (en format XML i, per extensió, en RDF). Necessitem doncs establir un vincle entre XML i els SGBD. El problema principal és que estem parlant de dues plataformes bàsicament incompatibles, creades per a propòsits diferents. Necessitarem mecanismes capaços de convertir les dades enregistrades en els SGBD en format XML, i que ens permetin al mateix temps enregistrar un document XML en una base dades.

Una possible solució seria treballar directament amb bases de dades natives de XML. És evident que aquests sistemes ens garanteixen un bon funcionament amb dades estructurades, però presenten la limitació de treballar només amb formats XML. Resultaria més interessant si poguéssim reutilitzar el SGBD que gestiona les nostres bases de dades i aplicacions actuals. Els SGBD són sistemes madurs, àmpliament utilitzats i entorn dels quals existeixen múltiples aplicacions. Seria menys costos afegir la capacitat de treballar amb dades estructurades (XML) com una funcionalitat d'un SGBD que no pas migrar totes les aplicacions existents a un sistema XML natiu (suposant que això fos possible).

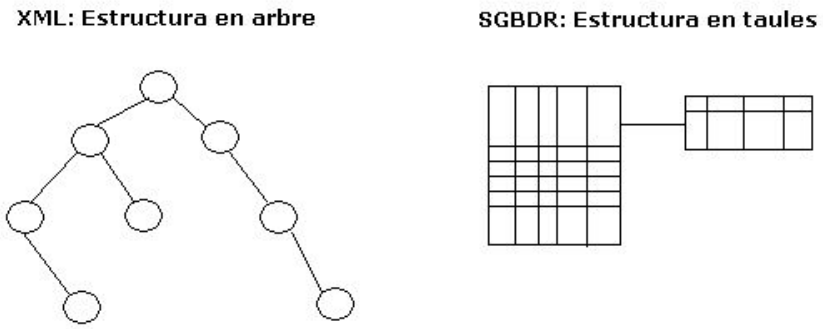



**Figura 5.2 Problemàtica XML – SGBDR**

En aquest sentit apareixen *middlewares* com: SAX (*Simple API for XML*)[124], que fou el primer "*parser*"[125] i entén la informació com una sèrie d'esdeveniments; DOM (*Document Object Model*)[126], que permet a les aplicacions accedir de forma dinàmica i actualitzar documents Web i representa la informació en una estructura d'arbre; XML-DBMS[127], és un paquet *open source* escrit en Java per la transcripció de documents XML a BD relacionals. Aquestes API's es poden aplicar per accedir a

[124] http://www.saxproject.org

[125] De l'anglès: "analitzador sintàctic".

[126] http://www.w3.org/DOM/

[127] http://www.rpbourret.com/xmldbms/index.htm

dades XML enregistrades en una BD o es pot construir una BD nativa XML amb API's com interfície d'accés i actualització de dades.

Existeixen BD amb capacitat de XML; és a dir, suporten la conversió automàtica del model relacional a XML. Els documents estructurats s'enregistren com si fossin BLOB (*Binary Large Object*)[128]. Aquestes BD utilitzen SQL com a llenguatge d'accés a les dades relacionals, i XQuery[129] per a consultar documents XML.

Recordem però els problemes que suposa l'estructuració de documents amb XML: cada usuari pot definir el sistema d'etiquetes que consideri oportú; no hi ha un vocabulari comú. Si volem que el SGBD comprengui un document en XML caldrà afegir-hi també el conjunt de normes que l'ajudaran a comprendre les etiquetes utilitzades. Aquests documents s'anomenen esquemes XML. Els més utilitzats són els DTD (*Document Type Definition*)[130] i XLST (*eXtensible Stylesheet Language Transformations*)[131].

Una manera d'evitar de treballar sempre amb esquemes associats a cada document XML poder utilitzar directament les "tripletes" de RDF i les ontologies com OWL. Malauradament aquest nivell de descripció encara no ha arribat als SGBD, tot i que hi ha alguns desenvolupadors que comencen a orientar els seus esforços en aquest sentit. Així comencem a trobar el mercat productes com Kowari[132], un sistema de bases de dades no-relacionals *Open Source* per a la gestió de metadades.

En el camp dels SGBDR, Oracle a constituït el seu *Semantic Technologies Center*[133], que comença a donar alguns fruits com un sistema de conversió de "tripletes" RDF en objectes emmagatzemables en taules, inclòs en la versió d'Oracle 10g Release 2:

> *«Oracle Spatial 10g Release 2 introduces the industry's first open, scalable, secure and reliable data management platform for RDF-based applications. New object types have been defined to manage RDF data in Oracle. Based on a graph data model, RDF triples are persisted, indexed and queried, similar to other object-relational data types. The Oracle 10g RDF database ensures that application developers benefit from the scalability of the Oracle database to deploy scalable semantic-based enterprise applications.»*[134]

---

[128] Conjunt de dades en binari enregistrades com una sola entitat en un sistema gestor de bases de dades. BLOBs són normalment imatges, vídeo i altres objectes multimèdia, tot i que també es pot tractar de trossos de codi.

[129] Veure aparat 3.6.2.

[130] http://www.xmlfiles.com/dtd/

[131] http://www.w3.org/TR/xslt

[132] http://www.kowari.org

[133] http://www.oracle.com/technology/tech/semantic_technologies/index.html

[134] http://www.oracle.com/technology/tech/semantic_technologies/index.html

# 6. Implementant Web Semàntica

En aquest capítol es dóna una visió més pràctica d'utilització dels SGBD per desar i recuperar descripcions en RDF. En aquest sentit, s'ha seguit un extensiu procés d'investigació per tal de conèixer les característiques i el funcionament de diferents productes relacionats de forma més o menys directa amb la Web Semàntica.

Abans de veure'n casos pràctics, però, cal entendre què significa "implementar Web Semàntica". Recordem que el concepte de web amb significat es basa en els conceptes de RDF i OWL, que són les eines que ens proporcionen documents estructurats, de manera que siguin comprensibles per als ordinadors. Necessitarem, per tant, en el primer nivell d'implementació, un programari que ens permeti crear, manipular i comprovar la sintaxis d'aquests tipus de dades; es tracta d'editors o "*parsers*" RDF.

Un cop generats els fitxers de partida, ens cal un sistema per a enregistrar aquesta informació. Podríem publicar directament les dades RDF en una pàgina web, però resultarà més interessant emmagatzemar tot el conjunt en bases de dades; d'aquesta manera podem afegir, modificar o consultar les dades directament i publicar a continuació els canvis en la Web. Aconseguirem una independència total entre les dades publicades i el disseny del nostre lloc web. Ens caldrà doncs triar un SGBD adequat per a treballar amb dades RDF.

Malauradament, si bé la majoria d'aquests productes inclouen utilitats per a treballar amb fitxers XML, l'ús de RDF encara no està prou estès en aquest entorn. Veurem més endavant que en la majoria de casos (a excepció feta potser d'Oracle 10g) estem obligats a utilitzar interfícies o "*middlewares*" que ens ajudaran en el procés d'importació/exportació de dades RDF dins els SGBD. Alguns d'aquests programaris han adquirit un grau d'implicació tal amb el concepte de Web Semàntica, que incorporen inclús editors, eines de consulta i utilitats per a la publicació dels resultats en la Web.

En aquest tema veurem primerament com generar i corregir fitxers RDF. A continuació es mostraran algunes de les solucions més esteses per al desenvolupament de Web Semàntica; en veurem el funcionament i les relacions amb els SGBD per el tractament de dades RDF.

Pel que fa a la publicació de les dades en la Web, actualment s'utilitza en general PHP, ja que la majoria de SGBD proporcionen un alt nivell de compatibilitat amb aquest llenguatge. Tot i això, de fet es pot utilitzar qualsevol llenguatge amb capacitat de gestionar dades enregistrades en bases de dades (gràcies a connexions de tipus ODBC o JDBC), com ara ASP[135], Phyton[136] o Perl[137].

---

[135] http://www.asp.net/

[136] http://www.python.org/

[137] http://www.perl.com/

Finalment podríem utilitzar encara una altra tecnologia: si bé constatem que encara és aviat per dir que els SGBD suporten RDF, sí podem assegurar que la majoria d'aquests sistemes estan preparats per treballar directament amb XML. Aleshores, podem utilitzar una eina que ens ajudi a transformar els fitxers RDF en XML (amb l'ajuda de les fulles d'estil XSLT[138], per exemple) i a partir d'aquí treballar amb XML per a les bases de dades i la publicació en la Web.

