



DADES OBERTES ENLLAÇADES BIBLIOGRÀFIQUES

Daniel Bosch Vega
Grau en Informàtica

Consultor: Joan Anton Pérez Braña

CURS 2016-2017 – Primer semestre

GNU Free Documentation License (GNU FDL)

Copyright © 2016 Daniel Bosch Vega.

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.

A copy of the license is included in the section entitled "GNU Free Documentation License".

FITXA DEL TREBALL FINAL

TÍTOL DEL TREBALL:	Dades obertes enllaçades bibliogràfiques
NOM DE L'AUTOR:	DANIEL BOSCH VEGA
NOM DEL CONSULTOR:	JOAN ANTON PÉREZ BRAÑA
DATA DE LLIURAMENT:	01/2017
ÀREA DEL TREBALL FINAL:	WEB SEMÀNTICA
TITULACIÓ:	GRAU EN INFORMÀTICA
RESUM DEL TREBALL:	
<p>Dades obertes enllaçades bibliogràfiques</p> <p>La Web Semàntica constitueix una evolució del Web tradicional, i gràcies a la filosofia de les dades obertes enllaçades introduïda per Tim Berners-Lee, la informació aporta un significat ben definit, permetent així una millor cooperació entre persones, i sobretot entre màquines, promovent la idea d'un món més col·laboratiu.</p> <p>Les dades enllaçades s'utilitzen, avui en dia, a molts i diferents camps d'acció essent un d'ells el de les referències bibliogràfiques en biblioteques, museus i arxius de qualsevol tipus de documentació per la creació i intercanvi de dades com alternativa als sistemes tradicionals.</p> <p>En aquest treball, es vol realitzar un estudi de l'estat de les arts de les dades enllaçades en el camp de les referències bibliogràfiques i aportar una visió global del seu estat actual i del seu futur desenvolupament. Aquest estudi es complementarà amb un exemple de la seva utilització real d'aquesta tecnologia mitjançant el desenvolupament d'un API que aprofita les dades enllaçades obertes per extreure la bibliografia d'un autor (o conjunt d'ells) de la web de dades amb consultes en SPARQL, i mitjançant la creació d'una petita aplicació exemple que utilitza aquesta API per mostrar un cas real d'ús.</p>	
<p>Datos abiertos enlazados bibliográficos</p> <p>La Web Semántica constituye una evolución del Web tradicional, y gracias a la filosofía de los datos abiertos enlazados introducida por Tim Berners-Lee, la información aporta un significado muy definido, permitiendo así una mejor cooperación entre las personas, y sobre todo entre las máquinas, promoviendo la idea de un mundo más colaborativo.</p> <p>Los datos enlazados se utilizan, hoy en día, en muchos y diferentes campos de acción siendo uno de ellos el de las referencias bibliográficas en bibliotecas, museos y archivos de cualquier tipo de documentación para la creación y intercambio de datos y como alternativa de los sistemas tradicionales.</p> <p>En este trabajo, se quiere realizar un estudio del estado de las artes de los datos enlazados en el campo de las referencias bibliográficas y aportar una visión global de su estado actual y posible futuro desarrollo. Este estudio se complementará con un ejemplo de la utilización real de dicha tecnología mediante el desarrollo de un API que aprovecha los datos enlazados abiertos para extraer la bibliografía de un autor (o conjunto de ellos) de la web de datos con consultas en SPARQL, y mediante la creación de una pequeña aplicación ejemplo que utiliza esta API para mostrar un caso de uso real.</p>	

Bibliographic open linked data

The Semantic Web constitutes an evolution of the traditional Web, and thanks to the philosophy of the open linked data entered by Tim Berners-Lee, the information includes a very clear-cut meaning, allowing a better cooperation between people, and especially between machines, promoting the idea of a more collaborative.

Today, the use of LOD has been extended on many and different fields, being one of them the field of bibliographic references in libraries, museums and archives of any kind of documentation for the creation and that exchange of data being an alternative to the traditional systems.

This project is centered on realizing a study of the state of the open data linked in the field of the bibliographic references and contribute to provide a global vision of its current state and probable future development. This study will be complemented with an example of a real utilization of this technology developing an API that uses LOD concepts to extract the bibliography of an author (or conjoint of them) of the web of data with queries in SPARQL, and creating and small application that uses the API to show a real case of use.

Paraules clau (entre 4 i 8):

Dades enllaçades obertes, web semàntica, SPARQL, RDF, XML, LOD, Open Linked Data, Bibliografia

ÍNDEX

ÍNDEX.....	1
LLISTA DE FIGURES	4
LLISTA DE TAULES	4
1. INTRODUCCIÓ	4
1.1. CONTEXT I JUSTIFICACIÓ DEL TREBALL	4
1.2. OBJECTIUS DEL TREBALL	5
1.3. ENFOCAMENT I MÈTODE SEGUIT	6
1.4. PLANIFICACIÓ DEL TREBALL	6
1.4.1. DISTRIBUCIÓ DE TASQUES	6
1.4.2. DIAGRAMA DE GANTT	8
1.5. RESULTATS ESPERATS.....	9
1.6. DESCRIPCIÓ DELS CAPÍTOLS	9
2. FONAMENTS TECNOLÒGICS EN EL MODELATGE DE DADES A LA WEB.....	9
2.1. DEFINICIONS.....	9
2.1.1. WEB SEMÀNTICA.....	10
2.1.2. METADADES	11
2.1.3. ONTOLOGIA	11
2.1.4. OPEN DATA	13
2.1.5. LINKED DATA	13
2.1.6. LINKED OPEN DATA.....	13
2.2. LENGUATGES PER DEFINIR MODELS DE DADES.....	13
2.2.1. XML (eXtensible Markup Language).....	13
2.2.1.1. CARACTERÍSTIQUES DE XML	14
2.2.1.2. ESTRUCTURA D'UN DOCUMENT XML.....	14
2.2.1.3. NAMESPACES	15
2.2.1.4. DOCUMENTS BEN FORMATS I VÀLIDS	15
2.2.1.4.1. DOCUMENTS BEN FORMATS.....	15
2.2.1.4.2. DOCUMENTS VÀLIDS	16
2.2.1.5. LIMITACIONS DE XML.....	16
2.2.2. RDF (Resource Description Framework)	16
2.2.2.1. CARACTERÍSTIQUES DE RDF	16
2.2.2.2. MODEL RDF	17
2.2.2.3 SINTAXI RDF I RDF/XML	18
2.2.2.3.1. SINTAXI SERIALITZADA	19
2.2.2.3.2. SINTAXI ABREUJADA	19
2.2.3. RDF-SCHEMA (RDFS).....	21
2.2.3.1 ESTRUCTURA D'UN RDF-SCHEMA.....	22

2.2.3.1.1. CLASSES	22
2.2.3.1.2. PROPIETATS	22
2.2.3.1.3. RESTRICCIONS.....	22
2.2.3.2 LIMITACIONS DEL RDF-SCHEMA	23
2.2.4. OWL (Ontology Web Language).....	23
3.DADES OBERTES ENLLAÇADES BIBLIOGRÀFIQUES	24
3.1. INTRODUCCIÓ.....	24
3.2. PRINCIPIS DE LOD.....	25
3.3. REPOSITORIS LOD	26
3.4. ESTAT DE LES ARTS DE LES DADES OBERTES ENLLAÇADES BIBLIOGRÀFIQUES	27
3.4.1. INICIATIVES I ESTÀNDARDS BIBLIOGRÀFIQUES	28
3.4.1.1. SCHEMA.ORG	28
3.4.1.2. FUNCTIONAL REQUERIMENTS FOR BIBLIOGRAPHIC RESOURCES (FRBR)	29
3.4.1.3. RESOURCE DESCRIPTION ACCESS (RDA)	30
3.4.2.4. ONLINE INFORMATION EXCHANGE (ONIX)	31
3.4.2.5. DUBLIN CORE METADATA INITIATIVE (DCMI)	31
3.4.2. REPOSITORIS LOD BIBLIOGRÀFICS	32
3.4.2.1. BRITISH LIBRARY (BL).....	32
3.4.2.1.1 Introducció	32
3.4.2.1.2. Estat	32
3.4.2.1.3. Model de dades	33
3.4.2.2. DEUTSCHE NATIONALBIBLIOTHEK (DNB)	33
3.4.2.2.1. Introducció	33
3.4.2.2.2. Estat	34
3.4.2.2.3. Model de dades	34
3.4.2.3. DATOS.BNE.ES.....	35
3.4.2.3.1. Introducció	35
3.4.2.3.2. Estat	35
3.4.2.3.3. Model de dades	36
3.4.2.4. OCLC/WorldCat	36
3.4.2.4.1. Introducció	36
3.4.2.4.2. Estat	37
3.4.2.4.3. Model de dades	37
3.4.2.5. DIGITAL BIBLIOGRAPHY & LIBRARY PROJECT (DBLP).....	38
3.4.2.5.1. Introducció	38
3.4.2.5.2. Estat	38
3.4.2.5.3. Model de dades	39
3.4.2.5.4. Accés a les dades.....	39
3.4.2.6. ALTRES	39

4. SPARQL	40
4.1. INTRODUCCIÓ	40
4.2. CONSULTES	41
4.2.1. PRÒLEG	42
4.2.2. CONSULTES SIMPLES	42
4.2.2.1. SELECT i WHERE	42
4.2.2.2. ORDER BY	43
4.2.2.3. LIMIT i OFFSET	43
4.2.2.4. DISTINCT	43
4.2.2.5. CONSULTES NIADES	43
4.2.3. OPERADORS SPARQL	44
4.2.3.1. FILTER	44
4.2.3.2. UNION	45
4.2.3.3. OPTIONALs	46
4.2.4. ALTRES TIPUS DE CONSULTES	46
4.2.4.1. CONSTRUCT	46
4.2.4.2. ASK	47
4.2.4.3. DESCRIBE	47
4.3. SPARQL ENDPOINTS	48
5. CAS PRÀCTIC	49
5.1. SELECCIÓ DE REPOSITORI DE DADES	49
5.2. ANÀLIS D'ESTRUCTURA DE REPOSITORI	49
5.3. CREACIÓ DE LES CONSULTES SPAQRL	49
5.4. CREACIÓ DE L'API	50
5.5. CREACIÓ DE L'APLICACIÓ	51
5.6. FUNCIONAMENT DE L'APLICACIÓ	52
5.7. FUTURES MILLORES	54
6. CONCLUSIONS	55
7. GLOSSARI	56
8. BIBLIOGRAFIA	58
8.1. CONCEPTES PREVIS	58
8.2. ESTAT DE LES ARTS OPEN LINKED DATA BIBLIOGRÀFICS	58
8.3. SPARQL	59
8.4. JAVASCRIPT PER CONSULTES SPARQL	59
8.5. MATERIAL ASSIGNATURA	59
8.6. TESIS I PROJECTES	59
8.7. LLIBRES I ARTICLES	60
9. ANNEXOS	60
9.1. CODI DE L'API	60
9.2. CODI DE L'APLICACIÓ	60

LLISTA DE FIGURES

<i>Figura 1. Diagrama de Gantt</i>	8
<i>Figura 2 Estructura de capes de la Web Semàntica</i>	10
<i>Figura 3. Model Subjecte-Predicat-Objecte</i>	18
<i>Figura 4. Repositoris LOD</i>	27
<i>Figura 5 Esquema del model FRBR</i>	30
<i>Figura 6. Entitats OCLC/Worldcat</i>	37
<i>Figura 7. Estructura de l'aplicació.</i>	52
<i>Figura 8. Formulari de cerca</i>	53
<i>Figura 9. Missatge d'espera</i>	53
<i>Figura 10. Execució sense resultat</i>	53
<i>Figura 11. Execució amb resultat</i>	54

LLISTA DE TAULES

<i>Taula 1. Distribució de tasques</i>	7
<i>Taula 2 Components d'una ontologia</i>	12
<i>Taula 3. Estructura d'un document XML</i>	14
<i>Taula 4. Parts del pròleg XML</i>	14
<i>Taula 5. Components bàsics d'una instància XML</i>	15
<i>Taula 6. Exemple 1 de sintaxi abreujada</i>	20
<i>Taula 7. Exemple 2 de sintaxi abreujada</i>	20
<i>Taula 8. Exemple 3 de sintaxi abreujada</i>	20
<i>Taula 9. Classes d'un document RDF</i>	22
<i>Taula 10. Propietats a un document RDF</i>	22
<i>Taula 11. Restriccions a un document RDF</i>	23
<i>Taula 12. Tipus de grafs patró</i>	41
<i>Taula 13. Tipus de consultes SPARQL</i>	41
<i>Taula 14. Estructura d'una consulta SPARQL</i>	41
<i>Taula 15. URIs més habituals abreujades a SPARQL</i>	42
<i>Taula 16. Taula de dades exemple SPARQL</i>	42
<i>Taula 17. Funcions de l'API</i>	51

1. INTRODUCCIÓ

1.1. CONTEXT I JUSTIFICACIÓ DEL TREBALL

La Web Semàntica és una extensió de la web actual en què la informació aporta un significat ben definit, permetent així una millor cooperació entre persones, i sobretot entre màquines.

