

Estudi dels llenguatges de consulta per a documents RDF

Estudiant: **Jordi López Ramot**

Consultor: **Òscar Celma Herrada**

Titulació: **Enginyeria Informàtica**

Gener de 2005

Agraïments

Agraeixo el suport de Òscar Celma Herrada per haver estat a l'altra banda del correu electrònic i atendre totes les consultes que he realitzat durant la redacció del treball.

Resum

La Web Semàntica es pot concebre com una extensió de la web actual on la cerca de continguts es realitza de forma molt més efectiva que a l'actualitat.

L'aspecte més rellevant de la Web Semàntica consisteix en que els continguts disposen d'informació descriptiva associada. L'existència d'aquesta informació descriptiva possibilita que els usuaris de la web cerquin i trobin de forma efectiva els recursos web del seu interès, siguin pàgines, imatges, àudio, vídeo o qualsevol altre tipus de recurs.

Dins el context de Web Semàntica, la informació descriptiva associada als recursos s'expressa seguint les recomanacions de l'estàndard RDF (*Resource Description Framework*). Donat el gran nombre de recursos existents a la web, la informació RDF associada a aquests s'emmagatzema en sistemes gestors de bases de dades, que possibiliten una gestió efectiva i eficient d'aquest gran volum d'informació.

Aquest document presenta un estudi de les diferents tecnologies i estàndards sobre els que es fonamenta la Web Semàntica, com son el llenguatge XML (*eXtended Markup Language*), el llenguatge RDF, els diferents sistemes gestors de bases de dades que permeten emmagatzemar informació RDF, així com els diversos llenguatges de consulta d'informació RDF existents.

Com a materialització pràctica dels conceptes descrits en aquest treball i com a part del projecte, s'ha desenvolupat una aplicació web que permet cercar cançons, artistes i editors de weblogs (*bloggers*). L'aplicació web permet realitzar consultes en un dels llenguatges estudiats (*SPARQL*) sobre les bases de dades RDF corresponents als tres tipus de recursos web esmentats.

1 Índex de continguts

1	Índex de continguts	2
2	Introducció.....	2
2.1	Justificació del treball i context en el qual es desenvolupa.....	2
2.2	Objectius del treball	2
2.3	Enfocament i mètode seguit.	2
2.4	Planificació del projecte	2
2.4.1	Continguts a desenvolupar	2
2.4.1.1	Memòria	2
2.4.1.2	Presentació	2
2.4.1.3	Aplicació Web	2
2.4.2	Anàlisi de Riscos	2
2.4.2.1	Tasques no representatives.....	2
2.4.2.2	Imprevistos laborals	2
2.4.2.3	Tasques massa generals	2
2.4.3	Organització del Treball.....	2
2.4.3.1	Disponibilitat.....	2
2.4.3.2	Relació de tasques.....	2
2.4.3.2.1	Redacció de la memòria	2
2.4.3.2.2	Implementació de l'aplicació web per realitzar cerques RDF.	2
2.4.3.2.3	Revisió i correcció de la memòria	2
2.4.3.2.4	Redacció de la presentació.....	2
2.4.4	Calendari de Treball	2
2.4.5	Fites principals.....	2
2.5	Productes Obtinguts	2
2.6	Descripció de capítols.....	2
3	La Web Semàntica.....	2
3.1	El problema amb la Web actual	2
3.2	La Web Semàntica.....	2
4	Introducció a XML	2
4.1	Orígens del llenguatge XML	2
4.2	Definició de XML.....	2
4.3	Regles sintàctiques. Documents ben formats.....	2
4.3.1	Components bàsics XML: Etiquetes, Elements i atributs.	2
4.3.2	Capçalera XML.....	2

4.3.3	Element arrel únic.....	2
4.3.4	Cal especificar les etiquetes de final d'element.....	2
4.3.5	Els elements no es poden superposar	2
4.3.6	La sintaxi XML distingeix entre majúscules i minúscules	2
4.3.7	El valor dels atributs ha d'anar entre cometes.....	2
4.4	Regles gramaticals. Documents vàlids.	2
4.4.1	Documents XML Schema i DTD.....	2
4.5	Aspectes semàntics dels documents XML	2
5	Introducció a RDF	2
5.1	Model RDF	2
5.2	Sintaxi RDF/XML	2
6	Característiques de les consultes RDF	2
6.1	Definició de consulta RDF	2
6.2	Propietats dels llenguatges de consulta RDF	2
6.2.1	Inferència (Entailment).....	2
6.2.2	Model Abstracte RDF	2
6.2.3	Tancament.....	2
6.2.4	Ortogonalitat	2
6.2.5	Seguretat	2
6.2.6	Expressions de camí (Path Expressions)	2
6.2.7	Expressions de camí opcional (Optional Path Expressions)	2
6.2.8	Selecció	2
6.2.9	Projecció	2
6.2.10	Producte	2
6.2.11	Unió	2
6.2.12	Diferència.....	2
6.2.13	Quantificació universal (Universal Quantification)	2
6.2.14	Agregació.....	2
6.2.15	Agrupació (Grouping)	2
6.2.16	Recursivitat	2
6.2.17	Materialització (Reification).....	2
6.2.18	Col·leccions	2
6.2.19	Espais de Noms (Namespaces)	2
6.2.20	Idioma (Language).....	2
6.2.21	Comparacions Lèxiques	2
6.2.22	Comparacions Valuades.....	2
6.2.23	Tipus de dades XML Schema.....	2
6.2.24	Operadors sobre tipus XMLSchema.....	2

6.2.25	Ordenació	2
6.2.26	Consulta simultània a múltiples models RDF	2
6.2.27	Sintaxi similar a SQL (Structured Query Language).....	2
7	Estudi de llenguatges de consulta per documents RDF	2
7.1	RQL.....	2
7.2	SeRQL	2
7.3	TRIPLE	2
7.4	RDQL	2
7.5	N3	2
7.6	Versa.....	2
7.7	RDFQL.....	2
7.8	RXPath	2
7.9	SPARQL	2
7.10	Taula comparativa	2
8	Estudi de SGBDs que treballen amb informació RDF	2
8.1	RQL.....	2
8.2	Sesame.....	2
8.3	rdfDB.....	2
8.4	Jena	2
9	Aplicació Web per cercar documents RDF	2
9.1	Objectiu.....	2
9.2	Decisions de disseny i arquitectura	2
9.3	Instal·lació	2
9.4	Notes sobre navegadors.....	2
10	Conclusions	2
11	Glossari	2
12	Referències	2

2 Introducció

2.1 Justificació del treball i context en el qual es desenvolupa

La **Web Semàntica** es percep com una de les temàtiques que tindran major repercussió en el futur de la web. Tot i aquest consens en la importància de la Web Semàntica, la informació disponible al respecte, sovint no és prou comprensible i és difícil de posar en context.

Aquest treball pretén ser una aportació a la documentació existent que permeti assolir una bona comprensió del concepte de Web Semàntica i tecnologies associades.

Per a assolir aquest objectiu, als següents apartats es parteix de zero amb material introductori bàsic i s'avança entrant en detall en conceptes claus dins el context de la Web Semàntica, com són el propi RDF (*Resource Description Framework*), els llenguatges de consulta RDF així com els sistemes gestors de bases de dades que treballen amb informació RDF.

Adicionalment a l'aportació teòrica, aquest treball posa a disposició una aplicació web que permet realitzar cerques RDF. L'aplicació web utilitza l'estàndard de llenguatge de consulta RDF més actual (*SPARQL*), sobre diverses bases de dades (artistes musicals, cançons i editors de *weblogs*). Hi ha prediccions que apunten a que SPARQL es convertirà en el llenguatge de consulta RDF més acceptat.

L'aportació de l'aplicació web és doble, doncs pot ésser utilitzada com a base per desenvolupar aplicacions web més complexes i també com a plataforma de pràctiques del llenguatge de consulta SPARQL.

2.2 Objectius del treball

Tal i com s'ha dit a l'apartat anterior, tot i que és relativament fàcil de trobar informació específica sobre un tema puntual, no ho és tant trobar documentació detallada que permeti comprendre la Web Semàntica conceptualment i alhora possibiliti la comprensió d'aspectes especialment rellevants, com ara l'estàndard RDF, els llenguatges de consulta RDF i les bases de dades RDF.

Aquest treball aporta una visió general de la Web Semàntica i alhora n'estudia en suficient detall les principals tecnologies relacionades.

Com a objectiu final per comprovar de forma pràctica la teoria descrita, l'aplicació web permet consultar bases de dades RDF mitjançant el llenguatge de consulta SPARQL.

2.3 Enfocament i mètode seguit.

Donada la manca de bibliografia específica, la única font d'informació consultada per a realitzar el treball ha estat la web. El mètode seguit ha consistit en realitzar recerca d'informació sobre temes específics i incloure la informació a la memòria, intentant sempre mantenir una evolució des de conceptes introductoris com ara la introducció a XML, fins a conceptes més avançats com ara la comparativa dels llenguatges de consulta RDF o la pròpia aplicació web.

2.4 Planificació del projecte

Aquest apartat enumera els continguts desenvolupats, així com la planificació seguida per a portar a terme el treball, tot seguint les fites esperades.

L'objectiu de la planificació descrita a continuació, és determinar la magnitud del treball així com servir de guia i de calendari a seguir, per tal d'obtenir els resultats esperats dins el període de temps establert.

2.4.1 Continguts a desenvolupar

El treball realitzat es materialitza en la present memòria i una presentació resum. Addicionalment, s'ha creat una aplicació web que serveix per comprovar, de forma pràctica, els principals resultats i conclusions teòriques del treball.

2.4.1.1 Memòria

La memòria consisteix en el present document, que descriu, de forma detallada, el resultat dels estudis i conclusions assolits durant la realització del treball.

2.4.1.2 Presentació

La presentació complementa la memòria i serveix com a resum comprensiu d'aquesta última. La presentació inclou els punts més rellevants del treball, així com les conclusions més representatives a les que s'ha arribat.

2.4.1.3 Aplicació Web

L'aplicació web permet comprovar de forma pràctica, punts presentats a la memòria amb un inherent component pràctic; l'aplicació web serveix també per reforçar els apartats de la documentació que fan referència a aquests punts.

2.4.2 Anàlisi de Riscos

A continuació es descriuen les possibles incidències i imprevistos que podien sorgir durant la realització del treball, així com possibles accions a dur a terme per mitigar-ne els efectes.

2.4.2.1 Tasques no representatives

Aquest risc sorgeix quan la subdivisió inicial del projecte en tasques no és realista i obliga a modificar el pla de treball durant l'evolució del projecte.

Impacte	Implica revisar el pla de treball i adequar-lo a la nova distribució de tasques. Requereix un temps de dedicació que redueix el temps disponible per a la resolució de les pròpies tasques. El risc més important sorgeix en la possibilitat de centrar esforços en resoldre tasques innecessàries o irrellevants.
Probabilitat	Alta
Accions de Mitigació	Posar esforços en revisar el pla de treball, especialment les tasques a realitzar, durant el període inicial del projecte.