## *6.1 Generar dades RDF*

Com ja s'ha vist en el capítol 3, els fitxers RDF i OWL es basen en la notació de XML. Aleshores, per a crear un fitxer que generi la informació que ens interessa, podem triar entre escriure completament el codi a partir d'un editor de text qualsevol, o utilitzar una interfície que ens ajudi a generar el mateix fitxer de forma visual.

Podem utilitzar un editor com *Altovolta SemanticWorks*[139] per a generar fitxers de dades RDF, construir ontologies en OWL i verificar-ne la sintaxis. *SemanticWorks 2006* permet:

- Crear i editar documents RDF, RDF-Schema (RDFS), OWL-Lite, OWL-DL i OWL-Full utilitzant una interfície de tipus *drag-and-drop*.
- Verificar la sintaxis, de manera que els documents generats siguin conformes a les especificacions de RDF / XML publicades per W3C.
- Generació i edició de "tripletes" RDF / OWL.
- Impressió de les representacions gràfiques dels documents RDF / OWL, per a documentar els projectes de Web Semàntica.

Tot i que *SemanticWorks* és una aplicació molt completa, existeixen en el mercat altres solucions com *Protégé*[140], *RDF Editor*[141], *The Validating RDF Parser* (VRP)[142], *Another RDF Parser* (ARP)[143], *Raptor RDF*[144]. Quasi tots aquests programaris són de lliure accés.

Veurem a continuació un exemple de gestió de RDF amb *SemanticWorks.* Partim d'una estructura ja generada que pertany al domini del món musical, on hi haurà les metadades de les cançons (el títol, el compositor, l'estil del grup de música...) i dels compositors (nom, nacionalitat, estil, etc). Tindrem doncs tres fitxers. El primer defineix l'ontologia: estableix les classes que hi intervenen (artistes i cançons). Els altres dos fitxers contenen les dades de treball que corresponen a cadascuna de les classes.

---

[138] http://www.w3.org/TR/xslt

[139] www.altova.com/products_semanticworks.html

[140] http://protege.stanford.edu

[141] http://www.cs.rpi.edu/~puninj/rdfeditor/

[142] http://139.91.183.30:9090/RDF/VRP/index.html

[143] http://www.hpl.hp.com/personal/jjc/arp/

[144] http://librdf.org/raptor/

Obrim primerament amb *SemanticWorks* el fitxer que conté l'ontologia: "*simac-ontology-0.1.rdf*". Obtenim la representació esquemàtica de les classes que hi intervenen, incloent-ne les propietats de cadascuna, així com el diagrama de classes.

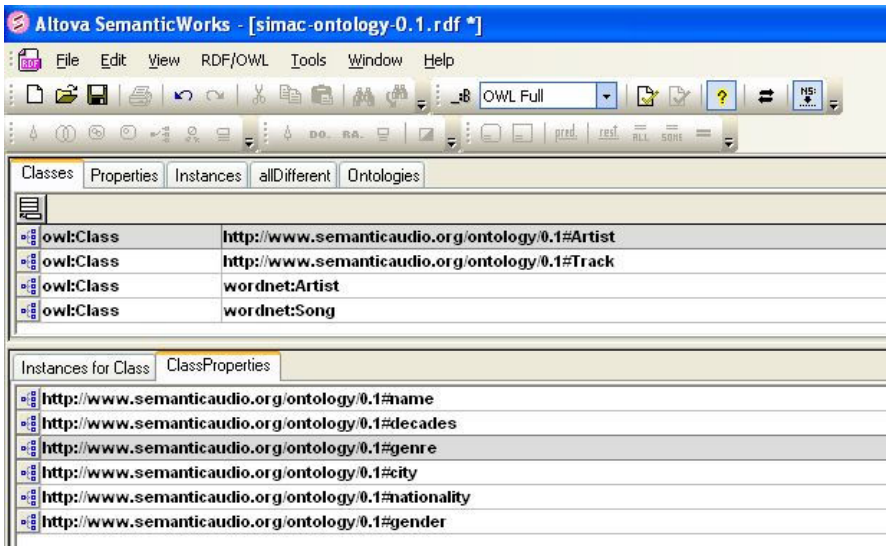



**Figura 6.1 Exemple d'ontologia OWL**

Com ja s'ha comentat, aquest editor ens permet detectar possibles errors en la definició de l'ontologia:

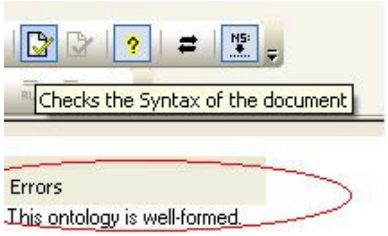



**Figura 6.2 Comprovar la sintaxis d'una ontologia**

D'altra banda, també podem verificar la sintaxis dels fitxers que contenen les dades RDF ("*artists.rdf*", per al cas dels artistes) i comprovar si hi ha cap error:

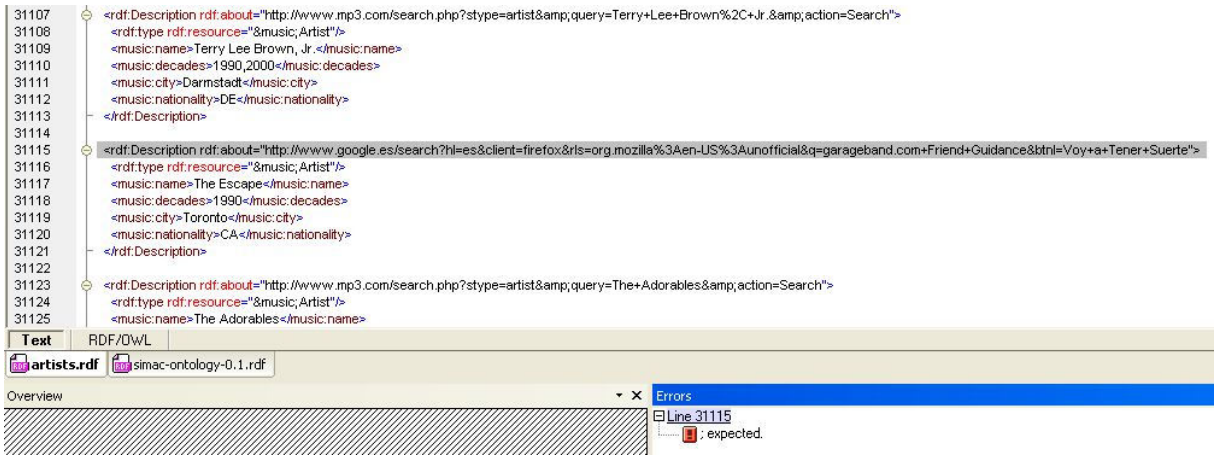



**Figura 6.3 Exemple de fitxer RDF**

*SemanticWorks* també ens permet obtenir l'esquema de l'ontologia. En la figura 6.4 es mostra la definició de la classe "*Artist*".

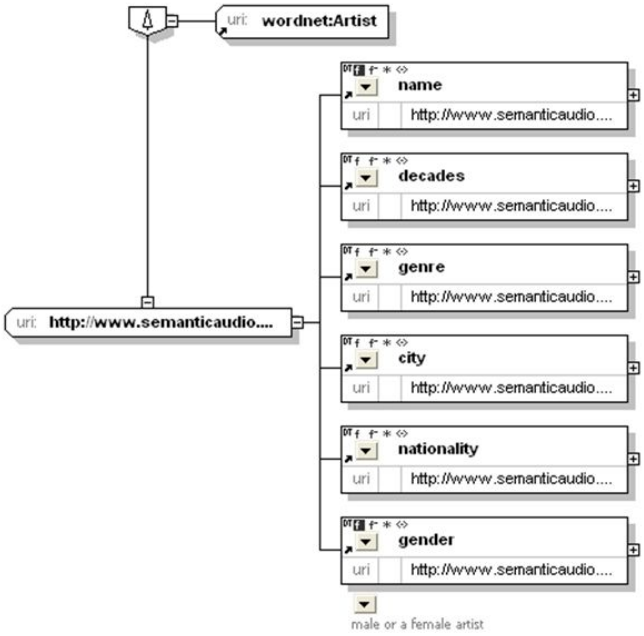



**Figura 6.4 Diagrama de classes dins l'ontologia**

Una vegada estem segurs que les dades que disposem són correctes, ens cal enregistrar aquesta informació en una bases de dades.

## *6.2 Diferents solucions de Web Semàntica*

Es troben en el mercat múltiples solucions i diferents tecnologies per a la implementació de Web Semàntica, algunes d'elles sota llicència i la majoria d'accés lliure. D'una banda existeixen solucions globals que inclouen tots els components necessaris, com *RDF Interface*. També existeixen "*middlewares*" que implementen una part de l'aplicació, basant-se en SGBD ja consolidats; és el cas de *Jena* o *Sesame*. Finalment, es poden desenvolupar solucions personalitzades aprofitant les capacitats de compatibilitat amb XML dels SGBD més importants del mercat.