La filosofia de les dades obertes enllaçades i la seva utilització s'estén avui en dia a molts i diferents camps d'acció essent un d'ells el de les referències bibliogràfiques. No només s'utilitzen en la creació i intercanvi de dades de les biblioteques, sinó també en museus i arxius de qualsevol tipus de documentació o obres, audiovisuals o escrites, constituint una extensió i una evolució dels models d'intercanvi i col·laboració tradicionals utilitzats per les biblioteques.

Tradicionalment, els estàndards per l'intercanvi de dades havien estat, per una banda, la norma ISO2709 com a format per d'intercanvi de registres bibliogràfics, i, per l'altra, el Z39.50 com estàndard i protocol de comunicacions obert dirigit a la recerca i recuperació d'informació en bases de dades amb diferent estructura usant una interfície comuna per realitzar la cerca.

Molts catàlegs de biblioteques, i bases de dades bibliogràfiques o de texts complets, ja siguin comercials com a gratuïtes, formen part del que es coneix com la "web invisible", és a dir, que no són aconseguïdes en els resultats de cerca dels cercador. A partir del sorgiment de les dades obertes enllaçades, les dades disponibles a la web invisible comencen a ser visibles, integrades i reutilitzades per altres serveis.

Aquest treball vol fer un repàs sobre l'estat actual de les dades enllaçades bibliogràfiques presentant i analitzant alguns dels repositoris de dades bibliogràfiques actuals més importants i explicar, mitjançant el desenvolupament d'un API de consulta i d'una petita aplicació, com es poden fer recerques de dades bibliogràfiques a sobre un d'aquests repositoris.

La memòria d'aquest treball inclourà una sèrie de definicions i de coneixements previs imprescindibles per a que el lector sigui capaç d'obtenir una base i el fonaments necessaris per a poder entendre tant l'abast de la informació presentada, com ser capaç de seguir el desenvolupament i explicació del cas pràctic presentat.

S'inclourà també en el treball un petit resum dels objectius que es pretenen assolir amb l'elaboració del mateix i s'exposaran les fases i les tasques a través de les quals s'ha dut a terme aquest projecte i la planificació detallada sobre la qual s'ha desenvolupat.

1.2. OBJECTIUS DEL TREBALL

En aquest punt es descriuran els objectius que es pretenen assolir amb l'elaboració d'aquest Treball de Fi de Grau (TFG) i els objectius singulars del projecte.

Durant aquest projecte es tractaran aspectes directament relacionats amb la web semàntica aprofundint sobre el camp de les dades enllaçades, adquirint els coneixements necessaris sobre aquest mètode de publicació de dades estructurades que permet la seva interconnexió i poder fer-les servir d'una forma útil per extreure'n dades d'elles a través d'un llenguatge específic de consulta.

L'objectiu principal d'aquesta proposta és la d'implementar una aplicació que utilitzi la informació de la web de dades i que, mitjançant una sèrie de consultes, permeti identificar la bibliografia completa d'un autor (o conjunt d'ells).

En primer lloc, i després d'analitzar l'estructura dels repositoris de dades en els quals es troba tota la informació necessària per a dur a terme les recerques, s'haurà de crear un API que, a través d'unes consultes en el llenguatge SPARQL, s'encarregui de consultar i extreure la informació requerida dels repositoris.

En segon lloc, i un cop implementada L'API, s'haurà de desenvolupar l'aplicació objecte d'aquest TFG que permetrà a l'usuari fer les recerques desitjades i veure els resultats en pantalla de les mateixes.

A més s'haurà de realitzar un estudi previ sobre el tema de les dades enllaçades (linked data) per tal d'adquirir tots els coneixements necessaris per a dur a terme aquest projecte, i, si possible, poder utilitzar aquesta informació o reaprofitar-ne part en la construcció de l'aplicació.

Així doncs, els objectius propis d'aquest projecte determinat serien els següents:

- Estudiar els conceptes bàsics de la web semàntica i entendre què són les dades enllaçades en l'àmbit de la web semàntica.
- Ser capaç d'analitzar *Open Linked Data* per identificar els repositoris de dades enllaçades a la web de dades que es puguin fer servir que incloguin la informació demandada.
- Conèixer, tant un dels llenguatges d'intercanvi de dades més utilitzats com és RDF, com el llenguatge de consulta SPARQL per a poder definir i realitzar consultes sobre els repositoris de la web de dades representats d'aquesta forma i extreure'n dades.
- Realització d'un cas pràctic utilitzant aquestes tecnologies, dissenyant i implementant la creació d'un API per a realitzar una sèrie de consultes predeterminades així com la construcció d'una aplicació que la utilitzi a través de la qual un usuari pugui obtenir, a partir del nom d'un recercaire, totes les seves publicacions incloent en el resultat: El nom dels autors, el títol i any de la seva publicació i, si s'escau, dades corresponents a la seva edició.

1.3. ENFOCAMENT I MÈTODE SEGUIT

L'abast d'aquest projecte compren des de l'estudi de les dades enllaçades i de les dades obertes enllaçades en l'àmbit de les referències bibliogràfiques per tal d'identificar repositoris d'informació necessaris per al projecte, fins a la creació d'un API de consulta i la implementació d'una aplicació funcional que permeti les recerques de la bibliografia d'un autor.

Per dur-lo a terme es realitzarà l'estudi esmentat centrant la recerca de repositoris exemple en els de tipus bibliogràfic que puguin fer-se servir per a donar resposta a les consultes que ha de formular l'aplicació que s'ha de desenvolupar.

Es triarà després un dels repositoris, s'analitzarà la seva estructura per a després crear les consultes, amb elles l'API i per últim la aplicació que la faci servir per el cas de prova.

1.4 PLANIFICACIÓ DEL TREBALL

A la següent taula es podrà veure el número d'hores que s'han assignat a cada tasca així com les dates planificades d' inici i de fi.

S'ha disposat el seu ordre segons les dependències entre les tasques com es podrà comprovar més endavant en el diagrama GANTT.

S'ha fet una planificació segons la distribució de les tasques que s'ha considerat oportuna i coherent per cada entrega per a poder acabar el projecte a temps.

1.4.1. DISTRIBUCIÓ DE TASQUES

Tasca	Data Inici	Data Fi	Hores
1.1.1 Descarregar documentació	21/09/16	21/09/16	1
1.1.2 Lectura de la documentació	21/09/16	21/09/16	1
1.1.3 Lectura i anàlisi del projecte	21/09/16	21/09/16	1
1.2.1 Recerca d'informació	21/09/16	24/09/16	8
1.2.2 Recopilar bibliogràfica	26/09/16	26/09/16	2
1.3.1. Definició del Pla de treball	26/09/16	01/10/16	9
1.3.2. Temporització del Pla de Treball	01/10/16	05/10/16	6
1.4 Entrega de la PAC1	05/10/16	05/10/16	0
2.1 Estudis dels conceptes d'ontologia	05/10/16	06/10/16	3

2.2.1 XML	06/10/16	08/10/16	4
2.2.2 RDF	10/10/16	12/10/16	5
2.2.3 SPARQL	12/10/16	17/10/16	10
2.2.4 Preparació del entorn de treball	19/10/16	19/10/16	2
2.3 Estudi del linked data	19/10/16	16/11/16	46
2.4 Entrega de la PAC2	16/11/16	16/11/16	0
3.1 Anàlisi dels repositoris de dades	16/11/16	19/11/16	7
3.2 Implementació de les consultes SPARQL	19/11/16	21/11/16	6
3.3 Disseny i implementació de l'API	21/11/16	26/11/16	12
3.4 Elaboració de l'aplicació i proves	26/11/16	07/12/16	20
3.5 Entrega de la PAC3	07/12/16	07/12/16	0
4.1 Síntesi de la informació per Memòria de treball	07/12/11	14/12/16	10
4.2 Elaboració de la Memòria del Treball	14/12/16	26/12/16	20
5.1 Síntesi de la informació per la presentació	26/12/16	31/12/16	7
5.2 Elaboració de la presentació	31/12/16	12/01/17	20
6 Entrega final	12/01/17	12/01/17	0
7.1 Seguiment del correu	22/01/17	25/01/17	16
7.2 Elaboració del document amb les respostes	22/01/17	25/01/17	5
TOTAL HORES DEDICADES	03/03/11	26/06/11	231