2.4.2.2 Imprevistos laborals

La dedicació temporal en el projecte es veu reduïda per situacions laborals.

Impacte	Terminis en el desenvolupament del projecte. Desviació de la temporització establerta en el calendari del projecte.
Probabilitat	Mitja
Accions de Mitigació	La meva condició d'autònom, sovint suposa pics de feina seguits de períodes de més (o total) disponibilitat. La disminució en les hores de dedicació durant un període de temps, es pot compensar en una major dedicació en un període anterior o posterior.

2.4.2.3 Tasques massa generals

La subdivisió inicial del projecte en tasques és massa general i provoca una temporització poc acurada per resoldre cada tasca.

Impacte	Terminis en el desenvolupament del projecte. Implica dedicar temps a revisar la planificació del projecte, reduint el temps assignat a la resolució de les tasques. De forma similar al impacte descrit a l'apartat 2.4.2.1, el risc més important consisteix en dedicar excessius recursos a una part de la tasca quan d'altres parts son més importants i requeriran un temps major.
Probabilitat	Alta
Accions de Mitigació	Evitar la manca de visibilitat de subtasques ocasionada per definir tasques massa generals. Posar esforços en revisar el pla de treball, especialment les tasques a realitzar, durant el període inicial del projecte.

2.4.3 Organització del Treball

2.4.3.1 Disponibilitat

La disponibilitat estimada per dedicar al treball ha estat de 1 persona x 1,5 hores al dia x 6 dies a la setmana. Això fa un total de 36 hores mensuals dedicades al projecte.

La temporització anterior, s'estima en un context laboral on es compagina la feina professional amb la feina dedicada al treball. Es consideren els períodes de vacances com a propicis per dedicar més recursos a la resolució del treball, possiblement com a mesures de mitigació davant l'aparició d'algun dels riscos exposats a l'apartat 2.4.2.

2.4.3.2 Relació de tasques

A continuació es detallen els objectius o apartats principals del treball, amb la relació de tasques que cal portar a terme per assolir-los.

A tots els apartats que segueixen, es considerarà que una de les tasques principals consistiria en fer **recerca d'informació**. Tal com es va preveure, aquesta tasca ha tingut un pes destacat dins cada apartat, tal i com es reflecteix al calendari de l'apartat 2.4.4.

2.4.3.2.1 Redacció de la memòria

La redacció de la memòria ha estat el centre del treball i l'activitat que ha requerit més recursos. La tasca principal ha estat la recerca d'informació a la web i posterior redacció dels apartats corresponents.

2.4.3.2.2 Implementació de l'aplicació web per realitzar cerques RDF.

La implementació de l'aplicació web s'ha realitzat simultàniament a la redacció de la memòria. Tal com es preveia, un cop estudiats els llenguatges de consulta RDF i les bases de dades RDF, ja es disposava d'una base de coneixement suficient per començar a implementar una interfície web que interactuï amb els *frameworks* RDF estudiats i serveixi per a reforçar aspectes descrits en el treball teòric.

2.4.3.2.3 Revisió i correcció de la memòria

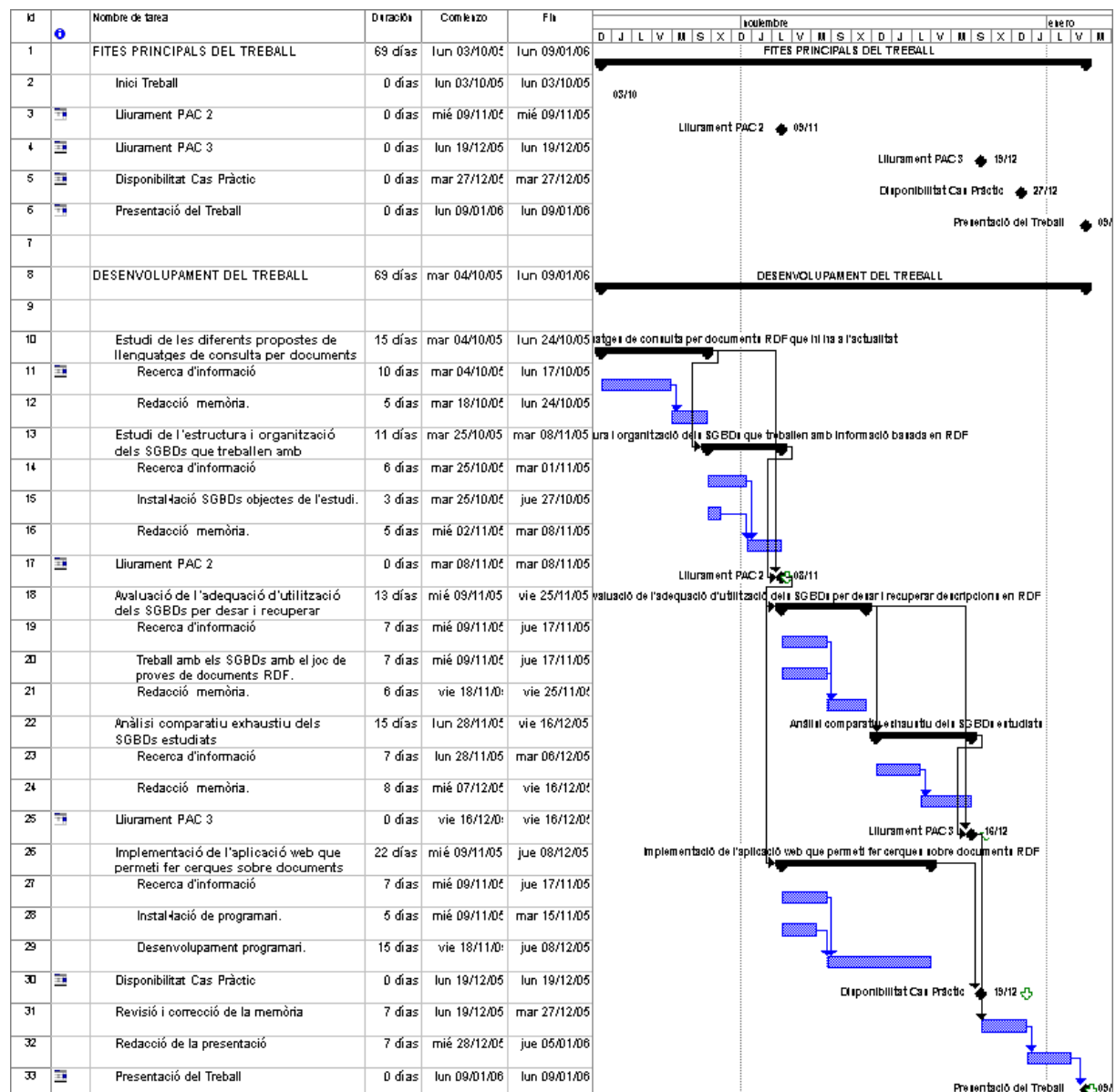
Aquest apartat consisteix en fer una revisió exhaustiva del contingut exposat a la memòria fent les correccions que siguin necessàries des de millores semàntiques fins a correccions sintàctiques.

2.4.3.2.4 Redacció de la presentació

Un cop finalitzada la memòria i partint de la versió final d'aquesta, s'ha treballat en la creació de la presentació - resum del treball.

2.4.4 Calendari de Treball

La següent figura mostra el diagrama de Gantt amb el calendari de treball que s'ha seguit.



2.4.5 Fites principals

Tal com s'observa al diagrama de l'apartat anterior, durant l'evolució del treball s'han considerat fites especialment importants les següents:

- PAC-2: Termini d'entrega el 09 de novembre de 2005
- PAC-3: Termini d'entrega el 19 de desembre de 2005
- Cas pràctic: Disponibilitat estimada pel 27 de desembre de 2005
- Memòria i presentació final: Termini d'entrega el 09 de gener de 2006

2.5 Productes Obtinguts

Com a resultat de la feina realitzada en aquest treball final de carrera, en resulta la present memòria, la presentació resum i l'aplicació web. L'aplicació web, de la qual se'n disposa del codi font documentat, ofereix la possibilitat d'utilitzar-la tant de plataforma de proves com de punt de partida en possibles aplicacions web futures.

2.6 Descripció de capítols

El cos del treball el formen els capítols 3 al 8, descrits a continuació, amb el capítol 9 dedicat a l'aplicació web.

El **capítol 3** es considera el capítol introductor que defineix el concepte de Web Semàntica, posant de manifest la problemàtica amb la web actual i presentant a la Web Semàntica i tecnologies relacionades com a possible solució.

El **capítol 4** consisteix en una introducció detallada al llenguatge XML, necessari per entendre la codificació RDF/XML, la més acceptada per a codificar informació RDF.

El **capítol 5** fa una introducció a l'estàndard RDF. Es descriu el model conceptual RDF així com la sintaxi RDF/XML.

El **capítol 6** introdueix el concepte de consulta RDF així com els trets bàsics necessaris per comprendre qualsevol consulta RDF, indiferentment al llenguatge de consulta en que estigui especificada. S'introdueixen conceptes comuns a altres tipus de consultes, com ara la projecció, i aspectes específics de les consultes RDF com ara la inferència.

El **capítol 7** consisteix en una definició dels principals llenguatges de consulta RDF actuals seguida d'una comparativa que, per cadascun dels llenguatges, n'avalua les característiques exposades al capítol 6. Els llenguatges que s'estudien son RQL, SeRQL, TRIPLE, RDQL, N3, Versa, RDFQL, RxPath i SPARQL.

El **capítol 8** descriu i enumera les característiques més rellevants dels sistemes gestors de bases de dades (SGBD) que treballen amb informació RDF.

El **capítol 9** està dedicat a l'aplicació web. S'hi descriuen les decisions de disseny i l'arquitectura així com les instruccions d'instal·lació.

3 La Web Semàntica

Originàriament, la *World Wide Web* es va concebre com un espai per a compartir informació entre humans. El paradigma d'utilització de la web consistia, per una banda en que una minoria publicés continguts comprensibles per a les persones (text o continguts multimèdia) i per l'altre en una majoria que accedia a aquests continguts, principalment per llegir-los, i més tard també per veure'ls o escoltar-los. La quantitat de continguts va començar a augmentar exponencialment i es va fer necessari un sistema que assistís a les persones en la recerca d'informació, els cercadors web.

Aquest model ha variat ben poc i ens ha acompanyat fins a l'actualitat, tot i que hi ha clars indicis de que la Web està començant a canviar i evoluciona consolidant-se amb el que ja es coneix com a Web 2.0.

Actualment es parla de la Web 2.0 com a evolució de la Web clàssica tal i com la coneixem (o Web 1.0). La principal diferència que aporta la Web 2.0, deixant de banda les innovacions tecnològiques que la possibiliten, és que les persones passen a tenir actituds més participatives i menys passives. La Web 2.0 també es coneix com a Web de lectura i escriptura (*Read-Write Web*)¹, ja que els usuaris –ja no una minoria- utilitzen regularment la Web no només per a obtenir-ne informació, sinó també per interactuar-hi, treballar-hi, aportar-hi coneixement i obtenir-ne resultats.