A continuació veurem el funcionament de cadascun dels tipus de solucions indicats. En el camp dels SGBD, destacarem el paper d'Oracle.

### 6.2.1 RDF Interface

*Intellidimension RDF Gateway (v2.2.4)*[145] es defineix com una plataforma per a la Web Semàntica. *RDF Gateway* és un aplicació client/servidor basada en HTTP amb una base de dades nativa RDF i OWL. Permet crear pàgines Web que consulten la base de dades i publiquen la informació en la Web. És capaç de recuperar la informació dels SGBD habituals i mostrar aquesta informació en la Web com a dades RDF.

[145] http://www.intellidimension.com/

Un cop l'aplicació instal·lada en el sistema, la consola d'administració és accessible mitjançant la URL http://localhost:7056/mgmt/:

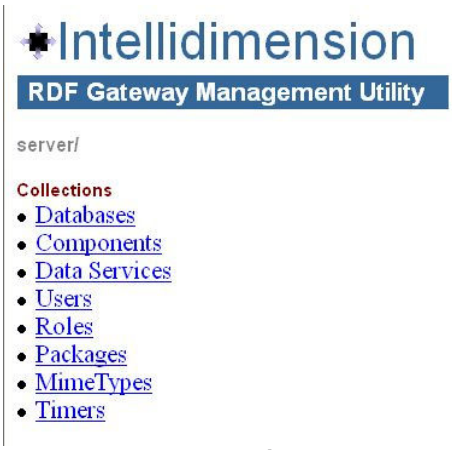



**Figura 6.5 Administració de RDF Interface**

Aquesta interfície ens permet gestionar i administrar les base de dades:

**Figura 6.6 Bases de dades amb RDF Interface**

*Intellidimension InferEd*[146] és un programari que permet l'edició de documents RDF i OWL. Permet a més la inserció de les dades RDF en documents Web.

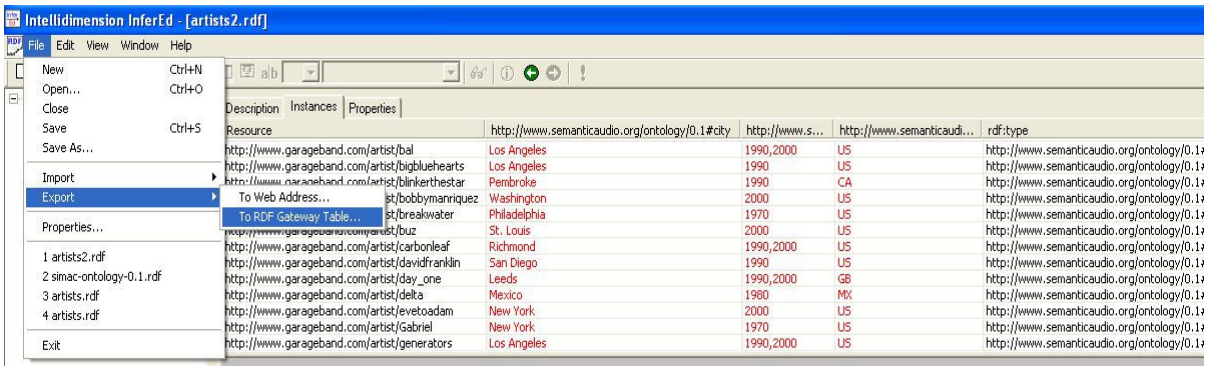



**Figura 6.7 Fitxers RDF amb InferEd**

*RDF Gateway* incorpora una eina per a la consulta de les dades RDF en la base de dades: *RDF Query Analyzer*. El sistema de consultes es basa en RDFQL:

146 http://www.intellidimension.com/default.rsp

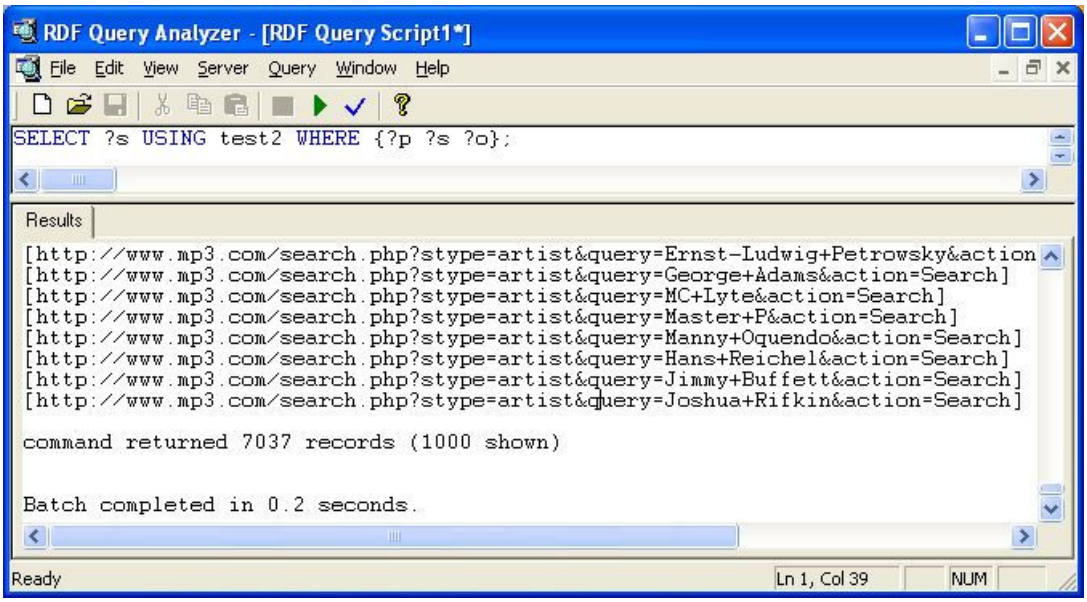



**Figura 6.8 Consultes amb RDF Query Analyzer**

### 6.2.2 Sesame

*Sesame*[147] és una base de dades *open source*, que suporta inferència i consultes sobre esquemes RDF. Fou desenvolupada per *Aduna*[148] (una empresa holandesa que abans s'anomenava *Aidministrator Nederland*).

Aquesta eina treballa sota entorns diferents (Linux, Windows, etc), pot col·laborar amb diversos gestors de bases de dades relacionals, i incorpora llenguatges de consulta variats, com SeRQL. Podem trobar el manual d'usuari al lloc http://www.openrdf.org/doc/sesame/users/.

*Sesame* és una interfície dissenyada amb Java, així que la instal·lació de Java és un requisit per al seu funcionament. A més, cal instal·lar un servidor HTTP del tipus Apache, com *TomCat Apache 5.0*[149]. *Sesame* necessita un SGBD per enregistrar les dades RDF; pot utilitzar PostgreSQL, MySQL, Oracle 9i i SQL Server (recordem que PostgreSQL i MySQL són d'accés lliure). Caldrà instal·lar en el directori de treball de *Sesame* els *drivers JDBC* corresponents al gestor que haguem triat.

Veiem a continuació el procés d'instal·lació i el funcionament de forma simplificada.

[147] http://www.openrdf.org/

[148] http://aduna.biz/index.html

[149] http://tomcat.apache.org/

- Instal·lació de Java 2 versió 1.4 (com a mínim):

**Figura 6.9 Instal·lació de Java 2**

- Instal·lació de TomCat, servidor Apache:

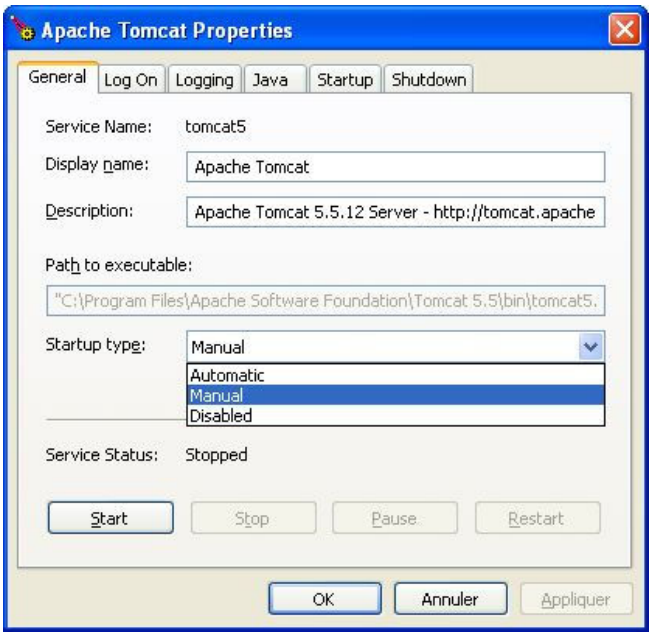



**Figura 6.10 Instal·lació d'un servidor Apache**

- Instal·lació i configuració del servidor TomCat segons el manual de Sesame. Si tot va bé, tindrem accés a Sesame mitjançant la URL: http://localhost:8080/sesame/. La instal·lació del servidor Apache és necessària per a la gestió de les pàgines PHP que publicaran les dades RDF.