Taula 1. Distribució de tasques

1.4.2. DIAGRAMA DE GANTT

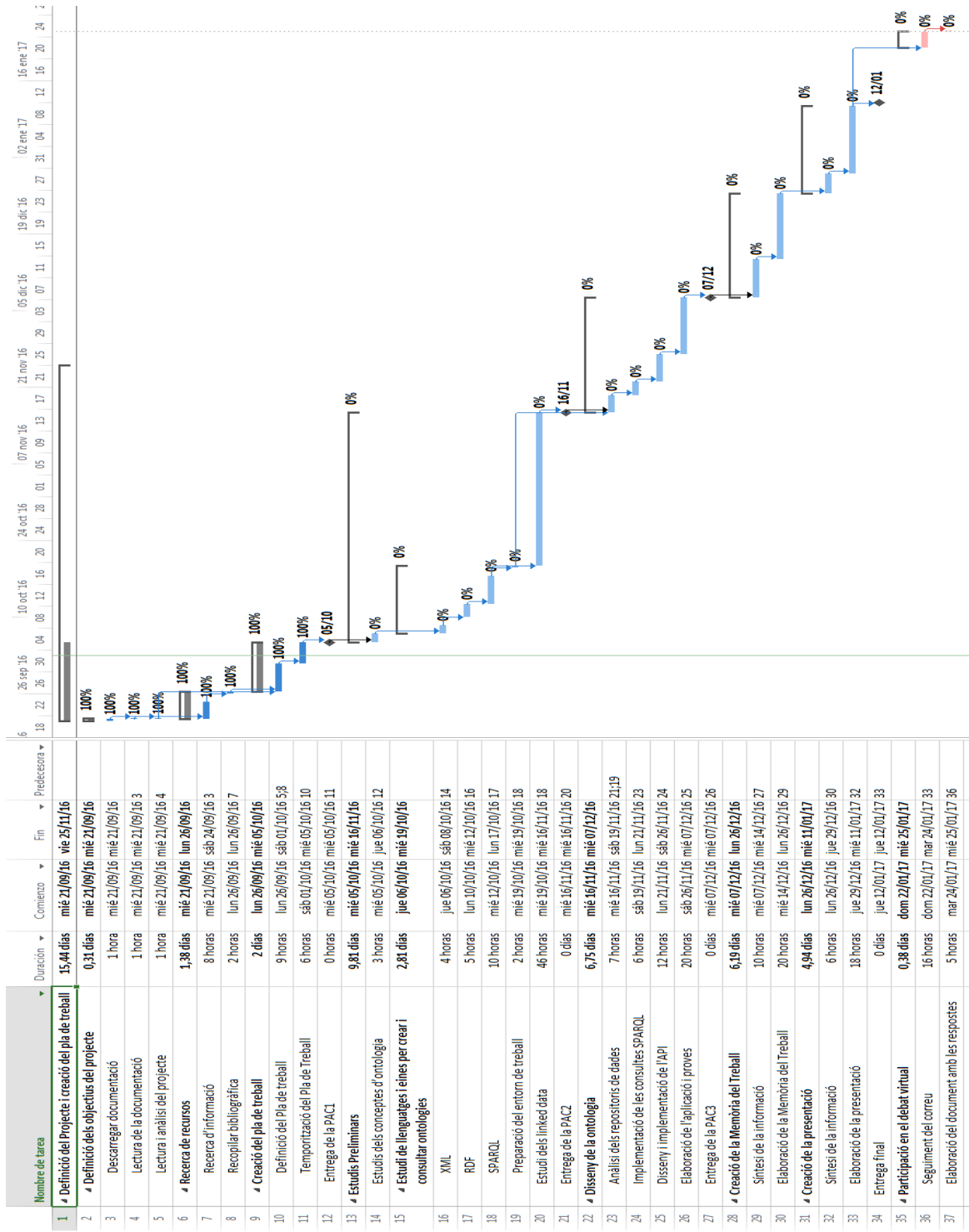


Figura 1. Diagrama de Gantt

1.5. RESULTATS ESPERATS

S'espera obtenir amb la realització d'aquest projecte per una banda una visió de l'estat actual de les dades enllaçades i les eines i recursos que existeixen per la seva utilització, i, per l'altra, la creació d'una aplicació que permeti realitzar recerques predeterminades sobre repositoris de dades publicats.

1.6. DESCRIPCIÓ DELS CAPÍTOLS

A continuació es farà un breu resum dels continguts de cada capítol d'aquesta memòria:

- **INTRODUCCIÓ:** En aquest capítol es presentaran els objectius, el mètode utilitzat, una planificació detallada i els resultats esperats d'aquest treball.
- **FONAMENTS TECNOLÒGICS EN EL MODELATGE DE DADES A LA WEB:** En aquest capítol es presentaran tots els conceptes i els fonaments tecnològics indispensables per al modelatge de dades a la web per tal d'establir una base ferma per entendre el concepte de dades enllaçades i com es creen els repositoris de dades d'aquest tipus.
- **DADES OBERTES ENLLAÇADES BIBLIOGRÀFIQUES:** En aquest capítol es presentaran el concepte i les bases de les dades obertes enllaçades i s'aprofundirà sobre l'estat de les arts dels LOD de tipus bibliogràfics fent un estudi d'algunes de les iniciatives i dels repositoris més destacats.
- **SPARQL:** En aquest capítol es presentarà el SPARQL, llenguatge de consulta més utilitzat per repositoris RDF, i s'explicarà com es poden construir consultes amb ell ja que seran necessàries pel desenvolupament del cas pràctic d'aquest projecte.
- **CAS PRÀCTIC:** En aquest capítol es presentarà un cas pràctic de l'ús i consulta contra un o varis repositoris de dades enllaçades de tipus bibliogràfics a mode d'exemple d'utilització d'aquestes tecnologies.
- **CONCLUSIONS:** En aquest capítol es presentaran les valoracions finals de l'autor d'aquest treball sobre les dades obertes enllaçades bibliogràfiques i sobre el projecte desenvolupat.
- **GLOSSARI:** En aquest capítol es recullen els termes més destacats utilitzats en aquesta memòria.
- **BIBLIOGRAFIA:** Aquest capítol es recull totes referències a la informació i a la documentació consultada per elaborar aquest projecte.
- **ANNEXOS:** Els annexos inclouran el codi de l'API i codi de l'aplicació

2. FONAMENTS TECNOLÒGICS EN EL MODELATGE DE DADES A LA WEB

2.1. DEFINICIONS

Abans d'entrar a exposar la idea de les dades obertes enllaçades d'àmbit bibliogràfic, s'explicaran en aquest capítol una sèrie de conceptes i de nocions imprescindibles que serviran de base per d'assimilar la posterior exposició i per a poder entendre l'abast d'aquest projecte.

2.1.1. WEB SEMÀNTICA

El concepte de Web semàntica va ser introduït al 2001 per Tim Berners-Lee, actual president del consorci W3C¹, amb la finalitat d'aconseguir que les màquines puguin entendre i, per tant utilitzar, els continguts de la Web.

La Web Semàntica proposa, per aconseguir el seu objectiu, descriure els recursos de la web mitjançant unes representacions que siguin comprensibles tan per a les persones com per programes dotant de més significat a la Web, és a dir, més semàntica, es poden obtenir millors solucions a problemes habituals en la recerca d'informació gràcies a l'ús d'una infraestructura comuna, mitjançant la qual compartir, processar o transmetre informació de forma senzilla.

Aquest concepte de web estesa basada en el significat se serveix de llenguatges universals per resoldre els problemes ocasionats per una web sense semàntica en la qual, de vegades, accedir a la informació desitjada es converteix en una tasca complexa i difícil i pretén que els ordinadors puguin comprendre el llenguatge humà de forma natural recuperant els conceptes d'ontologia i metadades amb l'objectiu de reagrupar la informació de manera útil a forma d'una enorme base de dades, a on tot estigui tot descrit en un llenguatge estructurat.

La Web Semàntica implementa els principis definits pel W3C en forma de capes de tecnologies estàndards i web tal com es pot veure en el següent esquema:

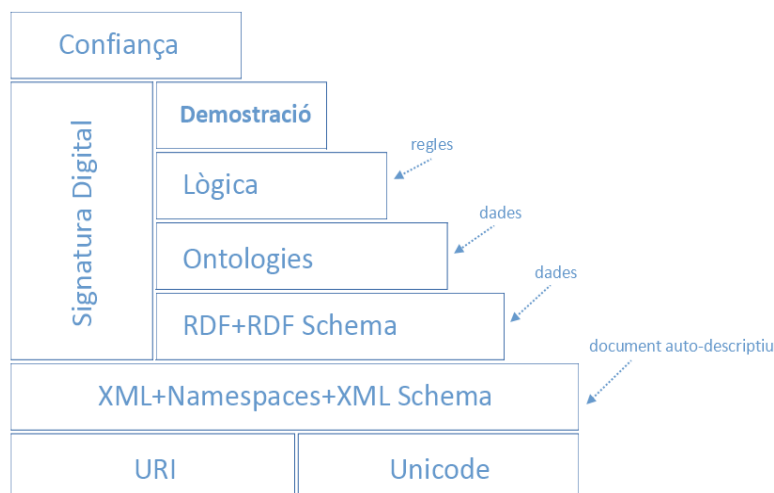


Figura 2 Estructura de capes de la Web Semàntica

- o A les capes inferiors Unicode i URI s'asseguren que es facin servir conjunts de caràcters internacionals i es proporcionin els mitjans per a poder identificar els objectes dins la Web Semàntica.
- o La capa XML inclou els espais de noms i les definicions d'esquemes i assegura la integració de les definicions fetes a la Web Semàntica amb d'altres estàndards basats en el llenguatge XML. *f*
- o La capa RDF i RDF-Schema permet la creació d'axiomes sobre els objectes i la definició de vocabularis que es poden identificar mitjançant una URI. Permet també assignar tipus als recursos i als enllaços. *f*
- o La capa Ontologies suporta l'evolució dels vocabularis definits a la capa inferior i permet la definició de relacions entre diferents conceptes. *f*
- o La capa de Signatura Digital permet detectar les modificacions en els documents i les capes Lògica, Demostració i Confiança estan encara en fase d'estudi. Avui dia només

¹ W3C (World Wide Web Consortium): <https://www.w3.org/Consortium/>

existeixen aplicacions simples que serveixen com a demostradores d'aquestes tecnologies.

D'igual forma que la web actual s'ha construït al voltant d'identificadors URI, del protocol HTTP i del llenguatge HTML, la Web Semàntica es basa en URI, HTTP i el llenguatge RDF. Gràcies a aquests llenguatge s'utilitza una infraestructura basada en metadades que permeten a la Web ser una infraestructura global per a la compartició de dades i documents.

Avui en dia els fonaments de la Web Semàntica s'utilitzen en diferents aspectes del món Web com poden ser la definició de recursos, la definició d'ontologies o tesaures, descriure imatges o agregació i sindicació de contingut. Són molts els projectes en els que s'estan utilitzant aquest conceptes i, al ser un tema en constant evolució, cal anar consultant els canals oficials i consultar activament les pàgines d'organismes com W3C al *Semantic Web Activity Statement*² on es publiquen les especificacions de les tecnologies que sorgeixen entorn de la Web Semàntica.

Els conceptes de metadades i d'ontologia s'explicaran a continuació en altres punts d'aquest mateix capítol de definicions i s'aprofundirà en llenguatge XML i derivats en capítols posteriors.

2.1.2. METADADES

Metadada (meta+dada) etimològicament significa "dades a propòsit de dades". Les metadades són doncs dades estructurades que descriuen informació. Poden indicar qui, què, quan, com, on i perquè, sobre les dades que es documenten.

Al haver-hi una quasi inabordable quantitat de dades contingudes a la web, cal posar una mica d'ordre a tota aquesta informació, i la manera de fer-ho és generant unes metadades addicionals que expliquen el contingut d'una pàgina. Les metadades són transparents per l'usuari però comprensibles pels agents que les utilitzen de forma que permet identificar correctament el tipus de dada en funció de la cerca a realitzar. Transparent vol dir que l'usuari pot veure i accedir a aquestes dades però no n'és conscient del seu àmbit o format real.

2.1.3. ONTOLOGIA

Ontologia és un concepte que té el seu origen a l'antiga filosofia grega però com a terme tècnic, van ser utilitzat primer per Jean Le Clerc al seu tractat "*Ontologia sive de ente in genere*" (1692) i després per Cristian Wolff a la seva obra "*Philosophia prima sive ontologia methodo scientifica pertracta, qua omnes cognitionis humanae principia continentur*" (1730) on la va definir com una "*scientia entis in genere, quatenus ens ens*" (ciència de l'ens en general, en quant que ens) afirmant que utilitza un mètode demostratiu que analitza els predicats de tots els ens com a tals. Totes aquestes definicions l'encasellen dins la branca de la Filosofia anomenada Metafísica, centrada en l'estudi de la naturalesa i organització de la realitat, és a dir el que existeix. Les ontologies en filosofia s'ocupen de l'establiment d'aquelles categories o modes generals de ser que tenen les coses, partint de l'estudi profund de les seves propietats, estructures i sistemes i en com els ens poden ser classificats de diferents formes dins d'unes jerarquies, i subdividits segons les similituds i diferències que presenten.