Exemples d'aquesta tendència son la proliferació dels últims anys de milions de *weblogs* (o *blogs*) que converteixen a persones fins llavors consumidores d'informació, a productors i editors d'informació. Altres exemples son llocs web com ara Flickr, que ha aconseguit que milers de persones publiquin i comparteixin fotografies. L'usuari de la Web 2.0 és un usuari que aporta continguts a la Web. Es passa d'un paradigma de lectura d'informació a escriptura d'informació a la Web. Exemples de llocs web considerats com a representatius de la Web 2.0 son **Flickr**, **del.icio.us** o **Wikipedia**.

Ens trobem doncs en un punt on, tot i que hi ha avanços i millores, hi ha un problema inherent que arrastrem fins i tot a la Web 2.0: **la gran quantitat d'informació**. Continua havent-hi un creixement espectacular pel que fa a quantitat i diversitat de continguts; especialment a la Web 2.0 on la creació de continguts passa a ser hàbit de masses. Això ha

¹ Per referir-se a la Web 2.0, [James Snell](#) utilitza el terme "*chmod 777 Web*".

portat a una situació crítica en que fins i tot els millors cercadors web actuals, que permeten trobar entre bilions de recursos web de forma pràcticament instantània aquells que contenen literalment les paraules que especifiquem, siguin cada cop més insuficients per cercar la informació que realment estem buscant.

La situació és fins i tot pitjor considerant que els cercadors han de tractar un nombre cada cop més elevat de recursos multimèdia com ara àudio i vídeo, cada cop més populars gràcies a les millores d'ample de banda i la capacitat de procés dels ordinadors domèstics.

Aquests tipus de recursos son un problema pels motors de cerca, ja que els cercadors actuen en base al contingut textual dels recursos, principalment pàgines web i arxius de continguts textuais com ara arxius de Microsoft Word o Adobe PDF.

Per tractar eficientment la informació disponible a la Web cal una evolució més enllà del que ens aporta la Web 2.0.

Els següents apartats descriuen la problemàtica en que ens trobem i descriuen la Web Semàntica com a possible solució.

3.1 El problema amb la Web actual

Suposem a mode d'exemple que anem al videoclub a llogar una pel·lícula. Per tal de trobar-la, realitzem una cerca sobre **informació associada a la pel·lícula** com ara el títol o la temàtica. Aquesta informació descriptiva associada a la pel·lícula confereix el que s'anomena **metadades** de la pel·lícula.

De forma similar, per trobar el telèfon d'un conegut, sovint realitzem una cerca sobre la informació associada al número que volem trobar a la guia telefònica. En aquest cas, informació com ara el municipi i nom de la persona conferirien les metadades del número de telèfon.

La idea important és que a la vida quotidiana som capaços de trobar coses a partir de les seves metadades.

A l'exemple anterior, si se sap el número de telèfon, es pot contactar amb la persona. De forma similar, a la Web si se sap la URL, es pot accedir al recurs desitjat. El repte a la web és el de ser capaços de trobar la URL del recurs desitjat mitjançant una cerca sobre les seves metadades.

Tot i que a la Web, les necessitats de trobar informació son les mateixes, la situació és diferent. A la Web hi ha infinitat de tipologies de recursos, el seu nombre es molt gran i en creixement continu. Per altra banda, les metadades associades a aquests recursos son gairebé inexistentes, tret de casos molt concrets.

Les eines actuals de què disposem per trobar informació (recursos) a la Web, majoritàriament consisteix en cercadors que, donat la manca de metadades, apliquen tècniques de força bruta per escanejar el contingut dels recursos. Un cop escanejats, s'indexen les paraules clau i això permet de realitzar cerques mostrant com a resultat una sèrie de recursos, candidats a ser el que realment busquem.

Per posar un exemple, volem trobar un hotel a Nova York. Utilitzem un dels cercadors més coneguts i introduïm la cerca "Hotel New York". El cercador retorna entre els candidats de cerca més rellevants "*New York-New York Hotel & Casino de Las Vegas*" i "*Hotel New York de Rotterdam*". La cerca no ens permet especificar un **component semàntic** que ens permeti discriminar si amb "New York" ens referim al nom de l'hotel o a la localització geogràfica d'aquest.

Per trobar un recurs a la web, els mecanismes de cerca que hi ha actualment tornen un elevat percentatge de recursos candidats que son totalment irrelevants i inconnexes amb el que realment estem buscant. El motiu de fons d'això és la falta de metadades i mecanismes de cerca que actuïn sobre aquestes.

L'actual sistema de cerca de recursos a la web s'ha de millorar. L'apartat següent descriu en què consisteix una possible sol·lució: la Web Semàntica.

3.2 La Web Semàntica

Possiblement la definició més simple, general i alhora més entenedora de què és la Web Semàntica, és la que proposa Tim Berners-Lee:

"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."

- Tim Berners-Lee

Com s'ha dit anteriorment, un tret comú tant de la Web 1.0 com de l'emergent Web 2.0 és la manca de metadades associades explícitament als recursos Web, o dit d'una altra manera, la manca de contingut semàntic explícit dels recursos web.

Es podria pensar que una possible solució passaria per dissenyar un sistema de metadatatge universal, adient per a descriure qualsevol recurs de la Web. En el context d'RDF aquesta solució consistiria en que tothom acordés utilitzar com a metadades un mateix conjunt de propietats o **vocabulari** per descriure qualsevol tipus de recurs.

Intuïtivament ja es veu que seria molt difícil cobrir les necessitats i preferències de tothom i no menys important, aconseguir consens. Continuant amb l'exemple de les biblioteques, tot i que comparteixen un únic tipus de recurs comú, el llibre, i tot i que les biblioteques existeixen des de fa molts anys, no totes les biblioteques utilitzen els mateixos sistemes de metadades per catalogar els llibres. Encara no hi ha hagut consens al respecte.

Donat que la web engloba infinites tipologies de recursos, es pot avançar que posar d'acord biblioteques, videoteques, hospitals i botigues d'animals de companyia en la utilització d'un sistema de metadatatge universal es una tasca utòpica o si més no, molt difícil de realitzar.

La promesa de la Web Semàntica és el naixement de multitud de vocabularis (no tan sols un vocabulari comú), creats específicament per un tipus de recurs concret pels propis interessats en el tipus de recurs (consumidors/publicadors de recursos).

Així doncs, i tornant a l'exemple de les biblioteques, hi haurà diversos vocabularis per catalogar llibres, el que sigui millor o més acceptat s'imposarà com a vocabulari comú i serà el més utilitzat. Si hi ha necessitats particulars, es realitzarà una extensió del vocabulari comú; per exemple si les biblioteques universitàries tenen necessitats diferents que les universitats municipals.

Ja hi ha vocabularis genèrics a partir dels quals ens podem basar per crear metadades de recursos i també per crear nous vocabularis extesos. El més conegut d'aquests vocabularis genèrics és el proposat per DCMI (*Dublín Core Metadata Initiative*).

El llenguatge per descriure recursos Web utilitzant els vocabularis esmentats, s'anomena Resource Description Framework o RDF (estàndard de *World Wide Web Consortium*).

Un aspecte consensuat és la recomanació que les metadades de recursos es creïn i es mantinguin localment pels propis creadors dels recursos metadatats. Això fa preveure multitud de bases de dades RDF distribuïdes per la xarxa. És doncs de vital importància que els motors de cerca RDF siguin capaços de consultar aquestes bases de dades distribuïdes de forma transparent als usuaris.

El futur passa per disposar de bilions de tuples propietat-valor relatives a recursos web (incloent, però no limitades a, pàgines web), d'acord amb uns vocabularis de propietats comuns i motors de cerca RDF *online* per poder trobar el que realment s'està buscant.

Aquesta evolució de la World Wide Web actual, sigui Web 1.0 o Web 2.0, una web amb recursos metadatats comprensibles per als ordinadors, juntament amb motors de cerca amb capacitats semàntiques, és el que es coneix com a **Web Semàntica**.

Els apartats que segueixen presenten els llenguatges XML i RDF, es descriu en què consisteixen les cerques RDF i es realitza un estudi d'alguns dels llenguatges de consulta RDF més rellevants, així com l'estructura dels SGBD (Sistemes Gestors de Bases de Dades) RDF que permeten emmagatzemar i consultar documents RDF.

4 Introducció a XML

En els darrers anys, XML (*eXtensible Markup Language*) s'ha establert com el llenguatge estàndard més àmpliament acceptat tant en la codificació de dades, la integració de sistemes informàtics o en la comunicació i interoperabilitat entre sistemes distribuïts.

En el context d'aquest treball, XML és especialment rellevant ja que una de les representacions més acceptada per representar models RDF es realitza utilitzant XML. La representació de models RDF utilitzant XML s'anomena **codificació RDF/XML**.

Aquest capítol és una introducció a XML. Es comença amb una perspectiva històrica dels orígens de l'estàndard XML i es continua amb una definició tècnica del llenguatge XML, des dels punts de vista sintàctic, gramatical i semàntic.

4.1 Orígens del llenguatge XML

A partir del moment en que es fa públic l'estàndard HTML (1991), el nombre de documents HTML publicats a Internet augmenta de forma exponencial. Aquest creixement va ser alimentat per la facilitat que tenien els autors, de publicar a Internet documents HTML, destinats a una audiència internacional i amb un cost molt baix.

A mitjans dels anys 90 la base documental existent, codificada en HTML, havia superat les expectatives més optimistes. La simplicitat de HTML el feia ideal per a la codificació i distribució via xarxa de documents, destinats a ser llegits davant d'un navegador web.

Moltes empreses van adonar-se del potencial d'utilitzar la infraestructura de la xarxa Internet per establir nous paradigmes de comunicació. Es buscava construir sistemes interoperables entre sí, on la comunicació, en comptes de ser ordinador – persona fos entre ordinadors.

Donat el gran volum d'informació HTML existent a Internet, inicialment es va intentar utilitzar HTML com a llenguatge estàndard per aquesta comunicació entre ordinadors. Aviat es va fer evident que es demanava a HTML que fes coses per les que no havia estat dissenyat.

HTML presentava una sèrie de limitacions que impossibilitaven la seva utilització per a la comunicació d'informació entre ordinadors. Aquestes limitacions eren i són:

- ❖ **Extensibilitat:** HTML no permet als usuaris d'especificar les seves pròpies etiquetes o atributs.
- ❖ **Estructura:** HTML no permet especificar estructures complexes d'informació.
- ❖ **Validació:** HTML no permet especificar restriccions sobre l'estructura i tipus de dades que representa.

Davant el potencial que oferia la xarxa Internet i el nombre creixent d'ordinadors connectats, es va decidir que s'havia de definir d'urgència un nou llenguatge que possibilités la comunicació ordinador – ordinador. Una de les premisses principals era que havia de ser el màxim de simple possible per garantir una acceptació massiva, similar a HTML, però alhora que cobrés les tres limitacions descrites que patia HTML.