- Instal·lació d'un SGBD, per exemple MySQL, d'accés lliure:

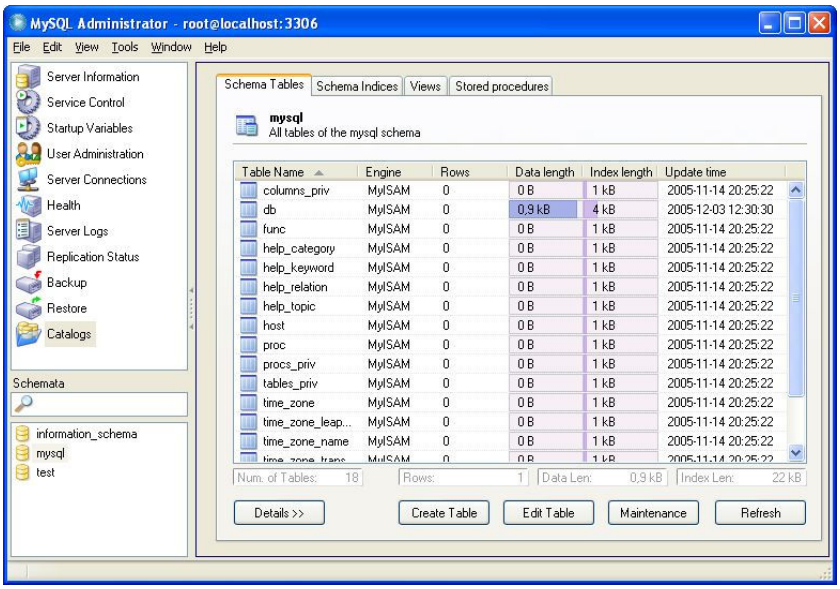



**Figura 6.11 Instal·lació de MySQL**

- Configuració i administració del servidor de Sesame, gestió d'usuaris i parametrització dels accessos als SGBD:

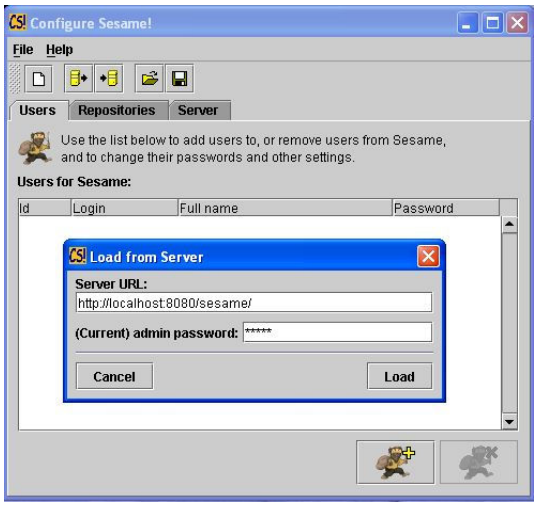



**Figura 6.12 Configuració de Sesame**

- Obrir una sessió i elecció del SGBD:

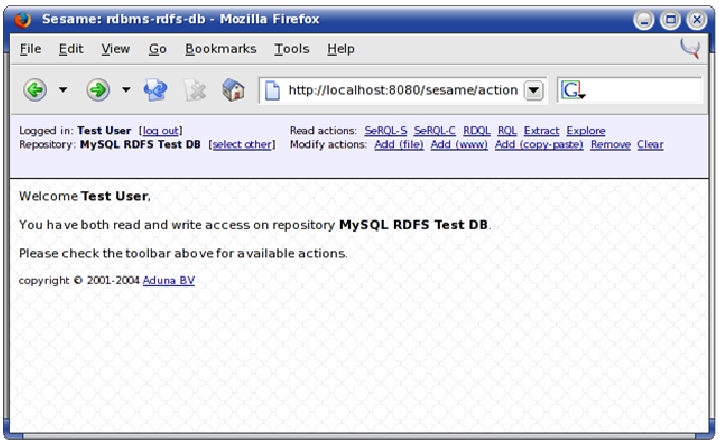



**Figura 6.13 Utilització de Sesame**

- Afegir dades fent click en l'opció "*Add (file)*": permet triar el fitxer RDF que es troba en el disc dur del sistema.

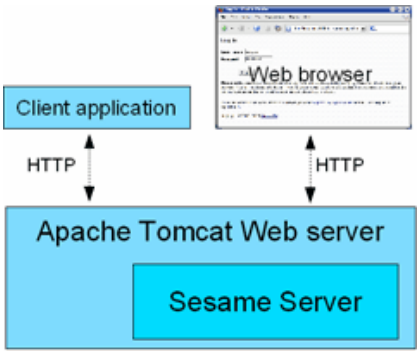



**Figura 6.14 Estructura de Sesame**

- Per a executar consultes sobre les dades RDF, podem utilitzar diferents llenguatges: Sesame suporta SeRQL[150] (*Sesame RDF Query Language*), RQL[151] i RDQL[152] (*RDF Data Query Language*).

### 6.2.3 JENA

Jena[153] és un entorn de treball desenvolupat en Java per a implementar aplicacions de Web Semàntica. Proporciona suport per RDF, RDFS i OWL, incloent un motor d'inferències. Es tracta d'una interfície "*open source*" que va néixer en l'entorn del *HP Labs Semantic Web Research Programme*[154]. Jena inclou una API per RDF, una altra per OWL, capacitat de lectura i escriptura de dades RDF, enregistrament persistent de dades i suport a RDQL.

*Another RDF Parser*[155] és el programari de Jena que permet crear i comprovar la validesa de fitxers de dades RDF. Pel que fa a la col·laboració amb SGBD, Jena suporta MySQL[156], Oracle[157] i PostgreSQL[158] i pot córrer sota Linux i WindowsXP. Per a publicar les dades RDF en la Web, Jena utilitza un servidor propi, anomenat *Joseki*[159]. Aquesta eina, a més de fer les funcions de servidor, proporciona una interfície HTTP per a executar consultes sobre les dades; fa la mateixa funció que un servidor Apache (com TomCat) en el cas de Sesame: conté l'aplicació Web generada a partir de la interfície Java i les dades RDF.

---

150 http://www.openrdf.org/doc/sesame/users/ch06.html

151 http://www.openrdf.org/doc/rql-tutorial.html

152 http://www.w3.org/Submission/2004/SUBM-RDQL-20040109/

153 http://jena.sourceforge.net/

154 http://www.hpl.hp.com/semweb/

155 http://www.hpl.hp.com/personal/jjc/arp/

156 http://jena.sourceforge.net/DB/mysql-howto.html

157 http://jena.sourceforge.net/DB/oracle-howto.html

158 http://jena.sourceforge.net/DB/postgresql-howto.html

159 http://www.joseki.org/

### 6.2.4 RDF Suite

*The ICS-FORTH RDFSuite*[160] és un entorn de desenvolupament d'aplicacions de Web Semàntica *open source* creat pel *FORTH Institute of Computer Science* que inclou: un editor/corrector sintàctic de documents RDF (*The Validating RDF Parser, VRP*)[161], un base de dades que genera representacions relacionals a partir de RDF (*The RDF Schema Specific DataBase, RSSDB*)[162] i una eina per a consultar les dades RDF mitjançant RQL[163]. Es tracta d'un projecte desenvolupat en Java i que treballa sobre PostgreSQL.