Des de la segona meitat del segle XX, s'han estat debatent a la comunitat filosòfica els mètodes possibles per a poder construir-les sense obtenir cap resultat satisfactori però al camp de la informàtica si que s'han pogut construir algunes de grans i robustes com Wordnet o Cic. Va ésser realment a principis de la dècada dels 90 quan van començar a proliferar definicions d'ontologia. La més estesa i acceptada, és la de T. R. Gruber (1993) continguda a la seva publicació "*Toward Principles for the Design of Ontologies Used for Knowledge Sharing*", i que es refereix al conjunt de conceptes i relacions entre conceptes que son rellevants en un domini o una aplicació. Gruber defineix una ontologia com una especificació explícita d'una conceptualització. Aquesta definició

² Semantic Web Activity Statement : <https://www.w3.org/2001/sw/Activity>

va esser després clarificada per Borst al 1997 definit una ontologia com una especificació formal d'una conceptualització compartida.

Unificant les dues definicions, es pot dir que una ontologia és una especificació formal i explícita d'una conceptualització compartida sobre la qual es poden inferir una sèrie d'idees clares. Un model abstracte de qualque fenomen del món en el que s'identifiquen els conceptes rellevants i s'especifiquen de forma explícita els seus conceptes i les restriccions existents entre ells, tot utilitzant un llenguatge formal, cosa que implica qualque tipus d'organització teòrica de termes i relacions que puguin esser utilitzats com a eina per l'anàlisi de conceptes d'un domini i pugui esser processat per computadors i no sols per persones. A més, es deixa clar que una ontologia és un coneixement compartit, fruit del consens dintre d'una comunitat.

Gruber afirmà que les ontologies sovint podien ser equiparables a les jerarquies taxonòmiques de les classes, les seves definicions i relacions entre elles i que les no poden limitar-se a aquestes formes, ni poden ser definides amb terminologies conservadores que no afegixen cap coneixement del món, i que, per tant, per especificar una conceptualització, existeix la necessitat de definir axiomes que no restringeixin les possibles interpretacions dels termes definits. A més, la creació d'una ontologia ofereix l'avantatge de que la categorització dels elements i relacions que intervenen en el model de coneixement es fa de forma explícita i, per tant, a més de que el model de coneixement pot esser editat i gestionat també possibilita la seva transmissió de forma que un sistema entengui la conceptualització que utilitzada per l'altre, donant lloc a un coneixement reutilitzable i mantenible.

Des del punt de vista de l'enginyeria, una ontologia és un artefacte construït utilitzant vocabulari específic que descriu una certa realitat i un conjunt de suposicions que tenen en compte el significat del vocabulari. Independentment a aquesta realitat i del vocabulari utilitzat, els components d'una ontologia solen esser sempre els mateixos:

Axiomes	Són elements que permeten el modelat de les asseveracions que es compleixen sempre dintre de la realitat establint així una sèrie de condicions i relacions amb la jerarquia de l'ontologia amb conceptes i atributs definits. Poden ser estructurals i/o no estructurals.
Regles	Són un tipus d'axioma específic que indiquen causalitat. Són del tipus: Si X -> Y i descriuen inferències lògiques.
Classe	Una classe és un conjunt o col·lecció d'objectes. Els objectes d'una classe es poden descriure especificant les propietats que han de satisfer per a pertànyer a aquella classe. Les classes són la base de la descripció del coneixement en les ontologies ja que descriuen els conceptes inclosos al domini. Una classe pot esser dividida en subclasses que representen conceptes més específics que la classe de la qual deriven. Una classe en la qual els seus components són classes es denomina superclasse o metaclasse.
Individus instàncies	Objectes que pertanyen a una classe. No poden ser dividits sense perdre la seva estructura i característiques funcionals.
Propietats, atributs slots	Conjunt d'aspectes, propietats, característiques o paràmetres poden tenir els objectes que en conjunt el descriuen. Les propietats d'una classe són heretades per les seves subclasses i pels individus que pertanyen a ella.
Restriccions	Rangs, especificacions i descripcions formals que han de complir les propietats o atributs perquè siguin entrades vàlides.
Relacions	S'estableixen entre els conceptes d'una ontologia i volen representar la forma en que interaccionen les classes i els individus entre si. Es solen definir generalment com el producte cartesià de n conjunts: R: C1 x C2 x ...x Cn.
Funcions	Tipus concret de relació on s'identifica un element mitjançant el càlcul d'una funció que considera diversos elements d'una ontologia.

Taula 2 Components d'una ontologia

Dissenyar i crear una ontologia es un procés complex i no n'existeix cap d'única per a representar un domini determinat sinó que depèn, en gran mesura, de múltiples variables com el context en que s'ha construït, la persona que l'hagi construït, etc.. Tampoc existeix una metodologia única per a la creació d'ontologies sinó que n'existeixen varies i encara que no s'ha establert cap mètode com el correcte, es coincideix en que es podrien diferenciar 4 fases diferents:

1. Definir les seves classes.
2. Col·locar las classes en un jerarquia de taxonomies (subclasse-superclasse).

3. Definir els atributs de les classes i les seves restriccions.
4. Establir els valors dels atributs amb exemples.

2.1.4. OPEN DATA

Per entendre el què són les dades obertes enllaçades, cal conèixer prèviament el concepte de Open Data (dades obertes).

Open Data és un moviment digital al qual s'estan adherint gradualment molts governs i institucions de tot el món amb l'objectiu de posar les dades que administren a disposició de totes les persones i institucions sense restriccions de copyright o, patents. A més, aquestes dades, com ja es veurà a pròxims capítols d'aquest projecte, es distribueixen en uns formats que permeten la seva fàcil reutilització per a qualsevol fi.

2.1.5. LINKED DATA

De la mà del terme Open Data està el de Linked Data (dades enllaçades o dades vinculades); dos termes relacionats però que no són el mateix. El concepte de Linked Data, és un més conjunt de bones pràctiques per a la publicació de dades que fan referència al fet que les dades estan enllaçades mitjançant tecnologies de web semàntica, en particular RDF³, llenguatge basat en XML i estàndard utilitzat per descriure els recursos web que s'explicarà en més profundament en un capítol posterior. Això no implica, ni que hagin d'esser gratuïtes, ni que hagin de ser obertes.

2.1.6. LINKED OPEN DATA

Un cop assimilades les definicions anteriors i tal i com ja s'explicarà en el seu capítol específic, de la unificació dels dos termes anteriors apareix el terme objecte d'aquest projecte, Linked Open Data o Dades obertes enllaçades, indica que són dades obertes i escrites en RDF. Això significa que l'usuari pot enllaçar dades provinents de diverses fonts, institucions o organitzacions, explorar i combinar aquestes dades de forma lliure i sense restriccions de copyright per a nous projectes de desenvolupaments. Aquest treball es centrarà en les dades enllaçades obertes utilitzades en l'àmbit de desenvolupament de bibliografies i referències de treball o obres escrites.

2.2. LLENGUATGES PER DEFINIR MODELS DE DADES

Al llarg d'aquest capítol es comentaran alguns dels llenguatges habituals empleats per a definir les models de dades i ontologies.

2.2.1. XML (eXtensible Markup Language)

S'introduiran breument en aquest apartat els conceptes principals del llenguatge XML⁴ fent un repàs a la seva estructura i a seva sintaxi ja que XML és la base d'altres llenguatge més estesos com RDF i OWL que si tenen la capacitat de afegir semàntica a les dades. Aquests llenguatges estesos s'hauran d'aprendre per a realitzar aquest projecte i s'abordaran als capítols següents.

El XML, va ser proposat per W3C per a resoldre els problemes d'intercomunicació entre ordinadors, aplicacions i bases de dades, i permet realitzar una estructuració lògica de les dades i dels documents en forma d'arbres d'etiquetes que tenen valor i atributs lo qual proporciona una sintaxi que facilita molt l'intercanvi d'informació entre programes i aplicacions. Aquesta estructura segueix unes regles determinades que permeten crear motors de cerca àgils i eficients que poden accedir ràpidament a les dades contingudes en un document i s'esdevé perfecte per

³ RDF (Resource Description Framework): <https://www.w3.org/RDF/>

⁴ XML: <https://www.w3.org/XML/>

l'intercanvi d'informació entre programes sent des de fa temps un estàndard reconegut utilitzat per la majoria dels fabricants de programari.

2.2.1.1. CARACTERÍSTIQUES DE XML

Es poden resumir les característiques més rellevants del XML en:

- És un dialecte simplificat del llenguatge SGML⁵ que pretén ser més simple i que hereta les seves propietats més importants i desitjables.
- Per ser un metallenguatge, permet la creació d'altres llenguatges de marques i és extensible al no estar predefinides les seves etiquetes i és fàcilment processable per humans i màquines, sent utilitzable amb qualsevol alfabet o llenguatge.
- Ofereix una representació estructurada de les dades fàcilment implementable i distribuïble i garanteix que les dades estructurades siguin uniformes i independents d'aplicacions o fabricants facilitat la interoperabilitat i la integració de dades.
- Codifica dades d'acord a un esquema semàntic i separa informació de presentació i format. A més, està basat en unes marques que compleixen una sèrie de regles lèxiques i normes sintàctiques i s'etiqueten els components del document d'acord amb elles.
- Se li pot associar una gramàtica per a definir un model d'estructura de document DTD-esquema.

2.2.1.2. ESTRUCTURA D'UN DOCUMENT XML

Un document XML té una estructura concreta amb dues parts diferenciades: Pròleg i Instància.

Així mateix, el pròleg consta de 3 seccions: Declaració de l'XML, Declaració del tipus de document i Referència a fulla d'estil externa:

Pròleg	Declaració del XML	<code><?xml version="1.0" encoding="" ISO-8859-1" standalone="no" ?></code>
	Declaració de tipus de document	<code><!DOCTYPE Classe SYSTEM "autors.dtd"</code>
	Fulla d'estil extern	<code><?xml-stylesheet href="autorshtml.xml" type="text/xsl" ?></code>
Instància		<pre> <autors> <autor> <nom>Miquel</nom> <cognoms>Pou Ramirez</cognoms> </autor> <autor> <nom>Joana</nom> <cognoms>Villar Berenguer</cognoms> </autor> </autors> </pre>

Taula 3. Estructura d'un document XML

Quant al pròleg, les 3 seccions defineixen:

Declaració del XML	Indica la versió o especificació de XML utilitzada al document; opcionalment, s'indiquen la codificació i la declaració de document aïllat (indica aquest document fa referència o no a una entitat externa).
Declaració del tipus de document	És opcional i definiria l'estructura i sintaxi que ha de seguir el document que sigui d'aquest tipus si la validació es fa amb un DTD fent referència a ell.
Referència a una fulla d'estil	És opcional i indicaria la forma de representar el document i es pot optar per seleccionar entre dues tecnologies: Fulles d'estil en cascada (CSS) o Fulles de transformacions XSL.