Científics de l'organització W3C (*World Wide Web Consortium*²) van entrar en escena proposant com alternativa l'estàndard SGML (*Standard Generalized Markup Language*). SGML ja feia anys que s'utilitzava amb èxit en grans empreses per codificar documents i complia amb els tres requeriments anteriors. Tot i així es va descartar perquè era massa complex i es preveia que no tindria acceptació.

Un dels principals interessats en el nou estàndard eren els fabricants de navegadors, que s'oposaren frontalment a SGML, perquè veien una relació cost/benefici molt desfavorable per incorporar suport SGML als navegadors existents.

Va ser llavors (1997), quan un enginyer de Sun Microsystems, **Jon Bosak**, juntament amb un grup de científics de W3C especialistes en SGML, van formar un grup de treball amb l'objectiu de simplificar l'especificació SGML. Intentarien retallar funcionalitat innecessària, però mantenint els principis d'Extensibilitat, Estructura i Validació.

Al cap d'uns mesos, el w3c anuncia el naixement del nou estàndard: **eXtensible Markup Language** o **XML**.

² <http://www.w3c.org>

4.2 Definició de XML

Estrictament parlant, XML és un *metallenguatge* (un llenguatge per definir altres llenguatges). Com ja s'ha dit, és un subconjunt de SGML on se n'ha eliminat gran part de la complexitat.

L'estàndard XML defineix únicament les regles sintàctiques que ha de complir qualsevol document XML³. Qualsevol aplicació que processi documents XML requereix que els documents XML siguin com a mínim sintàcticament correctes.

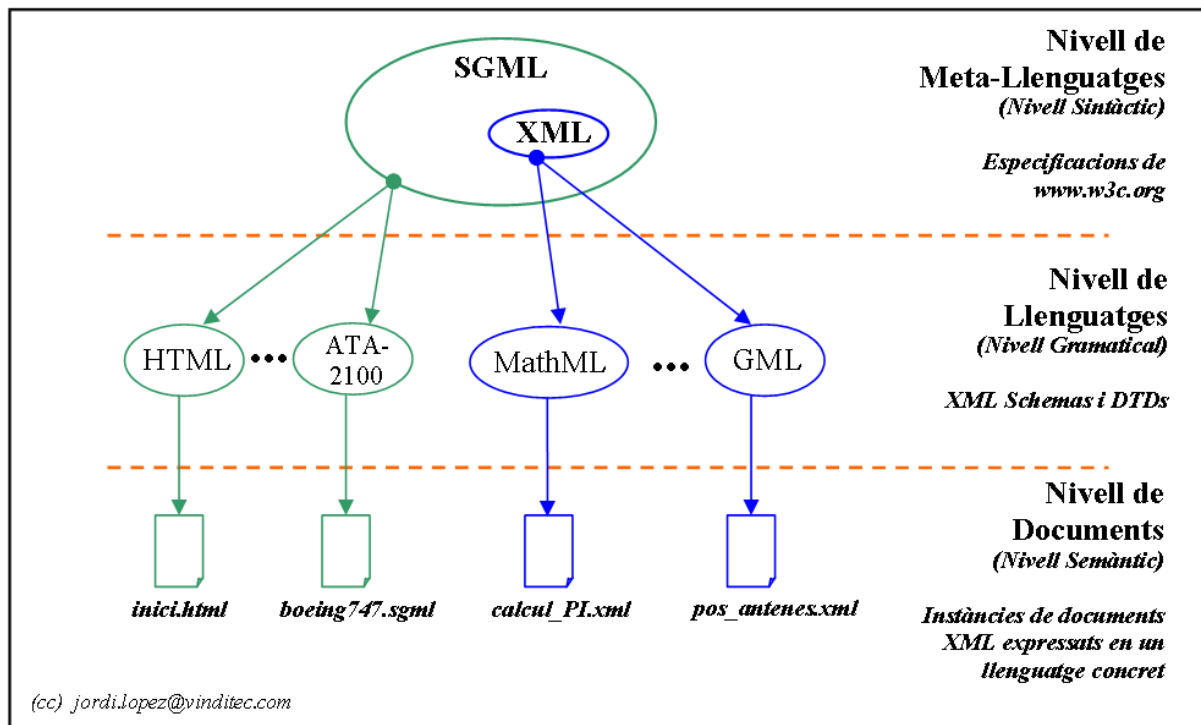
De forma opcional, tot document o instància XML pot complir amb una determinada gramàtica.

La sintaxi XML s'especifica a l'estàndard XML, disponible a la web de W3C, i consisteix en regles com ara "*tot document XML ha de tenir un únic element arrel*". Al següent apartat es descriuen detalladament les regles sintàctiques de XML.

La gramàtica s'especifica mitjançant documents de tipus XMLSchema o DTD (*Document Type Definition*) associats als documents XML subjectes de complir amb la gramàtica. Una regla gramatical pot determinar restriccions com ara '*l'element XML <edat> ha d'aparèixer com a element fill d'un element <alumne>*'. Val a dir que un document XML que no compleixi les regles gramaticals establertes al document XMLSchema o DTD associat, pot perfectament ser sintàcticament correcte.

³ També es pot utilitzar el terme **instància XML**, alternativament a **document XML**.

El següent gràfic pot ajudar a situar conceptualment a XML com a metallenguatge i als documents o instàncies XML en els nivells sintàctic i gramatical esmentats:



4.3 Regles sintàctiques. Documents ben formats.

Com s'ha dit, l'especificació XML defineix les regles sintàctiques bàsiques que tot document XML ha de respectar. Un document que no respecti aquestes regles es diu que és un document XML **no vàlid**. Quan un document XML compleix aquestes regles es diu que és un document XML **ben format**.

4.3.1 Components bàsics XML: Etiquetes, Elements i atributs.

Hi ha tres components bàsics per descriure parts d'un document XML: etiquetes, elements i atributs.

- Una **etiqueta** és el text que apareix entre els símbols de < i >. Hi ha etiquetes d'inici com ara <etiqueta> i etiquetes de final d'element com </etiqueta>.
- Un **element** es compon d'una etiqueta d'inici, una de final i tot el que hi hagi entremig.

- Un **atribut** és una tupla nom-valor dins de l'etiqueta d'inici d'un element.

Per exemple, del següent fragment XML:

```
<persona dni="40344533">  
  <nom>  
    Anna Rius  
  </nom>  
</persona>
```

Podem identificar l'etiqueta d'inici **<nom>** i la corresponent etiqueta de final **</nom>**.

L'element **nom** consisteix en les etiquetes esmentades d'inici i final juntament amb el text **Anna Rius**. Per altra banda, l'etiqueta d'inici **<persona>** conté un atribut 'dni' amb el valor '40344533'.

A continuació es descriuen les regles sintàctiques que tot document XML ha de complir.

4.3.2 Capçalera XML

Tot document XML ha de començar amb una capçalera on s'especifiqui que es tracta d'un document XML, així com la versió de l'estàndard XML en el que es basa. Aquesta informació és utilitzada per els programes o aplicacions que treballen amb documents XML. Una capçalera XML és la següent:

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

La versió de l'estàndard XML va començar i s'ha quedat a la versió 1.0. Addicionalment a la versió de l'estàndard, la capçalera pot opcionalment especificar la codificació (*encoding*) UNICODE que utilitza el document.

4.3.3 Element arrel únic

Tot document XML ha de contenir un únic element arrel, que conté la resta d'elements. Aquest element s'anomena "element arrel".

Ben Format	Invàlid
<pre><?xml version="1.0"?> <cotxe> Seat 600 </cotxe></pre>	<pre><?xml version="1.0"?> <cotxe> Seat 600 </cotxe> <cotxe> Renault 8 </cotxe></pre>

4.3.4 Cal especificar les etiquetes de final d'element

Sempre que s'obre una etiqueta d'inici d'element <element> cal especificar posteriorment l'etiqueta de final d'element </element>.

Ben Format	Invàlid
<pre><?xml version="1.0"?> <pagina> <p>Paràgraf A</p> <p>Paràgraf B</p> </pagina></pre>	<pre><?xml version="1.0"?> <pagina> <p>Paràgraf A <p>Paràgraf B </pagina></pre>

Si un element no conté cap informació s'anomena "element buit". En aquests casos es pot tancar l'element a la mateixa etiqueta d'inici. A l'exemple anterior si trobem un paràgraf buit, sense text, podem utilitzar alternativament <p></p> ó <p/>.

4.3.5 Els elements no es poden superposar

Quan s'obre un element, no es pot tancar fins que ho facin tots els elements que conté.

Ben Format	Invàlid
<pre><?xml version="1.0"?> <llibre> <titol>El alquimista</titol> <autor>Paulo Coelho</autor> </llibre></pre>	<pre><?xml version="1.0"?> <llibre> <titol>El alquimista <autor></titol>Paulo Coelho</autor> </llibre></pre>

4.3.6 La sintaxi XML distingeix entre majúscules i minúscules

Els elements <cotxe> i <Cotxe> es consideren elements diferents.

Ben Format	Invàlid
<pre><?xml version="1.0"?> <llibre> <titol>El alquimista</titol> <autor>(Paulo Coelho)</autor> </llibre></pre>	<pre><?xml version="1.0"?> <Llibre> <titol>El alquimista</Titol> <autor>(Paulo Coelho)</autor> </llibre></pre>

4.3.7 El valor dels atributs ha d'anar entre cometes

És indiferent utilitzar cometa doble o cometa simple mentre siguem consistents amb la nostra decisió. Per exemple, si el valor de l'atribut conté cometes dobles, aquest es pot tancar entre cometes simples.

Ben Format	Invàlid
<pre><?xml version="1.0"?> <llibre edició="tercera"> <titol>El alquimista</titol> <autor>(Paulo Coelho)</autor> </llibre></pre>	<pre><?xml version="1.0"?> <Llibre edició=tercera> <titol>El alquimista</Titol> <autor>(Paulo Coelho)</autor> </llibre></pre>

Els atributs sempre han de tenir un valor. Per exemple, no es considera vàlid un element `<llibre edició>` (on 'llibre' és el nom d'element i 'edició' el nom d'atribut). En aquest cas per indicar que es desconeix l'edició d'un llibre, es podria utilitzar la notació `<llibre edició="">`.

4.4 Regles gramaticals. Documents vàlids.

El compliment de les normes sintàctiques descrites a l'apartat anterior és un aspecte obligatori per considerar un document XML com a tal. Opcionalment, es pot anar més enllà de la sintaxi i definir el contingut gramatical que ha de seguir una determinada instància o document XML. Les regles gramaticals especifiquen aspectes com ara quins elements poden aparèixer a les instàncies XML que compleixen amb la gramàtica, en quina seqüència o ordre apareixen, quins han d'aparèixer com a contingut d'altres formant una jerarquia, i altres detalls bàsics relatius a l'estructuració del document XML.

Per especificar aquestes regles gramaticals s'utilitzen documents DTD (*Document Type Description*) o els cada cop més acceptats documents XML Schema. L'associació d'un DTD o XMLSchema a un document XML fa possible comprovar, mitjançant programari específic, si el document XML compleix amb les regles gramaticals que té associades.