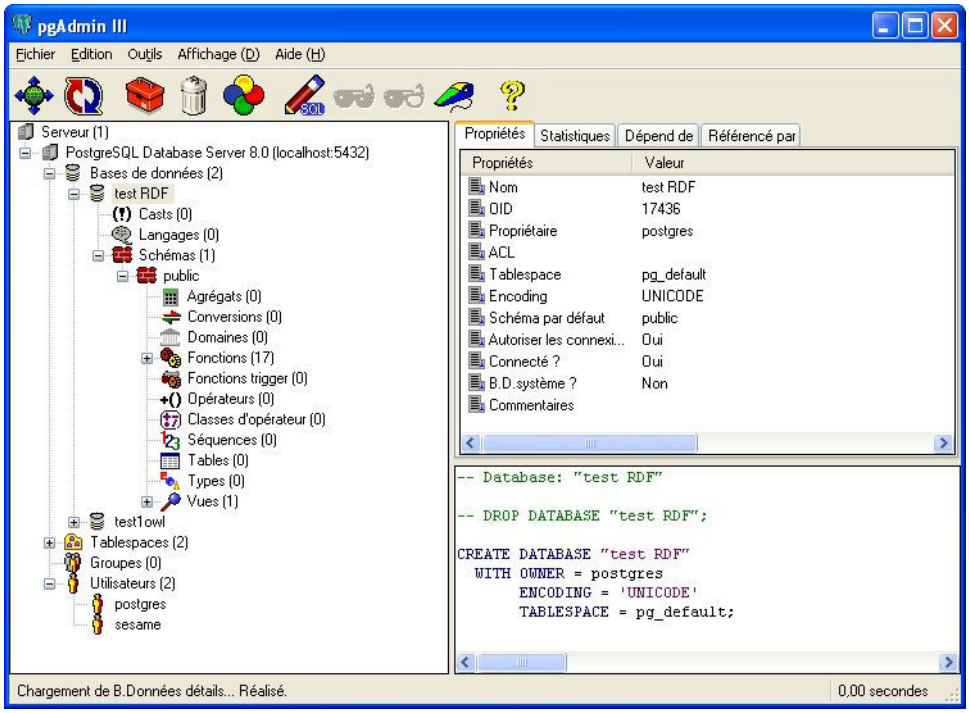



**Figura 6.15 Administració de PostgreSQL**

### 6.2.5 rdfDB

rdfDB[164] és una base de dades *open source* per a RDF. Aquest projecte pretén implementar un SGBD capaç de: enregistrar fitxers de dades RDF, suportar esquemes RDF, capacitat d'inferència, suport de "tripletes" i representació de grafs. Aquest programari s'ha desenvolupat en C, però permet integració amb altres llenguatges com Perl o Java; Eric van der Vlist[165], per exemple, ha dissenyat una interfície en Java per a rdfDB.

---

[160] http://139.91.183.30:9090/RDF/

[161] http://139.91.183.30:9090/RDF/VRP/index.html

[162] http://139.91.183.30:9090/RDF/RSSDB/index.html

[163] http://139.91.183.30:9090/RDF/RQL/index.html

[164] http://www.guha.com/rdfdb/

[165] http://xmlhack.com/read.php?item=785

### 6.2.6 Oracle i RDF

Un dels SGBD que més preocupació mostra per el desenvolupament de la Web Semàntica és el producte de la casa Oracle. Una clara prova d'aquest fet és l'existència d'un Centre de Tecnologies Semàntiques (*Semantic Technologies Center*)[166] dins el lloc web d'Oracle.

Les bases de dades d'Oracle són compatibles amb XML des de la versió 9i[167]. Es pot trobar complerta informació sobre en el lloc "*Using Oracle XML Database to Manage XML Documents*"[168]. Per a més informació sobre Oracle i XML, es pot consultar el lloc "Oracle XML FAQ": http://www.orafaq.com/faqxml.htm.

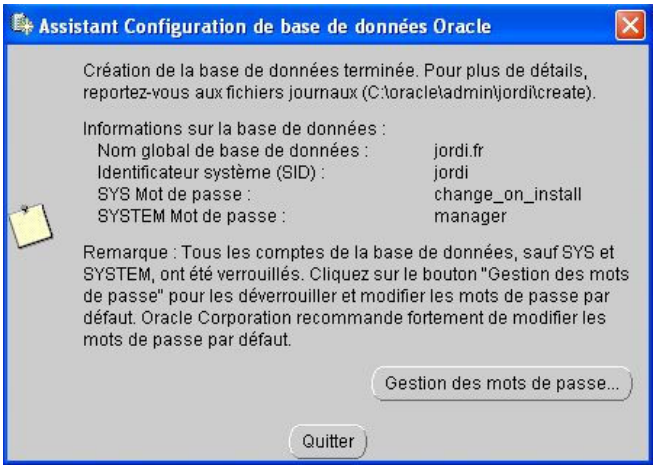



**Figura 6.16 Instal·lació d'Oracle 9i**

Però és sobretot amb l'aparició de *Oracle Database 10g* que es demostra la seva implicació en el concepte de Web Semàntica. *Oracle Spatial 10g Release 2*[169] és una plataforma de gestió de dades per a aplicacions basades en RDF.

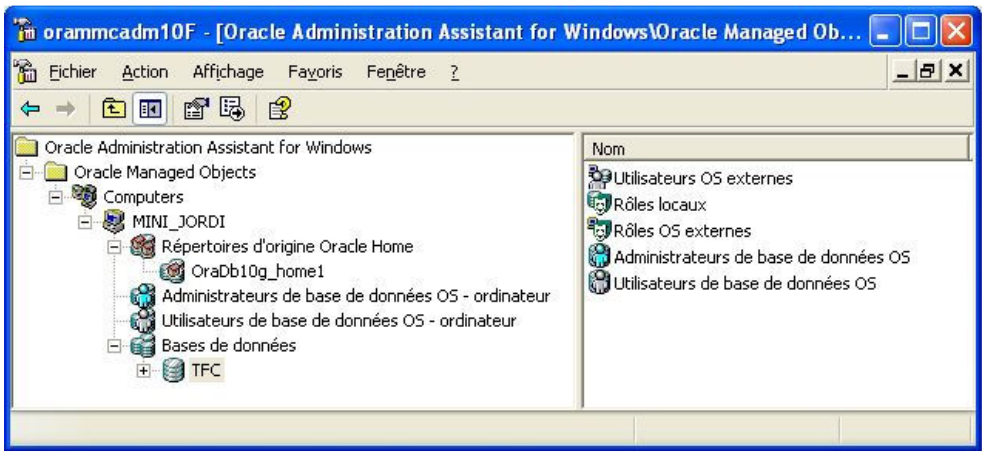



**Figura 6.17 Administració de Oracle 10g**

---

[166] http://www.oracle.com/technology/tech/semantic_technologies/index.html

[167] Per a més informació respecte la instal·lació d'Oracle9i, consultar el lloc: http://www.oracle.com/technology/products/oracle9i/htdocs/9iobe/OBE9i-Public/index.html

[168] http://www.oracle.com/technology/obe/obe9ir2/obe-cnt/xdb/xmldb.htm

[169] http://www.oracle.com/technology/products/spatial/index.html

Per tant, amb Oracle 10g ja no parlem de bases de dades XML, com fèiem amb Oracle 9i, sinó directament de bases de dades RDF. Per a gestionar les dades RDF, Oracle ha creat dos tipus d'objectes: *SDO_RDF_TRIPLE* i *SDO_RDF-TRIPLE_S*. *SDO_RDF_TRIPLE* serveix per a definir la representació de la "tripleta" RDF del tipus <recurs, enunciat, predicat>. *SDO_RDF_TRIPLE_S* enregistra les dades RDF en la base de dades. Aquestes classes inclouen diverses propietats i funcions per a gestionar correctament les dades RDF, com la funció *GET_TRIPLE(),* que retorna la "tripleta" definida en la classe *SDO_RDF_TRIPLE*.