Taula 4. Parts del pròleg XML

⁵ SGML (Standard Generalized Markup Language): https://en.wikipedia.org/wiki/Standard_Generalized_Markup_Language

A la instància del document, en canvi, es troba el contingut del document pròpiament dit. Comença amb una marca inici(<element>) d'un element (element arrel) i acaba amb la marca de fi d'aquest mateix element(</element>). L'arrel del document pot contenir més elements formant una estructura jeràrquica en forma d'arbre contenent. Els seus components bàsics són els elements, els atributs i les dades:

Elements	Constitueixen les unitats bàsiques dels documents i representen els components lògics que el conformen. Cadascun d'ells es descriu a través d'una etiqueta a on hi consta el nom del tipus d'element i una llista d'atributs[opcional] que indicaran les propietats de l'element. El contingut de l'element entre marques d'inici i fi. A l'exemple, es declara l'element arrel amb l'etiqueta <code>projecte</code> i es situa el seu contingut entre la marca d'inici <code><projecte [llista_tributs]></code> i la marca final <code></projecte></code> .
Atributs	Descriuen les propietats dels elements i formen part de l'etiqueta. Cadascun d'ells es representa a través d'una parella nom-valor en què aquest últim ha d'anar tancat entre cometes com es pot veure al exemple anterior al atribut nom de l'etiqueta del element <code><projecte></code> .
Dades	Els elements del document contenen, a més dels atributs, dades. Aquestes es troben entre les marques d'inici i de fi, i poden ser de diferents tipus, entre els quals hi ha d'altres elements. També poden ésser buits i no tenir-ne. En l'exemple, els continguts de l'etiqueta <code>descripció</code> , <code>nom</code> o <code>cognoms</code> .
Comentaris	Comentaris als documents XML a mode de notes o explicacions que no seran interpretats pels parsers.

Taula 5. Components bàsics d'una instància XML

2.2.1.3. NAMESPACES

Els namespaces (espais del noms) en els elements de XML proporcionen un mètode per a la unicitat, que marca les diferències entre diversos autors d'esquemes que utilitzen els mateixos noms d'elements i que seran bàsics a l'hora de definir altres llenguatges com el RDF i el OWL.

Segons la definició inclosa a la *W3C Namespaces Recommendation*,⁶ un espai de noms XML és una col·lecció de noms d'elements i atributs qualificats mitjançant un URI⁷, i bàsicament s'utilitzen per evitar conflictes de noms quan es combinen dades de diverses fonts XML i n'existeixen dos tipus: els namespaces per defecte i els explícits.

L'espai de nom per defecte es declara només un pic al document XML i després s'aplica a l'element en que ha estat declarat i a tots els elements que jeràrquicament siguin fills seus. Els espais de nom explícits, en canvi, només s'aplicaran a l'element on han estat declarats sense estendre's als seus elements fills. Així es poden combinar elements de diverses fonts.

2.2.1.4. DOCUMENTS BEN FORMATS I VÀLIDS

Als documents XML s'ha de tenir molt en compte dues característiques que són molt importants: Estar ben format i ser vàlids. Si es vol comprovar que un document XML compleix aquestes propietats, serà necessari el que s'anomena un parser, que serà *no validador*, si només comprova si el document compleix les especificacions, o *validador*, si a més contrasta el contingut del document amb les regles contingudes al esquema.

Per processar o compartir un document XML entre diferents aplicacions és necessari que estigui ben format i pot ésser interessant també que sigui vàlid si, per exemple, es vol fixar un format estricte per a que es comuniquin dues aplicacions i assegurar que el document que li envii una a l'altra estigui en un format determinat.

2.2.1.4.1. DOCUMENTS BEN FORMATS

Per ser un document XML ben format (*well-formed XML*), el document ha de complir rigorosament una sèrie de normes. El conjunt de totes les normes que ha de complir un document XML venen donades pel W3C en la XML 1.0 Recommendation⁸

6 W3C Namespaces Recommendation: <https://www.w3.org/TR/REC-xml-names/>

7 URI (Uniform Resource Identifier): <https://www.w3.org/wiki/URI>

8 XML 1.0 W3C Recommendation :<https://www.w3.org/TR/xml/>

2.2.1.4.2. DOCUMENTS VÀLIDS

S'ha explicat que per poder-se tractar i intercanviar entre aplicacions, seria convenient que el document XML fos també un document vàlid (*valid XML*) i que ho seria si la seva estructura interna compleix una sèrie de normes prèviament definides en un esquema.

Bàsicament l'esquema per efectuar la validació d'un document XML pot fer-se amb un DTD⁹ o amb un XML-Schema¹⁰ i és també un document de text. En ell es descriu l'estructura que pot tenir el document XML i si l'estructura del document coincideix amb la de l'esquema, es dirà que es tracta d'un document vàlid.

Cal tenir present que tot document vàlid és també un document ben format, però no necessàriament tot document ben format és vàlid.

2.2.1.5. LIMITACIONS DE XML

Els problemes principals que té XML són, tant la falta de flexibilitat, com la d'escalabilitat ja que, en la majoria de casos, l'ordre en que apareixen els elements a un document XML és significatiu i necessari cosa que no és lògica en el món de les metadades, que es poden definir com a les dades sobre les dades i descriuen les propietats i característiques sobre les dades, en que l'important és poder accedir-hi i no l'ordre de la seva presentació, ordre que pot provocar un gran cost de manteniment si el nombre d'elements és elevat. Les gramàtiques XML tampoc permeten validar les estructures de documents XML, ni interpretar de forma general les dades que contenen i per a fer-ho és necessari implementar una aplicació que faci.

A més, té el problema de que els arbres jeràrquics són difícils de combinar entre si i accedir a les dades cosa que pot esser problemàtica si existeixen vàries rutes d'accés a una mateixa dada provocant una baixa escalabilitat en els seus accessos si l'arbre és gran.

Per totes aquestes limitacions va aparèixer RDF que té un model de dades més flexible basat en triplets de subjecte+predicat+objecte en els que els seus elements estan identificats amb URIs que permeten l'accés a la seva semàntica i a la seva interpretació de les dades.

2.2.2. RDF (Resource Description Framework)

El llenguatge RDF és una de les tecnologies claus en el món de la web semàntica i, al igual que moltes d'altres, és una proposta del W3C basat en XML amb la finalitat de crear un format per aconseguir la compatibilitat entre els diferents sistemes de metadades existents. S'explicaran en aquest apartat els conceptes que han de permetre conèixer i comprendre tots els fonaments en que es base el llenguatge: la seva sintaxi, la seva estructura i model de dades que representa.

2.2.2.1. CARACTERÍSTIQUES DE RDF

RDF és un llenguatge que, bàsicament, proporciona un nou model d'estructuració de la informació que facilita poder definir un mecanisme que descriu recursos sense que s'assumeixi cap regla o norma sobre un domini d'aplicació en particular per defecte o que es defineixi prèviament la semàntica de cap domini d'aplicació. A més, posseeix una facilitat per processar metadades que permet la interacció entre diverses aplicacions i proporciona un mecanisme perfecte d'intercanvi d'informació a través de la web, destacant en multitud d'àmbits del entorn web.

Anteriorment s'ha parlat sobre els espais de noms com a una part fonamental de la infraestructura web ja que representen una part crítica de la infraestructura necessària per a desplegar sistemes de metadades modulars web fent possible la seva modularitat i extensibilitat. Això es degut a que qualsevol conjunt d'elements de metadades és un espai de noms limitat per les regles i convencions determinades per la seva pròpia agència de manteniment i a que les declaracions

⁹ DTD (Document Type Declaration) : <https://www.w3.org/TR/xml/#dt-markupdecl>

¹⁰ XML-Schema (Esquema XML): <https://www.w3.org/XML/Schema>

d'aquests espais de noms permetran als dissenyadors d'esquemes de metadades definir el context per a un terme en particular assegurant que la definició sigui única per al terme dintre dels límits de l'espai de noms declarat. Aquesta infraestructura permet als dissenyadors dels sistemes de metadades seleccionar els elements desitjats dins d'un conjunt d'elements de metadades anteriorment establert, evitant així tornar a inventar conjunts de noves metadades per a cada nou domini.

RDF representa doncs, un model de metadades que dirigeix per referència molts dels aspectes de la codificació que requereix l'emmagatzematge i transferència de fitxers. Existeixen molts d'aquests aspectes en que RDF utilitza XML; per això que molt sovint es diu que RDF i XML són complementaris ja que RDF és el model de metadada i solament dirigeix per referència molts dels aspectes de codificació que requereix el seu emmagatzemament i transferència d'arxius. Per aquests aspectes, RDF té com a suport el XML. La sintaxi en XML és solament una de les sintaxis possibles per a que surtin formes alternatives per a representar el mateix model de dades RDF.

XML proporciona a RDF la seva flexibilitat a l'hora de generar nous conjunts d'etiquetes, la seva orientació multiplataforma, i el mecanisme semàntic que el permet expressar la descripció de qualsevol tipus de recurs. És destacable però, que la sintaxi XML és només una de les possibles a utilitzar dintre de RDF, i que, actualment, es pot optar per altres alternatives, com N3, o el seu subconjunt, les N-Triples, o altres formes alternatives que puguin sorgir per a representar un model de dades RDF.

Resumint doncs, es pot dir que el llenguatge RDF especifica una gramàtica lògica que aconseguix que els dissenyadors de planes web puguin descriure les propietats semàntiques dels seus documents en una notació estàndard i comú per qualsevol tipus de metadades, i que per poder-ho fer es basa en tres pilars fonamentals com les metadades, els espais de noms i un llenguatge que permeti proporcionar una estructura a les dades i documents de la web.

Als següents apartats s'explicaran els conceptes bàsics del model a utilitzar per a representar les metadades mitjançant RDF, així com una sintaxi per a codificar i transmetre-les maximitzant la interoperabilitat entre els servidors i clients web desenvolupats independentment.

2.2.2.2. MODEL RDF

La base de RDF és, com s'ha explicat abans, la creació d'un model per a representar recursos, propietats i els valors corresponents a aquestes propietats, incloent-hi també, les relacions existents entre ells i es basa en una nomenclatura neutral i independent que serveix per a representar expressions RDF.

Bàsicament el model de dades bàsic el constitueixen tres tipus conceptes:

- **Recursos (resources):** Es denominen recursos a totes les coses descrites per expressions RDF. Un llibre, una persona, una plana web completa, una part d'una o una col·lecció d'objectes del món, poden ser recursos. Cadascun d'ells que s'utilitzi ha d'estar identificat inequívocament mitjançant un URI (*Uniform Resource Identifier*) que és un mecanisme simple i extensible per identificar un recurs i que consisteix en una cadena de caràcters que permet identificar un recurs independentment del context en què es trobi. Les URLs (Uniform Resource Locator) són un subconjunt de les URIs que, a més d'identificar un recurs, proporcionen els mitjans per localitzar-lo a través d'internet.
- **Propietats (properties):** Qualsevol característica, atribut o relació utilitzat per descriure recursos com a, per exemple, un autor o un títol. A cada cadascuna d'elles se li assigna un valor, defineix els valors permesos, es limita els tipus de recurs a descriure, així com se li determina les seves relacions amb altres propietats.

- **Sentències (statements):** Una combinació formada per un recurs específic, una propietat determinada i el valor assignat a ella. Les sentències RDF són les **tripletes o triples**¹¹ que consten de subjecte, predicat i objecte. El significat de les dades sol expressar-se per a un conjunt d'aquetes sentències utilitzant marques XML. El valor pot ser un literal o un bé un altre recurs. La sentència RDF es sol representar com un graf etiquetat dirigit utilitzant: ovals per a representar els subjectes, rectangles per a representar els objectes i fletxes que van des del subjecte al objecte per a representar els predicats.