Quan un document XML compleix amb les regles gramaticals especificades a un DTD o XMLSchema associat (per suposat també ha de complir amb les regles sintàctiques XML) es

diu que el document XML és **vàlid**.

Existeixen doncs tres tipus de documents XML, segons si compleixen o no les regles sintàctiques i gramaticals:

- Un document XML es considera **ben format** si compleix les regles sintàctiques però no té associat cap DTD o XMLSchema i per tant no té cap restricció gramatical.
- Un document XML és considera **no vàlid** si no segueix les regles sintàctiques XML, o és sintàcticament correcte però no compleix les regles gramaticals especificades en un document DTD o XMLSchema associat.
- Per últim, un document XML es considera **vàlid** si és sintàcticament correcte i compleix amb totes les regles gramaticals especificades en un DTD o XMLSchema associat.

4.4.1 Documents XML Schema i DTD

Els documents XML Schema es consideren una evolució i millora dels documents DTD.

Algunes avantatges de XML Schema sobre DTDs:

- XML Schema pot definir les mateixes regles estructurals que es poden definir amb DTD i a més pot definir tipus de dades i regles més complexes impossibles d'especificar utilitzant DTD. Per exemple XMLSchema suporta tipus de dades com ara dates, nombres amb decimals, enters i fins i tot expressions regulars. Per altra banda, mitjançant DTDs no es pot restringir continguts d'elements o atributs XML per que siguin d'aquests tipus.
- Amb XML Schema també és possible construir tipus de dades nous a partir dels tipus de dades basics. Aquesta extensibilitat tampoc es pot aconseguir treballant amb DTD.
- Un altre punt a favor de XMLSchema enfront dels DTDs és que els primers es representen en XML, mentre que els DTDs es codifiquen amb una notació particular, no XML.

Per nous treballs amb XML on no calgui compatibilitat amb dissenys antics, es recomana basar el disseny en XML Schema.

4.5 Aspectes semàntics dels documents XML

Els dos apartats precedents han tractat aspectes sintàctics i gramaticals dels documents XML. Pel que fa a la vessant semàntica, XML a l'igual que qualsevol altre tipus de recurs Web, manca de contingut semàntic explícit. Explícit en el sentit que sigui processable i comprensible per a programari. La semàntica de recursos XML s'ha de definir i compartir a nivell de les persones que hagin de tractar aquests recursos XML. Per exemple, saber que els elements amb una etiqueta <nom> contenen el nom d'un producte, és una tasca que es pot acordar a nivell de persones que utilitzen el document XML.

Un cercador web que processi recursos XML, d'entrada no pot deduir que l'element <nom> d'aquest document XML (tot i ser vàlid) correspon al concepte "nom de producte".

Així doncs, podem concloure que els recursos XML pateixen la mateixa manca de contingut semàntic que la resta de recursos Web. Tot i així, i dins el context de la Web Semàntica, XML s'utilitza per a codificar models RDF (*Resource Description Framework*), que estan destinats a dotar de contingut semàntic als recursos Web, inclosos documents XML.

El següent apartat és una introducció a RDF, base fonamental de la Web Semàntica.

5 Introducció a RDF

RDF (*Resource Description Framework*) és un estàndard del W3C per a descriure informació de **recursos** a la WWW (*World Wide Web*).

En el context de la WWW, un **recurs** és qualsevol cosa que pugui ser identificada unívocament mitjançant una URI (*Uniform Resource Identifier*).

5.1 Model RDF

Conceptualment, el model RDF és molt simple:

- Un **recurs** té **propietats** que el descriuen.
- Les propietats tenen **valors**, que poden ser valors literals o altres recursos.
- Tot recurs i tota propietat té una URI que l'identifica unívocament.

Partint d'aquesta base, els recursos es descriuen en RDF construint sentències que especifiquen les propietats i valors dels recursos.

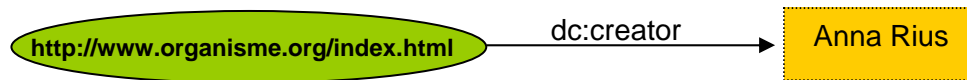
A mode d'exemple, i utilitzant llenguatge natural, podem fer la següent afirmació sobre qui ha creat una certa pàgina web:

La pàgina <http://www.organisme.org/index.html> té un **creador** que es diu **Anna Rius**.

S'hi observen els següents elements:

Recurs:	http://www.organisme.org/index.html
Propietat:	creador
Valor propietat:	“Anna Rius”

El model RDF es pot representar com un graf etiquetat i dirigit, on els nodes representen els recursos i els arcs les propietats. La mateixa sentència també es pot representar gràficament amb el següent graf:



Nota: En notació RDF, les URIs es representen com a nodes el·líptics i els atributs literals com a caixes rectangulars.

Una altra representació alternativa és en forma de triplet recurs-propietat-valor:

<http://www.organisme.org/index.html> dc:creator “Anna Rius”

Nota: En notació de triplet, les URIs es tanquen dins els símbols < i >. Els valors literals, en cas que tinguin espais, s'han de tancar amb doble cometes.

5.2 Sintaxi RDF/XML

Tot i que les tres representacions anteriors es poden utilitzar per construir sentències RDF de forma simple, la representació normativa d'RDF s'anomena RDF/XML i consisteix en una sintaxi XML (eXtensible Markup Language) per codificar sentències RDF. La sentència anterior representada en RDF/XML és:

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://www.purl.org/dc/elements/1.0/">
  <rdf:Description about="http://www.organisme.org/index.html">
    <dc:creator>Anna Rius</dc:creator>
  </rdf:Description>
</rdf:RDF>
```

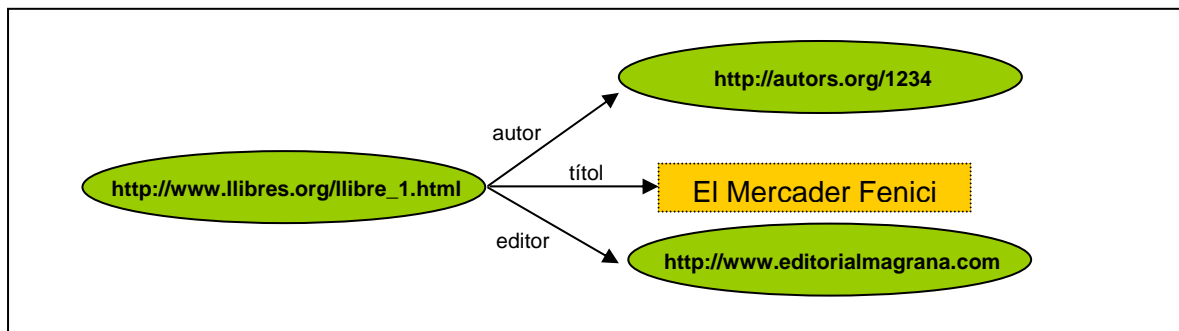
6 Característiques de les consultes RDF

Aquest apartat fa una introducció als trets característics de les cerques RDF, exposant una visió general del mecanisme utilitzat per identificar candidats i també la capacitat d'inferir informació, fins i tot sense aparèixer explícitament al model RDF a cercar.

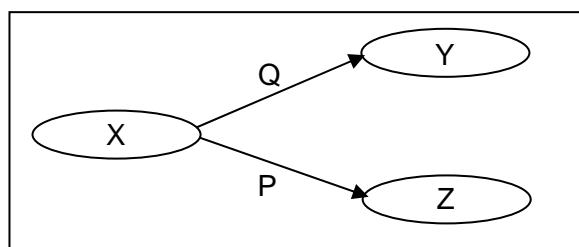
6.1 Definició de consulta RDF

Tal com s'exposa a l'apartat 5.1, els models RDF es poden representar com a grafs etiquetats. Els llenguatges de consulta RDF permeten descriure un graf RDF que encaixi com a sub-graf del graf a consultar.

Donat el graf següent, objecte de la cerca:



El graf:



Encaixa **estructuralment** de tres maneres al graf original:

a)

<code><http://www.llibres.org/llibre_1.html></code>	autor	<code><http://autors.org/1234></code>
<code><http://www.llibres.org/llibre_1.html></code>	títol	"El Mercader Fenici"

b)

<http://www.llibres.org/llibre_1.html>
<http://www.llibres.org/llibre_1.html>

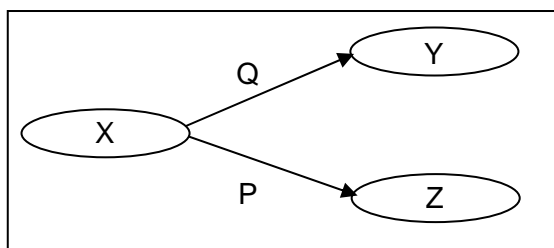
autor <http://autors.org/1234>
editor <http://www.editorialmagrana.com>

c)

<http://www.llibres.org/llibre_1.html>
<http://www.llibres.org/llibre_1.html>

títol "El Mercader Fenici"
editor <http://www.editorialmagrana.com>

Especificant etiquetes als enllaços i nodes, s'aconsegueix discriminar a quin dels tres subgrafs del model RDF ens referim.



Q = autor
P = títol
Z = IGUAL A "El Mercader Fenici"

Aquest cop no hi ha dubte que el subgraf referenciat és:

<http://www.llibres.org/llibre_1.html>
<http://www.llibres.org/llibre_1.html>

autor <http://autors.org/1234>
títol "El Mercader Fenici"

Els llenguatges de consulta per a documents RDF que s'estudien segueixen aquest paradigma alhora de cercar informació a documents RDF.

6.2 Propietats dels llenguatges de consulta RDF

Els següents apartats defineixen diverses propietats subjectes de ser suportades pels diferents llenguatges de consulta RDF. Aquestes propietats, s'utilitzaran en apartats posteriors per comparar els diferents llenguatges de consulta RDF.

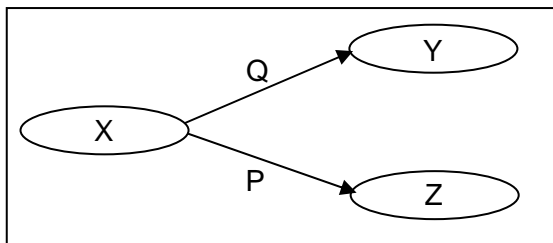
6.2.1 Inferència (Entailment)

El llenguatge RDF permet que les cerques infereixin o dedueixin informació a partir de sentències que no apareixen de forma explícita al graf RDF objecte de la consulta. RDF permet que s'augmenti el contingut semàntic d'un graf RDF mitjançant sentències externes.

Si a l'exemple exposat a 6.1, suposem que té associat un esquema amb la següent sentència:

autor **rdfs:subPropertyOf** persona

Vé a dir que tot autor és també una persona. La següent cerca sobre el graf de l'apartat 6.1:



Q = **persona**

P = **títol**

Z = IGUAL A "El Mercader Fenici"

S'obté el mateix resultat que cercant amb Q = **autor**. Veurem que no tots els llenguatges de consulta RDF suporten aquesta característica.

6.2.2 Model Abstracte RDF

La estructura abstracta de qualsevol model de dades RDF és un graf dirigit i etiquetat. Els llenguatges de consulta idealment han de treballar sobre aquesta representació abstracta i no dependre de la representació (serialització) del model.

Per exemple:

Els elements dins un document XML segueixen un ordre. Una cerca sobre un document RDF codificat en RDF/XML no hauria de tractar l'ordre en que la informació apareix al document XML. Per exemple, la serialització del graf següent:

```
<http://www.llibres.org/llibre_1.html>      autor <http://autors.org/1234>  
<http://www.llibres.org/llibre_1.html>      títol  "El Mercader Fenici"
```

Pot ocasionar dues representacions RDF/XML diferents, depenent de quin triplet es serialitza primer:

```
<rdf:Description about=' http://www.llibres.org/llibre_1.html '>  
  <autor>...</autor>  
  <títol>El Mercader Fenici</títol>  
</rdf:Description>
```

i

```
<rdf:Description about=' http://www.llibres.org/llibre_1.html '>  
  <títol>El Mercader Fenici</títol>  
  <autor>...</autor>  
</rdf:Description>
```

ambdues correctes.

Direm que un llenguatge de consulta RDF **actua sobre el model abstracte RDF** si ignora les peculiaritats presents a la serialització del model.

6.2.3 Tancament

Aquesta propietat requereix que el resultat d'una operació de cerca consisteixi en elements del model de dades a consultar. Si l'operació actua sobre un graf RDF, el resultat de cerca ha de continuar essent un graf RDF.

6.2.4 Ortogonalitat

Un llenguatge de consulta és ortogonal si totes les operacions que ofereix poden ser utilitzades independentment del context.

Les propietats de Tancament i Ortogonalitat permeten realitzar productes entre dades del model (*JOINS*).

6.2.5 Seguretat

Un llenguatge de consulta es considera segur si tota consulta sintàcticament correcta torna un nombre finit de resultats en un temps finit.

6.2.6 Expressions de camí (*Path Expressions*)

Una de les propietats considerades bàsiques dels llenguatges de consulta RDF és la de permetre descriure un graf de consulta que encaixi amb una part del graf a consultar. Per exemple:

“Obtenir el nom de l'autor d'una cançó determinada”.

6.2.7 Expressions de camí opcional (Optional Path Expressions)

El model de dades RDF permet representar informació de forma incompleta, on valors o nodes del graf no siguin especificats. Els llenguatges de consulta RDF han de preveure i suportar aquests casos de manca d'informació. Per exemple:

“Obtenir el nom i el correu electrònic, si en té, dels autors d'una cançó determinada”.

6.2.8 Selecció

Propietat bàsica que significa que es possible extreure informació del model donada una determinada restricció. Tots els llenguatges de consulta objectes d'aquest estudi compleixen aquesta propietat.

Per exemple:

“Obtenir el nom de l'autor d'una cançó determinada”.

6.2.9 Projecció

Propietat bàsica que significa que és possible variar la quantitat de tipus d'informació que s'extreu del model. Tots els llenguatges de consulta objectes d'aquest estudi compleixen aquesta propietat.

Exemple de dues projeccions, on la primera recupera els tipus d'informació 'nom' i 'email' i la segona només el 'nom':

“Obtenir el nom i el correu electrònic de l'autor d'una cançó determinada”.

“Obtenir el nom de l'autor d'una cançó determinada”.

6.2.10 Producte

Propietat que permet combinar informació de diverses triples d'un graf RDF. Tots els llenguatges de consulta objectes d'aquest estudi compleixen aquesta propietat.

Per exemple si un llibre té un autor i alhora l'autor té diverses adreces de correu electrònic, aquesta propietat ens permet obtenir el nom del llibre juntament amb totes les adreces de correu electrònic de l'autor d'un llibre concret.

6.2.11 Unió

La unió permet unir el resultat de dues consultes diferents. Per exemple:

“Obtenir els noms de tots els llibres i (unió) els noms de tots els autors”

6.2.12 Diferència

La diferència simula l'operador algebraic de resta. Per exemple:

“Obtenir els noms de tots els llibres que no continguin la cadena 'Potter'”

6.2.13 Quantificació universal (Universal Quantification)

Aquesta propietat implica que tots els valors han d'acomplir un cert predicat de quantificació.

“Obtenir els autors que han estat avaluats per **tots** els lectors”

6.2.14 Agregació

Les funcions d'agregació calculen un nombre a partir del conjunt total de valors resultants d'una cerca.

S'hi engloben funcions per calcular el Mínim (*Min*), Màxim (*Max*), Mitja (*Average*) i nombre de valors (*Count*) del conjunt de resultats.

6.2.15 Agrupació (Grouping)

Les funcions d'agregació actuen sobre la totalitat de valors especificats al conjunt de resultats. Les funcions d'agrupació es van introduir per tal que fos possible aplicar les funcions d'agregació sobre subgrups de valors dins un mateix conjunt de resultats.

6.2.16 Recursivitat

Aquesta propietat indica que el llenguatge de consulta permet utilitzar predicats recursius a les consultes. Aquesta propietat fa que sovint es pugui construir una consulta recursiva que no finalitzi mai i faci que el llenguatge no compleixi la propietat de Seguretat (6.2.5).

6.2.17 Materialització (Reification)

Aquesta propietat consisteix en la possibilitat de deduir una sentència i tractar-la com a part del model a partir d'una sentència implícita o abstracta. Això permet al motor de consulta tractar una abstracció deduïda del model com si fos part del model, o dit d'una altra manera, crear noves sentències a partir de les sentències existents.

6.2.18 Col·leccions

El llenguatge RDF permet definir grups d'entitats mitjançant col·leccions (*Bag*, *Sequence* i *Container*). El suport a col·leccions dels llenguatges de consulta consisteix en permetre recuperar elements de dins d'aquestes col·leccions. Per exemple :

“Obtenir el primer dels autors d'un llibre concret”

6.2.19 Espais de Noms (Namespaces)

La capacitat de consultar espais de noms permet avaluar expressions a les consultes sobre els espais de noms dels documents RDF.

“Obtenir tots els recursos amb un espai de noms que comenci per ‘http://www.uoc.edu’”

6.2.20 Idioma (Language)

El llenguatge RDF permet especificar l'idioma del valor textual dels literals. Els llenguatges de consulta que permeten tractar idioma, son capaços d'incorporar expressions d'idioma a les sentències. Per exemple, suposant que el títol d'un llibre apareix en diferents idiomes:

“Obtenir el títol en Anglès d'un llibre concret”

6.2.21 Comparacions Lèxiques

Les comparacions lèxiques consisteixen en comparar valors literals sense necessitat de tenir en compte el tipus de dades que representen. Per exemple per un llenguatge de consulta on els operadors sobre literals tractin els valors únicament a nivell lèxic, les següents expressions son certes: “9 és major que 10” ó “'004' és diferent que '04'”.

6.2.22 Comparacions Valuades

Els llenguatges de consulta RDF que suporten tipus de dades XML Schema (6.2.23) compleixen aquesta propietat si permeten avaluar els literals tenint en compte el seu tipus i no tan sols el seu valor lèxic. Per exemple un valor numèric escrit '003' si es compara amb un altre valor numèric escrit com '03', una comparació lèxica els trobarà diferents, mentre que si ambdós valors son del tipus enter, i el llenguatge de consulta suporta comparacions valuades, els veurà com a valors iguals.

Per exemple:

“Obtenir tots els llibres amb un nombre de pàgines major que 400 (valor enter)”

6.2.23 Tipus de dades XML Schema

El llenguatge RDF permet utilitzar tipus de dades definits a XML Schema. Entre altres es defineixen tipus de dades per cadenes de text, numèrics, dates i URIs.

6.2.24 Operadors sobre tipus XMLSchema

Els llenguatges de consulta que compleixen aquesta propietat permeten treballar amb els diferents operadors corresponents als tipus de dades XMLSchema.

6.2.25 Ordenació

Els llenguatges de consulta que compleixen aquesta propietat permeten d'obtenir els resultats de cerca ordenats segons algun criteri.

6.2.26 Consulta simultània a múltiples models RDF

Aquesta propietat consisteix en la capacitat dels llenguatges de consulta de resoldre una mateixa consulta contra diferents models RDF, per exemple diferents documents RDF.

6.2.27 Sintaxi similar a SQL (*Structured Query Language*)

Aquesta última característica, més que propietat, indica si la sintaxi de les consultes és similar a la sintaxi SQL. Es considera una característica rellevant perquè actualment SQL és un llenguatge de consulta molt estès i acceptat; és d'esperar que si els llenguatges de consulta RDF segueixen una sintaxi similar a SQL, tindran més acceptació que no si utilitzen una sintaxi especialitzada.

7 Estudi de llenguatges de consulta per documents RDF

Els següents apartats presenten una descripció, seguida d'una comparativa d'alguns dels principals llenguatges de consulta RDF actuals. En concret es consideren els llenguatges **RQL**, **SeRQL**, **TRIPLE**, **RDQL**, **N3**, **Versa**, **RDFQL**, **RXPath** i **SPARQL**.

7.1 RQL

RQL (RDF Query Language) és desenvolupat i mantingut al *Institut d'Informàtica FORTH*⁴ de Grècia, com a part del paquet RDF anomenat "*ICS-FORTH RDFSuite*". La totalitat de components de *RDFSuite* s'ofereixen com a codi obert.

El llenguatge de consulta RQL permet realitzar consultes tant a nivell de recursos com a nivell de propietats (nodes i enllaços) de les dades RDF. Compleix la propietat d'ortogonalitat, però no la de tancament ja que el resultat de les consultes no es torna com a grafs RDF.

RQL imposa una sèrie de restriccions als models RDF, com ara que cada propietat ha de tenir exactament un domini i un rang de valors especificat de forma explícita; es pot dir que la semàntica RQL no és totalment compatible amb la semàntica permesa per RDF.

7.2 SeRQL

SeRQL (Sesame RDF Query Language) és el llenguatge de consulta RDF de la base de dades RDF anomenada **Sesame**.

SeRQL és una evolució que incorpora trets destacables de llenguatges existents, principalment RDQL, N3 i RQL. És considerada un llenguatge àmpliament reconegut i utilitzat.

La quantitat i qualitat de la documentació disponible a la web de OpenRDF⁵, juntament al fet que es tracta d'una iniciativa de codi obert, son trets a destacar.

⁴ <http://www.ics.forth.gr/>

7.3 TRIPLE

TRIPLE és un llenguatge de consulta, inferència i transformació RDF, tal i com el defineixen els seus creadors. Per expressar les triples RDF, s'utilitza notació d'expressions F-Logic (Subjecte[Predicat --> Objecte] o S[P->O]).

TRIPLE retorna els resultats en forma tabular i per tant no compleix la propietat de tancament. TRIPLE tampoc compleix la propietat de Seguretat ja que permet crear sentències que, tot i ser sintàcticament correctes, fan entrar al sistema en bucles infinits.