A continuació es detallen els passos a seguir per a treballar amb dades RDF dins una base de dades Oracle:

- Crear un espai de treball: com que els fitxers RDF acostem a ser bastant grans, es aconsellable de crear un nou *TABLESPACE.*
- Crear una xarxa RDF: un cop creat l'espai de treball, cal crear una *Network Data Model* per a enregistrar RDF. Aquesta etapa és necessària per afegir el suport RDF a una base de dades Oracle.
- Crear una taula: a continuació podem ja crear un taula que contindrà les dades RDF. Aquesta taula ha de tenir una columna del tipus *SDO_RDF_TRIPLE_S,* que farà referència a totes les dades corresponents a un mateix model RDF.
- Crear un model RDF: un model RDF es defineix a partir d'un nom de model, de la taula que conté les referències RDF i del nom de la columna dins la taula que conté l'objecte *SDO_RDF_TRIPLE_S.*

Per a inserir les dades RDF en la base de dades, es poden utilitzar dues vies diferents: inserció mitjançant comandes SQL ('*INSERT INTO*'); inserció via Java, amb l'ajuda d'alguna de les interfícies ja vistes (Jena, Sesame, etc). Oracle també suporta RDF-Schema i permet la creació de normes (*rules*) d'inferència.

```
-- Basic steps:
-- After you have connected as a privileged user and called
-- SDO_RDF.CREATE_RDF_NETWORK to add RDF support,
-- connect as a regular database user and do the following.
-- 1. For each desired RDF model, create a table to hold its data.
-- 2. For each RDF model, create an RDF model (SDO_RDF.CREATE_RDF_MODEL).
-- 3. For each table to hold RDF data, insert data into the table.
-- 4. Use various subprograms and consructors.

-- Create the table to hold data for the RDF model.
CREATE TABLE articles_rdf_data (id NUMBER, triple SDO_RDF_TRIPLE_S);

-- Create the RDF model.
EXECUTE SDO_RDF.CREATE_RDF_MODEL('articles', 'articles_rdf_data', 'triple');

-- Information to be stored about some fictitious articles:
-- Article1, titled "All about XYZ" and written by Jane Smith, refers
--   to Article2 and Article3.
-- Article2, titled "A review of ABC" and written by Joe Bloggs,
--   refers to Article3.
-- Seven SQL statements to store the information. In each statement:
-- Each article is referred to by its complete URI The URIs in
--   this example are fictitious, although they are in the general
--   domain of the journal Nature (http://www.nature.com/nature/).
-- Each property is referred to by the URL for its definition, as
--   created by the Dublin Core Metadata Initiative.

-- Insert rows into the table.

-- Article1 has the title "All about XYZ".
INSERT INTO articles_rdf_data VALUES (1,
  sdo_rdf_triple_s ('articles','http://www.nature.com/nature/Article1',
    'http://purl.org/dc/elements/1.1/title','All about XYZ'));
```

Font: User Guide d'Oracle Spatial 10g Release 2

**Figura 6.18 Exemple d'inserció de dades RDF per Oracle 10g**

Oracle proporciona altres eines per al tractament de XML:

- XMLDB[170]: és la tecnologia de bases de dades inclosa en *Oracle9i Database Release 2* per a proporcionar un suport XML integrat.
- XSU (*XML-SQL Utility*)[171]: eina que proporciona intercanvi d'informació, permet transformació dades relacionals en XML o extreure dades "estructurables" en files i columnes a partir d'un document web XML.
- XML Developer's kits[172]: conjunt d'eines que integren Oracle amb tot un ventall d'utilitats excrites en Java, C, Perl, etc, entorn d'aplicacions XML.
- XSQL Servlet[173]: Convina SQL, XML, XSLT i el protocol HTTP per a obtenir plantilles de pàgines web rebre informació XML d'una base de dades i llençar consultes sobre dades XML des de la Web.

## *6.3 Publicació en la Web*

El sistema a utilitzar per a publicar dades RDF en la Web dependrà forçosament del tipus de tecnologia que s'hagi seguit per l'implementació de Web Semàntica.

Quan parlem de solucions globals o solucions natives RDF, ja hem vist que la majoria incorporen eines del tipus servidor Apache o similar.

Pel que fa a implementacions més o menys personificades basades en SGBD (MyQSL, posteSQL, Oracle), cal trobar una solució per a la generació de pàgines Web. Actualment el llenguatge web més utilitzat és PHP. Per a la publicació de pàgines PHP cal instal·lar un servidor Apache. En aquest sentit es pot optar per una solució del tipus EasyPHP[174]. Aquest programari instal·la i configura de forma automàtica servidor Apache per a la gestió de pàgines PHP i una base de dades MySQL, en un entorn de treball sota Windows. Podem trobar un exemple d'implementació de RDF amb PHP en el lloc:
http://xulfr.org/wiki/ApplisWeb/ExemplePhpRdf.

Pel que fa al cas particular d'Oracle, existeix un producte anomenat *Oracle iFS*[175], que permet connexions HTTP, FTP o inclús SMTP a bases de dades Oracle.

---

[170] http://www.oracle.com/technology/tech/xml/xmldb/9.2.0.2.0/index.html

[171] http://www.oracle.com/technology/tech/xml/xdk/doc/beta/doc/java/xsu/xsu_userguide.html

[172] http://www.oracle.com/technology/tech/xml/xdkhome.html

[173] http://www.oracle.com/technology/tech/xml/xdk/doc/beta/doc/java/xsql/xsql_userguide.html

[174] http://www.easyphp.org/

[175] http://www.orafaq.net/faqifs.htm

# 7. Valoració Econòmica

Com ja s'ha vist en el capítol anterior, es necessiten diferents elements per a l'implementació d'una aplicació web semàntica: un editor per a generar documents RDF, un sistema gestor de bases de dades per a enregistrar la informació estructurada, un sistema per a la publicació de les dades en format web.

Es poden trobar alguns "parsers" RDF gratuïts, però acostumen a ser editors força limitats, que no sempre poden donar resposta a les necessitats de l'usuari que decideix estructurar els seus documents. Altrament, existeix programaris molt complerts (a vegades fins i tot massa complerts) com *Altova XMLSpy 2006*, que tenen l'inconvenient d'haver de pagar una llicència (al voltant dels 600 €) per a poder-los utilitzar (tot i que es pot demanar una llicència de proves de manera gratuïta i amb una validesa de 30 dies).

Pel que fa als sistemes de bases de dades relacionals, ja hem vist en el capítol 5 que n'hi ha de lliure accés i de comercials (es necessita comprar una llicència). Segons les necessitats o requeriments del negoci de l'usuari caldrà decidir-se entre un o altre sistema, però res no justifica adquirir un sistema com Oracle (entre 300 € i 4.000 € segons la versió i el tipus de llicència) si es tracta simplement de treballar amb un SGBD específicament per a emmagatzemar dades RDF (es pot descarregar un gestor del tipus *open source* com MySql o PostgreSql).

Per últim, per a la publicació de les dades en la Web, cal un servidor HTTP amb capacitat d'accés al sistema gestor de bases de dades que haurem seleccionat. En el cas (més que probable) d'instal·lar un gestor de llicència lliure, s'acostuma a treballar amb un servidor Apache[176] i amb llenguatge PHP per a la generació de les pàgines web. En aquest sentit, tenim productes *open source* com Apache TomCat[177] o el paquet EasyPHP[178], que incorpora un servidor HTTP Apache i el gestor de bases de dades MySql.

[176] http://www.apache.org/

[177] http://tomcat.apache.org/

[178] http://www.easyphp.org/

# 8. Conclusió

La Web s'ha convertit en l'eina de treball i de cerca de molta gent, tal és la quantitat d'informació que conté i els serveis que s'han creat entorn d'ella. Però la cerca de determinades informacions en la Web actual demanen dedicar-hi un temps excessiu i poden inclús provocar frustració en l'usuari davant la incapacitat de trobar exactament la dada desitjada. Es amb la voluntat d'optimitzar la recerca d'informació que apareix el concepte de Web Semàntica.

Amb l'implementació de la Web Semàntica, els ordinadors seran capaços d'interpretar la informació expressada en un llenguatge humà: la màquina serà capaç de diferenciar el mot "demà" del concepte "avui", i prenent com a partida la data del sistema podrà inferir que avui és dilluns i, per tant, demà serà dimarts. Aquest fet és molt important perquè permetrà d'utilitzar expressions en els motors de cerca del tipus "vols per a Praga la setmana que ve", obtenint exactament els resultats esperats, sense ambigüitat, i sense necessitat d'especificar quin dia serà la setmana que ve ni quin és el punt de partida.

Per a que això sigui possible, s'han de complir el següent: que els documents publicats a la Web estiguin estructurats (amb l'ús de metadades) per a facilitar a les màquines la interpretació de les dades; l'establiment d'ontologies que defineixin cada concepte, les seves característiques i les relacions entre conceptes de manera inequívoca: si prenem per exemple el concepte "banc", ens cal una eina capaç de definir en quins àmbits serà vàlid parlar d'un "seient estret i llarg", i quan caldrà interpretar-ho com un "entitat on guardar els estalvis".

Com acostuma a passar quan apareix un nou concepte, sorgeixen diferents propostes de solucions per a la Web Semàntica, i el problema és que els diferents desenvolupadors i fabricants es posin d'acord per a establir un únic estàndard. Aquest és el paper del W3C (*World Wide Web Consortium*), que a definit un estàndard per a l'estructuració de les dades, la notació RDF, basada en la sintaxis XML. Igualment, per a la descripció d'ontologies, el model proposat s'anomena OWL, que estén les característiques de RDF per a afegir capacitat d'inferència. Les dades expressades en XML / RDF incorporen en paral·lel una avantatge addicional: es podrà compartir informació resident en sistemes diferents i provenint de formats no necessàriament compatibles; deixarà de tenir sentit el fet que una empresa utilitzi un sistema específic per a la comptabilitat i un altre per a les dades dels clients.