El valor d'una propietat pot ser un altre recurs o un simple literal. El diagrama bàsic de sentència és podria llegir com que el "subjecte" té com "predicat" a "objecte", o també: "objecte" és "predicat" del "subjecte" i tots els gràfics segueixen un model comú: la descripció de nodes mitjançant les propietats i personalitzant-los a través dels URI's.

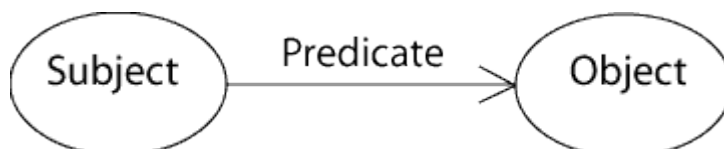


Figura 3. Model Subjecte-Predicat-Objecte

Un etiquetat dels URI's que qualifica als nodes i a les fletxes de manera unívoca és lo que permet la transformació d'aquestes representacions a RDF. El fet de que l'únic vocabulari que s'utilitza sigui el format per URI's permet l'ús d'una mateixa URI com a node i com a fletxa a més de permetre realitzar auto referències. Amb la combinació de les representacions descrites anteriorment es poden representar pràcticament totes les estructures o models de dades que es desitgin.

2.2.2.3 SINTAXI RDF I RDF/XML

Existeix més d'un tipus de sintaxis RDF i, actualment, es poden utilitzar tan l'N3, com el seu subconjunt, les N-Triples i XML. La més utilitzada és una serialització de XML coneguda com RDF/XML que va ésser dissenyada per a poder agrupar més d'una sentència sobre un mateix recurs, ja que generalment una sentència RDF no apareix aïllada, sinó que forma un agrupament de propietats per a qualificar a un determinat recurs.

A continuació es pot veure un exemple de com es plasmarien a un document RDF dues entrades de llibres d'una llibreria:

Títol	Autor	Editorial	ISBN	Preu	Any
EL TEMOR DE UN HOMBRE SABIO	PATRICK ROTHFUSS	DEBOLSILLO	9788499899619	10.90	2013
EL NOMBRE DEL VIENTO	PATRICK ROTHFUSS	DEBOLSILLO	9788499082479	9.90	2011

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:book="http://www.bookstore.fake/book#">
  <rdf:Description
    rdf:about="http://www.bookstore.fake/book/El Temor de un hombre Sabio">
    <book:títol>EL TEMOR DE UN HOMBRE SABIO</book:títol>
    <book:autor>PATRICK ROTHFUSS</book:autor>
    <book:editorial>DEBOLSILLO</cd:editorial>
    <book:ISBN>9788499899619</book:ISBN>
    <book:preu>10.90</book:preu>
    <book:any>2013</book:any>
  </rdf:Description>
  <rdf:Description
    rdf:about="http://www.bookstore.fake/book/El Temor de un hombre Sabio">
    <book:títol>EL NOMBRE DEL VIENTO</book:títol>
    <book:autor>PATRICK ROTHFUSS</book:autor>
```

¹¹ Triples RDF: <https://www.w3.org/TR/2004/REC-rdf-concepts-20040210/#section-triples>

```

<book:editorial>DEBOLSILLO</book:editorial>
<book:ISBN>9788499082479</book:ISBN>
<book:preu>9.90</book:preu>
<book:any>2011</book:any>
</rdf:Description>
</rdf:RDF>

```

RDF a la seves especificacions defineix dues sintaxis XML per a codificar una instància o objecte específic d'una categoria del model de dades . Poden utilitzar-se alhora a un mateix documents perquè els intèrprets RDF implementen les dues.

2.2.2.3.1. SINTAXI SERIALITZADA

La sintaxi serialitzada expressa les capacitats totals del model de dades d'una forma molt regular i estructurada i es en dos elements: *RDF* i *Description* que tenen el següent format segona per a descriure cadascun dels recursos que formen el document RDF. El format d'aquestes instruccions és el següent:

Element RDF

Format: [*<RDF>*] *Description** [*</RDF>*]

Aquest element s'utilitza per a indicar que es tracta d'un document RDF i és l'element arrel del document seguint el format d'un XML. Conté també una referència a l'espai de noms RDF. El contingut pot ser mapejat com una instància del model de dades. Si s'analitza aquesta instrucció del exemple anterior:

```

<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-
ns#
xmlns:book="http://www.bookstore.fake/book#">

```

Es pot veure la definició dels espais de noms, i els seus corresponents prefixes, que s'utilitzaran en el document. En aquest cas els prefixes són "*rdf*" i "*book*", que corresponen a document i esquema RDF respectivament.

Element Description

Format: *<Description Atributabout> Propietats***</Description>*

Aquest element s'utilitza per segona per a descriure cadascun dels recursos que formen el document RDF. A partir d'ells s'aniran definint i agrupant cadascuna de les propietats que han de qualificar a cadascun dels recursos en concret. *Description* pot tenir dos paràmetres que poden tenir orígens molt diversos:

- **Atribut about:** Identifica el recurs
- **Propietats:** A on es poden afegir i descriure les diferents propietats del recurs. Les propietats contingudes poden definir-se com atributs com al exemple anterior o com a recursos en que en lloc de tenir assignat un valor, té una referència a un recurs que conté informació sobre la propietat.

2.2.2.3.2. SINTAXI ABREUJADA

La sintaxi abreujada inclou termes addicionals per tal de proporcionar una forma més compacta d'un subconjunt del model de dades RDF d'una forma molt clara i fàcil d'entendre. Normalment és millor utilitzar una forma XML més compacta i per això es fan servir tres tipus d'abreviacions:

1. Si no existeixen propietats repetides dintre d'un element *Description* i els valors de les propietats són literals, les propietats es poden expressar com atributs XML de l'element *Description*. En aquest cas:

<pre><rdf:Description rdf:about="http://www.bookstore.fake/book/El Temor de un hombre Sabio"> <book:titol>EL TEMOR DE UN HOMBRE SABIO </book:titol> <book:autor>PATRICK ROTHFUSS</book:autor> <book:editorial>DEBOLSILLO</book:editorial> < book:ISBN>9788499899619</book:ISBN> < book:preu>10.90</book:preu> < book:any>2013</book:any> </rdf:Description></pre>	<pre><Description about="http://www.bookstore.fake/book/El Temor de un hombre Sabio" titol="EL TEMOR DE UN HOMBRE SABIO" autor="PATRICK ROTHFUSS" editorial="DEBOLSILLO" ISBN="9788499899619" preu="10.90" any="2013" /></pre>
---	--

Taula 6. Exemple 1 de sintaxi abreujada

- Una segona forma d'abreviacions es pot utilitzar amb sentències específiques a on l'objecte de la declaració és un altre recurs i el valor de qualsevol propietat donada per aquest segon recurs són cadenes de text. Afecta als elements *Description*. En aquest cas, s'aplica una transformació on les propietats de l'element *Description* d'aquest segon recurs poden posar-se com atributs XML de l'element *propertyElt* del primer recurs especificat:

Sintaxi serialitzada	Sintaxi abreujada
<pre><rdf:Description about="http://.../Programa-1"> <rdf:Arxiusresource="http://.../Arxiu-1"/> </rdf:Description> <rdf:Description about="http://.../Arxiu-1"> <rdf:Tipus>=indexat</rdf:Tipus> </rdf:Description></pre>	<pre><rdf:Description about="http://.../Programa-1"> <rdf:Arxiusresource="http://.../Arxiu-1" Tipus="indexat" /> </rdf:Description></pre>

Taula 7. Exemple 2 de sintaxi abreujada

- En el cas que un element *Description* contingui una propietat *type*, el tipus de recurs definit com a valor d'aquesta propietat pot utilitzar-se directament com a nom d'element en lloc de l'element *Description* en què està inclòs:

Sintaxi serialitzada	Sintaxi abreujada
<pre><rdf:Description rdf:about="http://www.bookstore.fake/book/El Temor de un hombre Sabio"> <type resource="http://.../schema/book" /> . . . </rdf:Description></pre>	<pre><rdf:book rdf:about="http://.../schema/book" /> . . . </rdf:book></pre>

Taula 8. Exemple 3 de sintaxi abreujada

Contenidors

Els elements contenidors s'utilitzen per a referir-se normalment a una col·lecció de recursos en lloc de a un recurs en concret, per exemple, és més habitual treballar amb la relació de programes informàtics que integren un inventari, que no treballar amb un programa en concret. Per poder mantenir aquestes llistes de recursos o literals. RDF n'utilitza de tres tipus:

- Bag(rdf:bag)**. Llista desordenada de recursos o literals. S'utilitzen per indicar que una propietat té múltiples valors, es valor poden estar en qualsevol ordre i se n'admeten duplicats:

```
<rdf:Description
rdf:about="http://www.bookstore.fake/autor/PATRICK ROTHFUSS MAXWELL">
  <book:autor>
  <rdf:Bag>
    <rdf:li>MAXWELL</rdf:li>
    <rdf:li>PATRICK</rdf:li>
    <rdf:li>ROTHFUSS</rdf:li>
```



```

</rdf:Bag>
</book:autor>
</rdf:Description>

```

2. **Sequence(rdf:Seq)**. Llista ordenada de recursos o literals. S'utilitza per a declarar que una propietat té múltiples valors però, al contrari que bag, l'ordre sí és significatiu. També admet valors duplicats:

```

<rdf:Description
rdf:about="http://www.bookstore.fake/autor/PATRICK ROTHFUSS MAXWELL">
  <book:autor>
    <rdf:Seq>
      <rdf:li>PATRICK</rdf:li>
      <rdf:li>ROTHFUSS</rdf:li>
      <rdf:li>MAXWELL</rdf:li>
    </rdf:Seq>
  </book:autor>
</rdf:Description>

```

3. **Alternative(rdf:Alt)**. Llista de recursos o literals que representa les possibles alternatives correctes per a un valor individual d'una propietat. En el següent exemple, format, només podria prendre 1 dels valors inclosos:

```

<rdf:Description rdf:about="http://www.bookstore.fake/book/El Temor de un hombre Sabio">
  <book:format>
    <rdf:Alt>
      <rdf:li>Tapa blanda </rdf:li>
      <rdf:li>Tapa dura</rdf:li>
      <rdf:li>Bolsillo</rdf:li>
    </rdf:Alt>
  </book:format>
</rdf:Description>

```

2.2.3. RDF-SCHEMA (RDFS)

El model de dades RDF, com ja s'ha explicat anteriorment, defineix un model simple per descriure els recursos i les relacions entre ells mitjançant la utilització d'unes propietats i uns valors assignats a elles que es podrien entendre com atributs del recurs corresponent-se en aquest sentit al típic parell atribut-valor. A més a més, les propietats RDF poden representar també relacions entre recursos i amb aquesta estructura, el model de dades RDF s'assemblaria molt a un diagrama entitat-relació però sense proporcionar mecanismes per a declarar aquestes propietats, ni tampoc per a definir les relacions entre les propietats i altres recursos. Per a poder fer-ho s'utilitzarà una extensió del RDF com és el RDF-*Schema*¹².

Un RDFS proporciona un marc de treball per a poder definir totes les propietats d'un recurs determinat i tots els tipus de recursos que es descriuran, per això, els esquemes RDF podrien equiparar-se als DTD's o als esquemes XML, però amb la diferència que permeten especificar restriccions en l'estructura d'un document XML i que proporcionen informació sobre la interpretació que s'ha de donar a una sentència en un model de dades concret. A més, igual que l'RDF, també utilitzen l'XML per a la seva codificació d'intercanvi i, per tant, amb ells es podran definir relacions jeràrquiques de classes de recursos, especificant propietats i relacions que s'admeten entre elles. En RDF, les classes, relacions i les pròpies sentències són també recursos i es poden tractar com a tals.

¹² RDF-*Schema*: <https://www.w3.org/TR/rdf-schema/>

2.2.3.1 ESTRUCTURA D'UN RDF-SCHEMA.

Com s'ha esmentat anteriorment, un RDFS és una col·lecció de recursos RDF que es poden utilitzar per descriure les propietats d'altres recursos RDF i que permeten definir vocabularis RDF per aplicacions específiques. L'RDFS seguirà una estructura definida i per especificar el nucli del vocabulari utilitzat en un document RDF, dins el propi document RDF s'indicarà l'espai de noms amb un URI. Per exemple a:

```
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-chemas#"
xml:base="http://www.bookstore.fake/book#">
```

La URI serà "<http://www.bookstore.fake/book>" i, a més, es pot veure a l'especificació que també es defineix un prefix *rdf* per a referir-se a l'espai de noms principal RDF : <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

Igual que s'ha descrit abans, en l'especificació del model i la sintaxi, els recursos poden ser instàncies d'una o més classes, i això s'indicava amb la propietat *type*. També s'ha vist que, en ocasions, les classes es poden organitzar de forma jeràrquica donant lloc a classes filles o subclasses *subClassOf*. El sistema d'escriptura dels RDFS, en realitat, es bastant similar al dels llenguatges de programació orientada a objectes amb la diferència que RDF, en lloc de definir una classe en termes de propietats, utilitza l'RDFS per definir les propietats en termes de classes dels recursos als que s'apliquen i ho fa a través de les restriccions *domain* i de *range*.

2.2.3.1.1. CLASSES

Són classes que formen part del vocabulari en un RDFS, i que cada model RDF que es crea sobre l'espai de noms de l'RDFS inclouen implícitament:

Resource	Totes les coses descrites per expressions RDF s'anomenen recursos, i es poden considerar instàncies de la classe Resource.
Property	Representa el subconjunt de recursos RDF que són propietats.
Class	Al igual que en els llenguatges orientats a objectes, es correspon amb el concepte genèric d'un tipus (Type) o d'una categoria (Category). En el moment que a un esquema es defineix una nova classe, el recurs que representa aquesta classe ha de tenir una propietat <i>type</i> , el valor de la qual és el recurs Class.

Taula 9. Classes d'un document RDF

2.2.3.1.2. PROPIETATS

Són objectes específics d'una categoria de la classe *Property* i proporcionen un mecanisme per a expressar les relacions entre les diferents classes:

type	Indica que un recurs és membre d'una classe i, per tant, té totes les característiques previstes per a un membre d'aquesta classe.
subClassOf	Especifica una relació pare-fill entre classes.
subPropertyOf	És una instància de <i>Property</i> que s'utilitza per a especificar que una propietat és una especialització d'una altra.
seeAlso	Especifica un recurs que podria proporcionar informació addicional sobre el subjecte.
isDefinedBy	És una subpropietat de <i>seeAlso</i> , i indica el recurs que defineix el subjecte.

Taula 10. Propietats a un document RDF

2.2.3.1.3. RESTRICCIONS

En l'RDFS, les restriccions s'utilitzen per a poder descriure limitacions classes i propietats. La nomenclatura de les restriccions més habituals és:

ConstraintResource	Defineix una subclasse de <i>Resource</i> , les instàncies de la qual són construccions de l'esquema RDF implicades en l'expressió de les restriccions.
---------------------------	---

ConstraintProperty	Defineix una subclasse de <i>Property</i> a on totes les instàncies són propietats utilitzades per especificar restriccions.
range	Les seves restriccions només s'apliquen a les propietats a la qual se li defineixen un rang de valors que ha de complir. Només pot haver 1 restricció de rang.
domain	S'utilitza per indicar la classe. Els seus membres poden utilitzar una propietat que pot tenir cap, una, o més d'una classe com a domini. Si no existeix una restricció de domini, pot utilitzar-se amb qualsevol recurs.

Taula 11. Restriccions a un document RDF

2.2.3.2 LIMITACIONS DEL RDF-SCHEMA

Una de les limitacions d'RDFS és la seva nula capacitat per a poder descriure semàntiques de conceptes o relacions més allà de les li proveu els mecanismes de l'herència i disposa solament de les primitives bàsiques per definir ontologies, cosa per la qual és un llenguatge dèbil per a representar de qualsevol sistema basat en coneixement. Si bé té suport per al tractament de conceptes ontològics i relacions, no en té per als axiomes que són un ingredient clau en la definició d'una ontologia ja que RDFS no va ser concebut per ser una resposta definitiva en la representació del coneixement en un domini particular sinó que es va pensar per a ser també per a ser un nucli ampliable sobre el que basar-se.

2.2.4. OWL (Ontology Web Language)

Com que no s'ha de construir cap ontologia durant el desenvolupament d'aquest projecte, només s'explicaran les nocions bàsiques d'aquest llenguatge utilitzat a la Web Semàntica per a definir ontologies.

OWL¹³ va ser definit directament entre diverses comunitats i, també, és una recomanació del W3C per a poder oferir una integració i interoperabilitat de les dades molt més completa que les existents i així poder les limitacions d'expressivitat que tenen RDF/RDFS. És compatible, com s'ha dit, amb la Web Semàntica, ja que es va pensar per ser utilitzat quan la informació continguda en els documents necessita ser processada per aplicacions i es pot utilitzar per representar explícitament el significat de termes en vocabularis i les seves interrelacions, és a dir ontologies. OWL té major capacitat per expressar significat i semàntica que XML, RDF, o RDFS i es construeix sobre els dos darrers però, encara que es basa URIs per a definir els noms dels recursos, millora la seva capacitat per expressiva al afegir més vocabulari per descriure propietats i classes. Entre altres, pot descriure: relacions entre classes, cardinalitat, igualtat entre les classes o instàncies, nous tipus de propietats, característiques per a les propietats i classes enumerades.

Aquestes noves característiques suposen també una limitació de la capacitat dels raonadors per a inferir nou coneixement no garantint que el raonador pugui trobar totes les conclusions vàlides (completitud computacional), ni que pugui trobar la resposta adequada per a qualsevol entrada en un temps finit(decidibilitat computacional) i, per això, es van crear tres subllenguatges variants a partir de OWL i cadascuna d'elles presenta un diferent grau en la seva relació expressivitat/decidibilitat i cadascun d'ells està pensat per esser utilitzat per diverses comunitats de desenvolupadors i usuaris. De menor a major capacitat expressivitat, que implica de major a menor eficiència de raonament es tenen: OWL Lite, OWL DL i OWL Full. Cadascun d'ells és una extensió del seu predecessor i garanteix el tractament de qualsevol ontologia o estructura creada correctament per ell.

L'estructura d'un document OWL estarà formada també per una capçalera o pròleg opcional (més d'una o cap capçalera) que inclou una sèrie de metadades que es poden fer servir per identificar l'ontologia, i per una sèrie de classes, propietats i restriccions, que són les que serviran per a declarar-la.

13 OWL(Ontology Web Language): <https://www.w3.org/OWL/>

Les classes seran bàsiques i proporcionen un mecanisme d'abstracció per a poder agrupar recursos que tinguin característiques similars i, com passava a RDF, cada classe OWL té associat un conjunt d'elements individuals que en formen l'extensió. En OWL Lite i en OWL DL, els elements no poden ser al mateix temps d'un altra classe però en OWL Full no s'aplica aquesta restricció.

Totes les classes son subclasses de la subclasse principal *Thing* i poden descriure de diferents formes :

- Amb un identificador o URI.
- Amb una enumeració de paràmetres individuals que formen la instància de la classe.
- Amb restriccions. Que permeten filtrar les condicions que han de complir els elements individuals de la classe.
- Amb la intersecció de dos o més descripcions de classes ja existents.
- Amb la unió de dos o més descripcions de classes ja existents.
- Amb el complement d'una descripció de classe. Que permet descriure una classe amb els elements individuals que no formen part d'una altra classe ja existent.

Si mitjançant RDF es poden crear pàgines de la nova Web Semàntica expressant-ne el seu significat, a través dels llenguatges RDF Schema i OWL es poden definir les ontologies.

Un cop explicades les bases de la Web Semàntica, en el capítol següent es farà una aproximació a com es poden realitzar consultes sobre aquesta nova web de tal forma que realment millorin les recerques sobre la web actual utilitzant el llenguatge de consultes SPARQL és un dels llenguatges per a les bases de dades que guarden dades RDF.

3.DADES OBERTES ENLLAÇADES BIBLIOGRÀFIQUES

3.1. INTRODUCCIÓ

Les Dades Obertes Enllaçades (LOD), és la forma que té la Web Semàntica de vincular les diferents dades que estan distribuïdes a la Web, referenciant-les de la mateixa forma que ho fan els enllaços de les pàgines web. No es tracta únicament de la publicació de dades a la Web, sinó que aquests es puguin vincular les unes amb les altres, de forma persones i màquines puguin explorar la web de les dades i arribin a informació relacionada que es fa referència des d'altres dades inicials.

De la mateixa forma que la web de la Web tradicional, la Web de les dades es construeix mitjançant documents a la web. No obstant això, i a diferència de la Web tradicional, on els enllaços són relacions entre punts dels documents escrits en HTML, les dades enllacen coses arbitràries que es descriuen en RDF.

Per exemple, suposant que un directori d'empreses publica informació especialitzada relativa a les organitzacions, com la seva grandària o àrea professional, és possible que desitgi indicar també informació sobre la localització. Ja que a la web existeixen llocs amb grans bases de dades geogràfiques, amb informació detallada sobre les localitzacions, el directori d'empreses pot fer referència a les dades geogràfiques que estan disposats per aquesta font externa. D'aquesta forma, les dades inicials de l'organització s'enriqueixen amb informació que ofereixen els experts en l'àmbit geogràfic.

En informàtica, les dades enllaçades descriu un mètode de publicació de dades estructurades perquè puguin ser interconnectades entre elles i resultin més útils. Es basa en URIs, HTTP¹⁴ i RDF i derivats.