TRIPLE consisteix no tan sols en el llenguatge de consulta sinó també en l'entorn d'execució complet per utilitzar el llenguatge. El sistema TRIPLE emmagatzema el model RDF realitzant canvis al model original, adaptant-lo a una codificació pròpia. Per un mateix model RDF, TRIPLE pot donar resultats no esperats o que difereixin dels resultats obtinguts consultant el mateix model amb d'altres sistemes.

Un tret a destacar de TRIPLE és que permet consultar diferents models RDF alhora de forma simultània.

7.4 RDQL

El llenguatge RDQL es basa en la informació explícita del graf RDF objecte de la consulta. RDQL no tracta informació externa al propi graf RDF, com ara documents RDF Schema.

El resultat de les consultes es torna en forma tabular i per tant no compleix les propietats de Tancament ni Ortogonalitat. RDQL compleix la propietat de Seguretat.

Per nous treballs és suggereix començar amb SPARQL en comptes de RDQL, ja que SPARQL deixa obsolet RDQL tal com indica el propi autor de RDQL, *Andy Seaborne*, al *tutorial de RDQL* ⁶

⁵ <http://www.openrdf.org>

⁶ <http://jena.sourceforge.net/tutorial/RDQL/>

7.5 N3

N3 o Notation3 consisteix en una sintaxi en text pla (no XML) per RDF. El va idear Tim Berners Lee com a alternativa a la codificació XML, per tal de simplificar el treball amb informació RDF.

No és per si mateix un llenguatge de consulta, però suporta el que s'anomena regles (*rules*). Les regles no es poden crear o executar de forma interactiva sinó que cal emmagatzemar-les en un fitxer i llavors es pot extreure les dades RDF que compleixen les regles mitjançant utilitats com ara el *parser* CWM, part de la plataforma SWAP (*Semantic Web Application Platform*).

N3 compleix les propietats de Ortogonalitat, Tancament i Seguretat.

7.6 Versa

En comptes de dissenyar la sintaxi de Versa perquè fos proper a la sintaxi SQL, es va dissenyar de nou una sintaxi orientada a consultar bases de dades consistents en triples RDF. La sintaxi de Versa té influència de XPath. .

El punt més rellevant de Versa és que permet operacions d'agrupament a l'estil GROUP BY de SQL i també ordenació de resultats; característiques que la majoria de llenguatges avaluats no disposen.

7.7 RDFQL

RDFQL és un llenguatge totalment integrat dins el producte comercial anomenat RDF Gateway de l'empresa *Intellidimension*⁷.

Un aspecte a destacar de RDFQL és que suporta un bon nombre de funcions, incloent funcions com *count()*, *min()*, *max()*, *sum()*, *avg()* que actuen sobre dades agrupades.

RDF Gateway inclou eines com ara un editor RDFQL. Disposa de molt bona documentació online. L'empresa ofereix la opció d'obtenir una versió del producte gratuïta per ús no comercial.

7.8 RxPath

RXPath és un llenguatge de consulta RDF que utilitza la sintaxi XPath. Forma part del conjunt d'utilitats RDF anomenat Rx4RDF.

Realitza un mapeig del model abstracte RDF a un model més adient per ser consultat mitjançant XPath. El model utilitzat per RxPath consisteix en un document XML on els nodes de primer nivell representen els recursos RDF, els fills dels nodes de primer nivell representen les propietats i alhora els seus fills els valors de les propietats.

7.9 SPARQL

SPARQL és la proposta més recent de llenguatge de consulta RDF portada a terme per W3C (World Wide Web Consortium). SPARQL es considera un dels candidats a convertir-se en l'estàndard de llenguatge de consulta RDF.

SPARQL no suporta inferència i per tant es centra en consultes sobre la informació explícita del graf RDF. L'argument per no incorporar aquesta funcionalitat és que amb SPARQL s'ha

⁷ <http://www.intellidimension.com/>

buscat la simplicitat i l'acceptació per part d'usuaris externs als ambients RDF.

Punts a destacar de SPARQL vers altres llenguatges son la capacitat del llenguatge d'obtenir resultats de cerques en format tabular, subgraf RDF del graf consultat o fins i tot construir un graf RDF nou a partir dels resultats.

La capacitat de resoldre cerques contra múltiples orígens RDF alhora és també una característica a ressaltar.

Tot i ser una especificació recent (21 de juliol de 2005), tant la versió més recent de Sesame (v2.0), com el framework RDF Jena, ja suporten SPARQL com a llenguatge de consulta.

7.10 Taula comparativa

Ampliant el que s'ha exposat als apartats anteriors, la següent taula mostra per cadascun dels llenguatges de consulta avaluats, si compleix o no les diferents propietats enumerades a l'apartat 6.2.

Hi ha propietats que determinats llenguatges les poden complir mitjançant consultes intermèdies o bé de forma indirecta. Les propietats que no es poden assolir de forma directa, s'han considerat com a no suportades. En aquests casos es mostra un símbol de menor (-) dins la cel·la.

	RQL	SeRQL	TRIPLE	RDQL	N3	Versa	RDFQL	RXPath	SPARQL
Inferència	Sí	Sí	No -	No	No -	No	Sí	Sí	No
Model Abstracte RDF	Sí	Sí	No	Sí	Sí	Sí	Sí	Sí	Sí
Tancament	No	Sí	No	No	Sí	Sí	Sí	Sí	Sí
Ortogonalitat	Sí	Sí	Sí	No	Sí	Sí	Sí	Sí	Sí
Seguretat	Sí	No	No	Sí	Sí	Sí	Sí	Sí	Sí
Expressions de camí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí
Expr. de camí opcional	No -	Sí	No	No	No	Sí	Sí	No -	Sí
Selecció	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí
Projecció	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí
Producte	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí
Unió	Sí	Sí	Sí	No	Sí	Sí	Sí	Sí	Sí
Diferència	Sí	Sí	No	No	No	No -	Sí	Sí	Sí
Quantificació Universal	Sí	Sí	No -	No	No	No	Sí	No	No
Agregació	Sí	No	No	No	Sí	Sí	Sí	Sí	No
Agrupació	No	No	No	No	No	Sí	Sí	No	No
Recursivitat	No	No	Sí	No	Sí	Sí	Sí	Sí	No
Materialització	Sí	Sí	No -	No	No	No -	Sí	Sí	Sí
Col·leccions	Sí	Sí	No -	No	No -	No -	Sí	Sí	Sí
Espais de Noms	Sí	Sí	No	No	Sí	No	Sí	Sí	Sí
Idioma	No	Sí	No	No	No	No	Sí	Sí	Sí
Comp. Lèxiques.	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí
Comp. Valuades	Sí	Sí	No	No	No	Sí	Sí	Sí	Sí
Tipus XMLSchema	Sí	Sí	No	No	No	No	No	No	Sí
Operacions XMLSchema	Sí	Sí	No	No	No	No	No	No	Sí
Ordenació	No	No	No	No	No	Sí	Sí	Sí	Sí
Múltiples models RDF	No	No	Sí	No	No	No	No	No	Sí
Sintaxi similar a SQL	Sí	Sí	No	Sí	No	No	Sí	No	Sí

8 Estudi de SGBDs que treballen amb informació RDF

Hi ha diferents tendències per emmagatzemar models abstractes RDF de manera que sigui possible de realitzar-hi consultes de forma eficient. Hi ha sistemes que opten per utilitzar un SGBD Relacional, com ara RQL, RDFQL o Jena. D'altres sistemes transformen el model RDF en una representació adient per a les necessitats d'un llenguatge de consulta específic, com ara RxPath.

A continuació es descriuen alguns dels SGBDs que treballen amb els llenguatges descrits a l'apartat anterior. En concret s'ha seleccionat **RQL** i **Jena** com a exemples de *frameworks* que utilitzen SGBDs de tipus relacional, i **Sesame** en la variant de SGBD natiu XML.

Adicionalment es descriu el SGBD utilitzat a **rdfDB**⁸ com a sistema alternatiu d'emmagatzematge utilitzant la base de dades no SQL *SleepyCat*⁹, que emmagatzema la informació en forma de tuples clau-valor.

8.1 RQL

La base de dades sobre la que actua RQL, utilitza el sistema gestor de bases de dades relacional (SGBDR) PostgreSQL per emmagatzemar la informació RDF.

Sobre aquesta base, l'interpret RQL actua parsejant la sintaxi de les consultes RQL (*Parser*), construint el graf de la consulta (*Graph Constructor*), transformant el graf resultant a consultes SQL (*SQL Translator*) i finalment accedint a la base de dades amb les sentències SQL resultants (*Evaluation Engine*).

La presentació de resultats segueix un camí diferent ja que, com s'ha dit a la comparativa de l'apartat 7, RQL no compleix la propietat de tancament i per tant no es construeix un graf RDF a partir dels registres resultants de la consulta SQL, sinó que es presenta el resultat mostrant el valor de les variables projectades de forma tabular.

⁸ <http://www.guha.com/rdfdb/>

⁹ <http://www.sleepycat.com>

8.2 Sesame

La base de dades Sesame es recolza en sistemes d'emmagatzematge existents diversos, com ara SGBDR com MySQL i PostgreSQL, sistemes d'arxius i fins i tot estructures en memòria.

Sesame utilitza una capa d'abstracció anomenada SAIL (Storage And Inference Layer) per interactuar amb el sistema d'emmagatzematge seleccionat. Dóna flexibilitat pel que fa al repositori de dades i actua com a una capa que transforma tant la lògica de consulta com les dades RDF, cap i des dels diferents sistemes d'emmagatzematge.

8.3 rdfDB

La base de dades rdfDB emmagatzema la informació RDF en tres repositoris de tuples clau-valor Sleepycat: **arcindex.db**, **findex.db** i **rindex.db**.

El contingut de cadascun dels tres repositoris Sleepycat és prou senzill i es descriu a continuació.

Donada una sentència RDF on **S₁** és el recurs, **A₁** és l'arc o propietat i **T₁** l'objecte o valor.

Definim la variable entera **Tipus** que pren els valors **1** si T₁ és un recurs, **2** si és un enter i **3** si és una cadena de caràcters.

Definim la variable entera **Font** que pren valors numèrics que permeten al sistema identificar la URL de on prové la sentència RDF.

- L'estructura del repositori **findex.db** és la següent:

La clau consisteix en la concatenació de **S₁ + "|" + A₁**

El valor consisteix en la concatenació de **Tipus + "\t" + Font + "\n" + T₁**.

- L'estructura del repositori **rindex.db** és el següent:

Clau = **T**₁ + "|" + **A**₁

Valor = **Tipus** + "\t" + **Font** + "\n" + **S**₁

- I l'estructura del repositori **arcindex.db**:

Clau = **A**₁

Valor = **Tipus** + "\t" + **Font** + "\n" + **S**₁

Tot i que rdfDB requereix de la base de dades Sleepycat instal·lada, actua com un servidor a l'escolta d'un port TCP, extremadament lleuger i que es pot accedir des de qualsevol aplicació client.

8.4 Jena

Jena és un entorn o *framework* per crear aplicacions per la Web Semàntica utilitzant el llenguatge de programació Java.

Jena es distribueix com a codi obert i inclou:

- Una interfície de programació d'aplicacions (API) RDF.
- Lectura i escriptura de documents RDF codificats en RDF/XML i N3.
- Una interfície de programació d'aplicacions OWL.
- Possibilitat d'emmagatzemar informació RDF a disc o en memòria.
- Suport pel llenguatge de consulta RDQL de forma nativa i SPARQL via el mòdul ARQ.