La majoria de les dades publicades en la Web estan emmagatzemades en bases de dades. Per a que la Web Semàntica sigui possible, es necessiten sistemes gestors de bases de dades capaços de gestionar dades en format RDF i OWL. En l'estat actual, la majoria de SGBD permeten treballar amb dades XML, però no encara amb notacions RDF directament. No obstant, es poden utilitzar diferents solucions *middlewares* (com Jena o Sesame) per a importar / exportar dades RDF en els SGBD més importants del mercat.

Una solució alternativa per al tractament de dades estructurades consisteix a utilitzar bases de dades natives RDF. Això imposa l'inconvenient de treballar amb SGBD diferents segons les necessitats de cada projecte, enlloc de centralitzar tota la informació en un sol gestor de bases de dades.

Oracle ha esdevingut el SGBD capdavanter en tecnologies semàntiques, incorporant compatibilitat amb dades RDF en la seva versió Oracle 10g Release 2. A més, està preparant la versió 11x Database, que inclourà eines per al tractament d'ontologies OWL. És clar que seria encara més interessant per a la major integració de la Web Semàntica, que gestors *open source* com MySql o PostgreSql s'incorporin en la carrera encapçalada per Oracle.

Independentment de la compatibilitat de RDF amb els diferents SGBD, és innegable que la semàntica comença a arrelar en l'entorn de la Web; cada cop trobem més aplicacions que recuperen els conceptes de RDF en la Web actual. El cas més estès és el de RSS (o la sindicació de continguts), una utilitat que permet descriure els recursos web de manera que es puguin compartir sense necessitat de consultar sistemàticament les URL's on es publiquen aquests recursos. Una altra aplicació és FOAF, xarxa semàntica de pàgines web que descriuen persones, els vincles entre elles, i coses que fan o creen. Una de les últimes novetats en la Web és l'anomenat *Podcasting*, que de fet no és més que una extensió de RSS en el camp multimèdia, permetent la sindicació de fitxers d'àudio relacionats amb transmissions de diferents emissores de ràdio.

De cara al futur, la incorporació de semàntica a la Web implicarà de manera immediata l'optimització de les cerques d'informació, i la possibilitat d'automatització de tasques repetitives o massa complicades per l'usuari, delegant la responsabilitat d'aquests processos a agents semàntics capaços de trobar la millor solució treballant en col·laboració amb d'altres màquines i/o agents, i tot plegat a partir d'un nombre mínim de dades de partida definides per l'usuari i aprofitant el seu perfil (que s'haurà generat a partir de cerques anteriors). D'aquesta manera s'arribarà a un punt on serà la pròpia màquina qui proposarà la informació a la persona; per exemple, serà capaç de decidir si l'últim gran èxit cinematogràfic serà de l'interès de l'usuari o si preferirà, en canvi, anar a sopar al restaurant vegetarià que s'acaba d'incorporar en la guia electrònica de restaurants de la nostra ciutat.

# Glossari

**AI** (*Artificial Intelligence*). Intel·ligència Artificial.

**CERN** (*Centre Européen pour la Recherche Nucléaire*)
El Centre Europeu de Recerca Nuclear, seu de la primera conferència de la World Wide Web i considerat com el lloc de naixement de la tecnologia WWW. Els treballs de recerca sobre la tecnologia WWW i l'establiment d'estàndards s'han traspassat a la World Wide Web Organization (W3O, a w3.org).

**HTML** (*Hypertext Markup Language*)
Llenguatge de marcatge d'hipertext. Llenguatge de marques (hipertext) amb el qual es dóna format a les pàgines web i es distribueix la informació de la Web.

**HTTP** *(Hyper Text Transfer Protocol)*
Protocol de Transferència d'HiperText. El mètode mitjançant el qual els documents es transfereixen de l'ordinador del sistema principal o del servidor als exploradors i usuaris individuals.

**RDF** (*Resource Description Framework*)
Model per a la representació simplificada de metadades, basat en "tripletes" recurs – propietat – entitat.

**Terabyte** *(TB)*
Unitat de mesura d'informació. 1 TB = $10^{12}$ Bytes. 1 Byte = 8 bits = 1 caràcter.

**UML** (*Unified Modeling Language*)
Llenguatge de propòsit general per a la construcció de software orientat a objectes.

**URI** (*Uniform Resource Identifiers*)
Identificador Unificat de Recursos. Text curt que identifica unívocament qualsevol recurs web. Consta de dues parts: protocol d'accés al recurs, nom del recurs.

**URL** (*Uniform Resource Locator*)
Localitzador Uniforme de Recursos. Cadena de caràcters amb la qual s'assigna una adreça única a cadascun dels recursos d'informació disponibles a Internet. El format general d'una URL és: "protocol://màquina/directori/fitxer".

**SW** (Semantic Web). Web Semàntica.

**XML** (*eXtensive Markup Language*)
Llenguatje de Marques Ampliable. Una forma flexible de crear formats d'informació i compartir tant el format com les dades en la Web.

**W3C** (*World Wide Web Consortium*). Consorci World Wide Web.

**WS** (*Web Services*). Serveis Web.

**WWW** (*World Wide Web*)
La Web. Sistema d'hipertext que funciona sobre Internet per a l'enllaç mundial de documents multimèdia que fa que la part d'informació que és comú entre dos documents tingui un fàcil accés, i que sigui completament independent de l'emplaçament físic.

# Bibliografia[179]

[Abrams 1998] M. Abrams, ed., *World Wide Web - Beyond the Basics*, Prentice Hall, 1998. http://ei.cs.vt.edu/~wwwbtb/book/.

[Berners-Lee 1989] T. Berners-Lee, "Information Management: A Proposal. Internal Project Proposal", CERN, 1989. http://www.w3.org/History/1989/proposal.html.

[Berners-Lee 2001] T. Berners-Lee, J. Hendler, O Lassila, "The Semantic Web", *Scientific American*, Maig 2001. http://www.scientificamerican.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&catID=2

[Cámara 2002] Juan Carlos Cámara, "Learning Metadata Standards", UPF, Barcelona, 2002. http://www.iua.upf.es/~jblat/material/doctorat/students/jccbis/Sumario.html

[Castells 2003] P. Castells, "La web semántica", *in* C. Bravo, M. A. Redondo (Eds.). *Sistemas Interactivos y Colaborativos en la Web*, Ediciones de la Universidad de Castilla - La Mancha, 2003, pp. 195-212. http://www.ii.uam.es/~castells/publications/castells-uclm03.pdf

[Connolly 2000] D. Connolly, "A Little History of the World Wide Web", *W3C Consortium*, 2000. http://www.w3.org/History.html.

[D.Ll.C. 1975] *Diccionari de la Llengua Catalana*, Barcelona, Albertí Editor, 1975.

[Gruber 1993] T. R. Gruber, "A Translation Approach to Portable Ontology Specifications". *Knowledge Acquisition*, 5(2), pp. 199-220, 1993.

[Labra 2005] Jose Emilio Labra Gayo, "Tecnologías XML y Web Semántica", Escuela Universitaria de Ingeniería Técnica Informática de Oviedo, Juliol 2005. http://di002.edv.uniovi.es/~labra/cursos/ver05/

[Lyman 2003] Peter Lyman & Hal R. Varian, "How Much Information? 2003", *SIMS, School of Information, Management and Systems.* Universitat de California, Berkeley, Octubre 2003. http://www.sims.berkeley.edu/research/projects/how-much-info-2003/execsum.htm#summary

[O'Neill 2003] E. T. O'Neill, B. F. Lavoie, R. Bennett, "Trends in the Evolution of the Public Web". *D-Lib Magazine*, Volume 9 Number 4, Abril 2003. http://www.dlib.org/dlib/april03/lavoie/04lavoie.html

[Provost 2002] W. Provost, "UML For W3C XML Schema Design", *XML.com*. Agost 2002. http://www.xml.com/pub/a/2002/08/07/wxs_uml.html

[Weigand 1997] H. Weigand, "Multilingual Ontology-Based Lexicon for News Filtering – The TREVI Project", *K. Mahesh*, 1997.

[179] En els casos aplicables, s'adjunta la URL de la versió en línia.

# Llocs Web

http://dublincore.org/documents/dces/

http://ei.cs.vt.edu

http://lg.msn.com/intl/ca/tutorial/glossary.htm

http://simile.mit.edu

http://www.dlib.org

http://www.google.es

http://www.google.fr

http://www.ii.uam.es/~castells/publications/

http://www.isoco.com

http://www.kartoo.com

http://www.openrdf.org

http://www.oracle.com

http://www.rpbourret.com/xml/ProdsNative.htm

http://www.schemaweb.info

http://www.scientificamerican.com

http://www.sedic.es

http://www.semanticweb.org

http://www.sidar.org

http://www.sims.berkeley.edu/research/projects/how-much-info/

http://www.w3.org

http://www.w3c.es

http://www.wsindex.org

http://www.xml.com/semweb/

# ANNEX 1: El problema de la Web actualment

La cerca de continguts mitjançant mots clau, no sempre ens aporta el resultat adequat d'una manera directa. A vegades això implica certes tasques esdevinguin molt complexes o gairebé impossibles.

Exemple de cerca del grup musical "Ketchup" amb el cercador *Google*:

CONSUMER.es EROSKI ::: Salsa **ketchup**
El diario del consumidor: Noticias diarias sobre consumo y vida cotidiana, reportajes especiales, Canal de Bricolaje, de Economía y de Nutrición, **...**
www.consumer.es/web/es/alimentacion/aprender_a_comer_bien/ alimentos_a_debate/ 2001/09/18/35369.php - 42k - En caché - Páginas similares

Donostia-Gastronomía
**KETCHUP** Ingredientes (para 1 l. y 1/2) 5 kg. y 1/2 de tomates muy maduros. 2 cebollas. 1/2 cucharadita de cayena en polvo. 4 dientes de ajo. **...**
www.paisvasco.net/recetas/**ketchup**.html - 5k - En caché - Páginas similares

Las **Ketchup**
Las noches del verano suenan al ritmo del 'Aserejé', el primer single de estas tres hermanas cordobesas que presumen de ser Las hijas del Tomate, **...**
www.hola.com/2002/07/08/im-las**ketchup**/ - 30k - 2 Dic 2005 - En caché - Páginas similares

Las **Ketchup**
APRENDE LA LETRA DEL ASEREJÉ'Hijas del tomate'Contiene: 'Asereje', 'Kusha Las Payas', 'Un de vez en.
www.hola.com/2002/07/10/de-**ketchup**/ - 22k - En caché - Páginas similares

Melodías Las **Ketchup**, Logos Las **Ketchup**, Melodías Polifónicas Las **...**
Encontradas 1 Melodías Las **Ketchup** Nokia **...** Encontradas 2 Melodías Polifónicas Las **Ketchup** Nokia **...** Encontrados 2 Logos Las **Ketchup** Nokia **...**
www.mrtones.es/busca/las_**ketchup**/Artista/ - 23k - 2 Dic 2005 - En caché - Páginas similares

Calvé **ketchup** bocabajo : comparativas, precios y compras online
Lee opiniones de consumidores y compara precios de Calvé **ketchup** bocabajo.
www.ciao.es/Calve_**ketchup**_bocabajo__145146 - 28k - En caché - Páginas similares

IBLNEWS : 'W **Ketchup**' para los repúblicanos, Heinz para los demócratas
"No apoyas a los demócratas ¿Por qué debes hacerlo con su **ketchup**? **...** El **ketchup** Heinz se le 'atraganta' a muchos republicanos desde que heredó el imperio **...**
iblnews.com/noticias/07/111334.html - 47k - 3 Dic 2005 - En caché - Páginas similares

**Ketchup** - Alimentos TRE, SA - **Ketchup** en todos nuestros formatos
**Ketchup** Fabricante de salsas y encurtidos of Spain. Nuestros productos: **ketchup**, mayonesa, mostaza, salsas gourmet, pepinillos, cebollitas, alcaparras, **...**
www.tresa.es/tre/spain/**ketchup**.htm - 15k - En caché - Páginas similares

KETCHUP
Aqui encontrarás letras de canciones españolas,un poco de todo.Entra y le echas un vistazo por si está la que buscas.
usuarios.lycos.es/atlanti6/newpage52.html - 136k - En caché - Páginas similares

Las **Ketchup** regresan a lo grande · Música · Terra
Canal de Música de Terra. Noticias, reportajes, entrevistas, agenda de conciertos y concursos. Descarga de canciones, álbumes y melodías para el móvil. ,Las **...**
musica.terra.es/articulo/html/mus2172.htm - 22k - 2 Dic 2005 - En caché - Páginas similares

# ANNEX 2: Projectes i eines relacionats amb la SW

Es presenten a continuació alguns dels projectes i eines més representatius generats al voltant de la Web Semàntica:

Aniceto[180]

Projecte de la Universitat Autònoma de Madrid. Aplicació de tecnologies de la Web Semàntica a la gestió d'informació financera i econòmica. Defineix una ontologia per el domini financer a partir de l'editor Protégé[181]. La informació s'enregistra en una base de dades MySQL, amb l'ajuda de la interfície Jena. Les consultes es fan via RDQL.

Annotea[182]

Projecte de W3C gestionat pel SWAD (*Semantic Web Advanced Development*) que pretén establir un sistema d' anotacions (metadades) en RDF sobre qualsevol contingut web. El primer resultat d'aquest projecte és la creació d'un editor i cercador anomenat Amaya[183], que inclou una interfície per a generar metadades en RDF.

Chump[184]

IRC que enregistra totes les discussions en format XML, de manera que es pot consultar fàcilment tot l'historial de diàlegs anteriors. Inclou un canal anomenat "*Semantic Web Interest Group IRC Scratchpad*"[185].

DOAP[186]

Descripció d'un Projecte (*Description Of A Project*). Projecte per a crear un vocabulari RDF que descrigui projectes de software lliure. Pretén proporcionar eines que suportin un projecte en els llenguatges de programació més populars. Doap-a-matic[187] és un producte resultant del projecte DOAP: una plantilla web que serveix per a generar descripcions RDF d'un projecte.

ESPERONTO[188]

Projecte liderat pel *European Information Society Technologies* que intenta establir un pont entre la Web actual i la Web Semàntica. Intenta unificar ontologies de continguts i llengües heterogènies.

---

[180] http://nets.ii.uam.es/aniceto

[181] http://protege.stanford.edu/protege-rdf/protege-rdf.html

[182] http://www.w3.org/2001/Annotea/

[183] http://www.w3.org/Amaya/

[184] http://rdfig.xmlhack.com/

[185] http://rdfig.xmlhack.com/

[186] http://usefulinc.com/doap/

[187] http://doapy.bonjourlesmouettes.org/doap-a-matic

[188] http://www.esperonto.net/semanticportal

FOAF[189]

Amic d'un amic (*Friend of a Friend*). FOAF és un vocabulari RDF que serveix per a descriure persones i les relacions entre elles. Foaf-a-matic[190] és una plantilla web que permet crear un fitxer RDF-FOAF. Foaf Explorer[191] és un cercador de dades FOAF.

Jena[192]

Interfície escrita en Java per Hewlett Packard, per al desenvolupament d'aplicacions de Web Semàntica. Suporta RDF, RDFS i OWL.

KAON[193]

Interfície escrita en Java per al desenvolupament d'ontologies. Producte fruit del projecte *WonderWeb*[194], que té per missió crear la infraestructura ontològica per a la Web Semàntica, com l'aprovació del protocol OWL.

Kowari[195]

Base de dades *Open Source* escrita en Java, construïda expressament per a enregistrar i consultar metadades. Es una base de dades nativa de RDF i OWL.

OntoWeb[196]

Portal web d'intercanvi d'informació basada en les ontologies per a la gestió del coneixement i el comerç electrònic.

Redland[197]

Conjunt de programaris lliures que donen suport a RDF. Una d'aquestes interfícies és Raptor RDF, escrita en C, permet verificar la sintaxis de fitxers RDF i generar-ne les "tripletes" relacionades.

SIMILE[198]

*Semantic Interoperability of Metadata and Information in unLike Environments*. Un projecte que compte amb la col·laboració del W3C que pretén optimitzar la inter-operabilitat entre vocabularis/ontologies, metadades, i serveis.

SESAME[199]

Interfície de codi lliure en Java que permet generar, enregistrar i consultar documents RDF i RDF-Schema. Es pot utilitzar com a base de dades nativa de RDF.

---

189 http://f14web.com.ar/inkel/2003/01/27/foaf.html

190 http://www.ldodds.com/foaf/foaf-a-matic.es.html

191 http://www.foafnaut.org/

192 http://jena.sourceforge.net/

193 http://kaon.semanticweb.org/

194 http://wonderweb.semanticweb.org/

195 http://www.kowari.org/

196 http://www.ontoweb.org/

197 http://librdf.org/

198 http://simile.mit.edu/

199 http://www.openrdf.org/doc/sesame/users/