Jena és el framework en que es basa l'aplicació Web que es descriu a l'apartat següent.

9 Aplicació Web per cercar documents RDF

9.1 Objectiu

Amb la construcció de l'aplicació web es busca assolir una aplicació pràctica a tot el material exposat en aquest document. L'aplicació web ha de permetre realitzar cerques sobre una base de dades RDF i mostrar-ne els resultats per pantalla de forma interactiva.

Per a facilitar les proves amb l'aplicació web, s'ofereixen consultes predefinides que es poden utilitzar també com a plantilles per a crear-ne de més complexes.

Les consultes predefinides permeten comprovar aspectes de SPARQL com ara ordenació de resultats de cerca, cerca sobre dades amb sentències opcionals i filtrar resultats utilitzant expressions regulars.

9.2 Decisions de disseny i arquitectura

Les dades RDF objecte de les cerques, consisteixen en la informació RDF posada a disposició del projecte sobre artistes i cançons (**artists.rdf** i **tracks.rdf**). Addicionalment, i per tal de demostrar consultes més complexes sobre models RDF menys uniformes (com ara amb sentències opcionals), l'aplicació web ofereix la possibilitat de consultar dades RDF sobre creadors de weblogs (**bloggers.rdf**), disponible a la web *planetRDF.com*¹⁰.

Tot i que alguns dels SGBD RDF comercials estudiats ofereixen molt bona documentació (com ara RDF Gateway), s'ha optat per utilitzar únicament alternatives de codi obert per construir l'aplicació web de principi a fi.

L'aplicació web utilitza el *framework* RDF **Jena** amb el mòdul **ARQ** que permet especificar consultes en llenguatge SPARQL.

Com a servidor d'aplicacions web s'utilitza el contenidor de servlets i pàgines JSP **Tomcat**, de *Apache Software Foundation*.

Durant l'inici i l'execució de l'aplicació web, a la consola del sistema operatiu es mostren missatges de *debug* que donen informació del que l'aplicació web està realitzant.

¹⁰ <http://www.planetrdf.com>

9.3 Instal·lació

L'aplicació web es distribueix com un únic arxiu WAR (*Web Application Archive*) desplegable en qualsevol servidor d'aplicacions Java que suporti l'estàndard Servlet 2.3.

Per instal·lar l'aplicació web simplement cal desplegar (fer un *deploy*) de l'arxiu **cercadorRDF.war** al servidor d'aplicacions. Sovint, només cal copiar l'arxiu .war al directori 'webapps' del servidor d'aplicacions i fer una aturada i engegada del servidor.

No cal descarregar ni instal·lar Jena ja que totes les dependències necessàries estan incloses a l'arxiu WAR. L'arxiu WAR també inclou les dades RDF objectes de les consultes. Un cop desplegada l'aplicació web, les dades RDF es troben al directori webapps/WEB-INF/dades/.

9.4 Notes sobre navegadors

L'aplicació web ha estat provada amb èxit utilitzant tant el navegador web Internet Explorer versió 6 com Mozilla Firefox versió 1.0.4; tot i que hauria de funcionar també amb qualsevol navegador que permeti obrir documents XML i que suporti Javascript.

La única incidència detectada ha estat en mostrar els resultats de les cerques en format XML. En aquest cas, i només amb les dades provinents de **tracks.rdf**, Internet Explorer mostra un error perquè l'XML resultant de la cerca conté algun caràcter que el navegador interpreta com a erroni. En aquest cas per veure els resultats de la cerca es pot optar per obtenir el resultat de la cerca en format taula, que es mostra correctament. Aquest problema no s'ha detectat utilitzant Mozilla Firefox, tot i que triga considerablement més en mostrar l'XML resultant.

10 Conclusions

El concepte de Web Semàntica, juntament amb les tecnologies que la possibiliten, confereixen el marc futur, tant evolutiu com tecnològic, de la web tal i com la coneixem actualment.

Les previsions apunten a que en el futur es disposarà d'un gran nombre de metadades RDF relatives a recursos web, metadades que seguiran vocabularis comuns i que es trobaran emmagatzemades en multitud de bases de dades RDF distribuïdes per la web i que seran cercades mitjançant motors de cerca RDF *online*, que permetran descobrir recursos web amb uns nivells de fiabilitat molt més alts que a l'actualitat.

S'espera que el fet de dotar de contingut semàntic als recursos web, possibiliti uns alts nivells de precisió en les cerques i en definitiva, una major col·laboració i enteniment entre persones i ordinadors.

Com a nota personal, aquest treball m'ha permès descobrir i aprofundir en tots aquests temes, la majoria completament nous per a mi, fins al punt d'estar en disposició de crear l'aplicació web que m'ha permès consolidar i posar en pràctica bona part dels continguts teòrics presentats.

La Web Semàntica té per endavant un llarg camí per recórrer; fet que ens ofereix a tots la possibilitat de participar en aquesta evolució aportant coneixement i treball. Espero que aquest treball final de carrera serveixi d'introducció a la Web Semàntica i es pugui utilitzar com a punt de partida per a treballs futurs.

11 Glossari

A continuació segueix una breu definició de diversos termes que apareixen al text (en ordre d'aparició):

World Wide Web: La World Wide Web és un sistema d'hipertext que funciona sobre la xarxa física Internet i que possibilita l'accés a infinitat de recursos mitjançant identificadors específics.

Web: Veure *World Wide Web*.

WWW: Veure *World Wide Web*.

Web Semàntica: Projecte que té per objectiu crear un medi universal d'intercanvi d'informació, mitjançant l'associació de contingut semàntic als recursos web comprensible tant per a persones com per ordinadors.

RDF: Acrònim de **R**esource **D**escription **F**ramework. Especificació publicada per W3C i que defineix un model de metadades per a descriure informació de recursos a la web.

Contingut web: Qualsevol tipus d'informació disponible i accessible a la web, siguin pàgines, imatges, vídeos o qualsevol altre tipus de contingut.

Recurs web: Veure Contingut web.

XML: Acrònim d'e**X**tensive **M**arkup **L**anguage.

Online: Terme que s'utilitza com a sinònim de 'formar part de la web'.

Weblog: Diari online, sovint mantingut per particulars, al que se subscriuen lectors interessats en la temàtica del diari, formant un canal de distribució d'idees sense restriccions i accessible a tothom.

Blogger: Es coneix amb aquest nom a l'editor d'un diari online.

Internet: Xarxa física mundial sobre la que es construeix la Web.

SPARQL: Acrònim de *SPARQL Protocol and RDF Query Language*. És un nou estàndard per a consulta de dades RDF emès recentment pel consorci W3C.

W3C: Acrònim per **World Wide Web Consortium**. És l'organisme que edita els estàndards més rellevants i d'ús més general de la web.

Aplicació Web: Una aplicació web consisteix en un programa que s'executa en un servidor que forma part de la web i que presenta una interfície d'usuari accessible des d'un navegador web estàndard.

Base de Dades RDF: Base de dades amb la particularitat que emmagatzema i permet gestionar informació RDF.

Framework: Estructura que comprèn programari i eines de suport i que serveix com a base per a la creació de projectes de programari de més alt nivell.

RDF/XML: Codificació de dades RDF utilitzant XML.

Cercador Web: Aplicació web que permet cercar recursos a la web a partir de paraules clau. Els resultats dels cercadors web s'utilitzen per part d'usuaris humans ja que els resultats requereixen d'interpretació per part d'un usuari humà per tal de determinar-ne la rellevància.

Blog: Veure Weblog.

Metadades: Informació descriptiva associada a un recurs web.

URL: Acrònim d'**Uniform Resource Locator**. Una URL és una seqüència de caràcters que identifica de forma unívoca un recurs a la web. També es coneix com 'adreça web' o 'identificador de recursos web'.

DCMI: Acrònim de **Dublín Core Metadata Initiative**. DCMI proposa un vocabulari estàndard de termes subjectes de ser utilitzats per a descriure objectes digitals, inclosos recursos web.

Metallenguatge: Llenguatge utilitzat per definir altres llenguatges. XML és un metallenguatge.

UNICODE: estàndard que té per objectiu definir la codificació digital de text especificat en qualsevol llenguatge.

WAR: Acrònim de **W**eb **A**pplication **aR**chive. Es tracta d'un fitxer comprimit que conté una aplicació web, instal·lable a qualsevol servidor d'aplicacions Java.

12 Referències

El treball exposat en aquest document ha estat possible gràcies a les següents fonts d'informació:

1. **XML for the absolute beginner**,
<http://www.javaworld.com/javaworld/jw-04-1999/jw-04-xml.html>
2. **Introducció a XML**,
<http://www.ibm.com/developerworks/edu/x-dw-xmlintro-i.html>
3. **A Comparison of RDF Query Languages**,
<http://www.aifb.uni-karlsruhe.de/WBS/pha/rdf-query/rdfquery.pdf>
4. **Towards an RDF Query Language – Comments on an Emerging Standard**,
http://www.sigsemis.org/columns/rdf/Querying/document_view
5. **The RQL Query Language**,
<http://139.91.183.30:9090/RDF/RQL/>
6. **The SeRQL query language**,
<http://www.openrdf.org/doc/sesame/users/ch06.html>
7. **RXPath Specification Proposal**,
<http://rx4rdf.liminalzone.org/RXPath>
8. **SPARQL. Query Language for RDF**,
<http://www.w3.org/TR/rdf-sparql-query/>
9. **rdfDB: An RDF Database**,
<http://www.guha.com/rdfdb/>
10. **Berkeley Sleepycat DB**,
<http://www.sleepycat.com/>
11. **The Sesame Framework**,

<http://www.openrdf.org/>

12. **The Versa Query language,**

<http://www.xml.com/pub/a/2005/07/20/versa.html>

13. **The N3 Query language,**

<http://logicerror.com/notation3>

14. **PostgreSQL,**

<http://www.postgresql.org/>

15. **MySQL,**

<http://www.mysql.org/>

16. **Sesame: A Generic Architecture for Storing and Querying RDF and RDF Schema,**

<http://www.openrdf.org/doc/papers/Sesame-ISWC2002.pdf>

17. **Contenedor de Servlets Apache Tomcat,**

<http://tomcat.apache.org/>

18. **Search RDF data with SPARQL,**

<http://www-128.ibm.com/developerworks/xml/library/j-sparql/>

19. **Jena Semantic Web Framework,**

<http://jena.sourceforge.net/>

20. **Introducció a Jena,**

<http://www-128.ibm.com/developerworks/java/library/j-jena/>

21. **Introducció a ARQ (extensió de Jena per treballar amb SPARQL),**

<http://jena.sourceforge.net/ARQ/documentation.html>

22. **Part de les dades RDF utilitzades a l'aplicació web provenen de PlanetRDF.com,**

<http://www.planetrdf.com/>