El Linked Data tracta de connectar informació procedent de diverses fonts en un mateix espai web i es va pensar per la comunitat per tal d'aconseguir una Web més pràctica publicant, compartint i connectant dades a través de web semàntica mitjançant l'ús de URIs i RDF fent que aquestes dades publicades estiguin pensades per a la seva lectura, interpretació i processament automàtic per part de programes informàtics.

La Open Knowledge Foundation és una organització que es dedica a promoure la creació, distribució i aplicació de coneixement obert en l'era digital. Posseeix diversos grups de treball i entre els que es poden esmentar: el Open GLAM (Galleries, Libraries Arxivis and Museaums) que té com a objectiu fomentar l'accés al patrimoni cultural digital de galeries, biblioteques, arxius i museus i el Open Bibliography amb l'objectiu de proveir informació, documentació i solucions respecte dades bibliogràfiques obertes o d'altres, encara en formació, com el Open Textbooks i Open Legislation.

3.2. PRINCIPIS DE LOD

El LOD (Linked Open Data) és un moviment que es va fundar el gener del 2007 pel SWEO (*W3C Semantic Web Education and Outreach Group*)¹⁵ amb el principal objectiu de que les dades estiguin en obert i a l'abast de tothom.

Tim Berners-Lee va ser la persona que va encunyar el terme de Open Linked Data en 2006 a on va dir que les dades obertes enllaçades eren bàsiques per connectar la Web Semàntica i, primer en una presentació al respecte al congrés TED del 2009 i després en 2010 en una revisió del document presentat 4 anys enrere, va suggerir un esquema per categoritzar les dades obertes per verificar si eren dades "5 estrelles". Les categories són incrementals i inclouen les característiques de les dades de la categoria anterior:

Categoria	Propietats de les dades
*	Han d'estar disponibles al Web amb una llicència oberta (El format és irrellevant).
**	Han d'estar disponibles en un format estructurat llegible per una màquina.
***	Han d'estar emmagatzemats en un format no propietari.
****	S'han d'utilitzar URIs per identificar les coses per a poder referenciar les dades.
*****	S'han de vincular amb dades d'altres persones per a proporcionar un context..

Les regles definides o principis per a la publicació de dades enllaçades segons Tim Bernes-Lee són quatre:

1. S'han d'utilitzar una direcció única (URI) per identificar el tipus d'objecte o concepte (el recurs) publicat al Web.

Significa que s'hauria d'intentar exposar tanta informació com sigui possible utilitzant URIs i que s'hauria de fer no només per pàgines Web, sinó també per documents de sol·licituds d'atenció al públic, dades i metadades de bases de dades, dades personals, transaccional logs, polítiques i normes empresarials, i fins i tot serveis. Si compartir qualsevol tipus d'informació pot resultar útil, s'hauria d'assignar-li una URIs i fer-ho.

Amb relació a la seguretat, no s'ha de tenir por ni fa falta confiar només en les aplicacions tradicionals per a protegir les dades. Cada dia milions de persones utilitzen el Web per a

¹⁴ HTTP (Hypertext Transfer Protocol): <https://www.w3.org/Protocols/>

¹⁵ SWEO (W3C Semantic Web Education and Outreach Group): <https://www.w3.org/blog/SWEO/>

realitzar operacions bancàries, comprar bitllets d'avió o a botigues en línia. El Web ha provat ser segur sempre que es facin cas a les seves bones pràctiques d'ús.

2. Els recursos cal posar-los en línia aprofitant el HTTP per a poder-los consultar i vincular.

Significa que s'haurien de deixar de banda esquemes d'ID obscurs (fins i tot esquemes d'URI) i aferrar-se només al HTTP, que s'ha fet servir tan bé al Web durant tot aquest temps. S'asseguraria així una varietat d'eines i de recursos accessibles molt més amplia.

3. Els recursos cal que proporcionin dades útils i facin servir estàndards com RDF o SPARQL.

Significa que les dades que es proporcionin a la gent que hi accedeixi per URIs hauria d'estar en un format comú adequat per compartir en el Web. XML, com s'ha vist als capítols anteriors, es queda curt i es necessita un llenguatge amb majors capacitat semàntica i que descriu les dades d'una forma més rica que pugui ser processat millor per una màquina. RDF és el format elegit i el que s'utilitza principalment en la comunitat de LOD. Ofereix una alta transparència semàntica i pot ser consultat mitjançant SPARQL.

4. Les dades han d'enllaçar a altres dades en la web (mitjançant el corresponent URI) per tal de potenciar el descobriment de nova informació.

Significa que s'ha d'intentar compartir la riquesa que ens proporciona aquest mètode. Si els tres primers principis animen a crear punters Web a dades i a maximitzar la utilitat de les dades compartides, una vegada que es tenen no s'hauria de ser tímid utilitzant-los i s'haurien de proporcionar dades enllaçades sempre que es pugui perquè no se sap mai si algú o alguna màquina utilitzarà aquestes dades ni per a què ho farà. L'objectiu de LOD és facilitar la reutilització de les dades amb objectius diferents per les que es van concebre.

3.3. REPOSITORIS LOD

Els primers que van començar a participar en aquest projecte van ser els investigadors de laboratoris de recerca de les universitats i petites empreses. Des de llavors el projecte ha crescut considerablement, i ara compta amb la col·laboració de moltes organitzacions com la BBC¹⁶, Thomson Reuters¹⁷ i la Library of Congress¹⁸.

Aquest augment de la participació es deu a la naturalesa oberta del projecte, i que qualsevol hi pot participar simplement ha de publicar dades acord amb les normes del Linked Data i interrelacionar-los amb grups de dades ja existents.

Hi ha moltes i diverses webs i repositoris formats a partir de LOD, com Wikipedia(DBpedia), Wikibooks, Music GeoNames, Brainz, WordNet, DBLP Bibliography i moltes altres publicades sota la llicència Creative Commons o similars. La mateixa comunitat LOD s'encarrega de mantenir un diagrama complet dels repositoris de dades públic disponibles:

¹⁶ BBC (British Broadcasting Corporation): www.bbc.co.uk/

¹⁷ Thomson Reuters: <http://thomsonreuters.com/en.html>

¹⁸ Library of Congress(LC): <https://www.loc.gov>

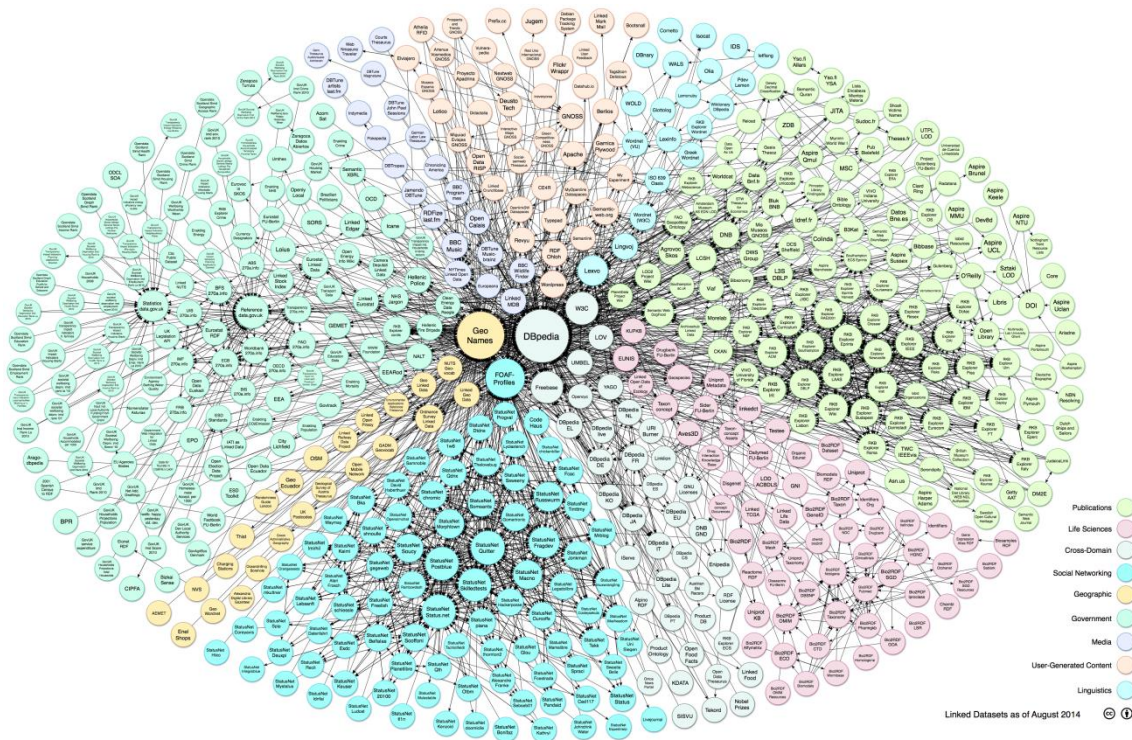


Figura 4. Repositoris LOD

Alguns d'interessants són:

- Project Gutenberg¹⁹: una iniciativa per aglutinar texts sense copyright .
- FOAF²⁰, una aproximació RDF al social networking.
- DBPedia²¹, el LOD al voltant dels articles de la Wikipedia.
- WorldCat²², És la llibreria en línia més gran del món de continguts i serveis.
- DBLP²³, Informació bibliogràfica d'articles científics.

A la figura 3 només es mostren els repositoris que han estat publicats respectant el format de les Dades Enllaçades i, encara que la versió més actual del diagrama és del 2014 i només que inclou els que contenen més de 1000 triples, ja es pot inferir que, avui en dia, el nombre de repositoris que la compondrien és molt superior i es va incrementant diàriament. La mida de cada bombolla és una indicació de la quantitat de dades en aquell conjunt de dada.

Malgrat el seu gran èxit, encara es pot fer molt augmentar aquest nombre i convertir el Web en una web de dades semànticament transparentes. LOD és bàsicament per aconseguir-ho.

3.4. ESTAT DE LES ARTS DE LES DADES OBERTES ENLLAÇADES BIBLIOGRÀFIQUES

Com ja s'ha explicat, les dades obertes enllaçades s'utilitzen per a la creació i intercanvi de dades en molts camps de recerca i un d'ells és el de les referències bibliogràfiques de llibres, revistes, articles i, en definitiva, qualsevol tipus de treball o d'obra escrita a biblioteques, museus,

19 Project Gutenberg:

20 FOAF: <http://www.foaf-project.org/>

21 DBPedia: <http://es.dbpedia.org/>

22 WorldCat: <https://www.worldcat.org/>

23 DBLP: dblp.uni-trier.de/

universitats o arxius traslladant tots els avantatges que proporcionen les LOD a les persones i institucions a aquest àmbit particular.

Tradicionalment, els estàndards per l'intercanvi de dades havien estat, per una banda, la norma ISO2709 com a format per a l'intercanvi de registres bibliogràfics, i per l'altra, el Z39.50 com estàndard i protocol de comunicacions obert dirigit a la recerca i recuperació d'informació en bases de dades amb diferent estructura usant una interfície comuna per realitzar la cerca.

Molts catàlegs de biblioteques, i bases de dades bibliogràfiques o bases de dades de texts complets, ja siguin comercials com a gratuïtes, formen part del que es coneix com la "web invisible", és a dir, que no són aconseguïdes en els resultats de cerca dels cercadors.

A partir del sorgiment de les dades obertes enllaçades, les dades disponibles a la web invisible comencen a ser visibles, integrats i reutilitzats per altres serveis.

El *Library Linked Data Incubator Group*²⁴, grup de treball definit sota el W3C, va manifestar la importància que la comunitat de bibliotecaris comenci a participar en la web semàntica no només com a usuària sinó com a productora de dades i considera les biblioteques com un proveïdor de continguts amb gran potencial per la seva gran qualitat d'informació bibliogràfica de tipus referencial, que es genera mitjançant normatives consolidades que assegurin la seva qualitat, capacitat de compartir i sostenibilitat. A més, com treballen amb vocabularis controlats i amb fitxers normalitzats d'autoritats, nom de persona, entitats, etc., resulta de gran utilitat per enriquir altres dades i per afavorir el correcte desenvolupament de la web semàntica.

En l'actualitat, el sector de la documentació treballa amb normalitat amb protocols estandarditzats per gestionar els seus fons, per la qual cosa a la Comunitat li resulta fàcil participar amb els seus continguts en la web semàntica a través de l'ús de dades enllaçades per publicar les seves dades.

Les dades obertes enllaçades constitueixen un nou desafiament en aquest camp i al llarg d'aquests capítols es farà un recull, primer d'algunes iniciatives LOD en l'àmbit bibliogràfic i després d'alguns dels projectes i repositoris bibliogràfics més importants analitzant-los.

3.4.1. INICIATIVES I ESTÀNDARDS BIBLIOGRÀFIQUES

3.4.1.1. SCHEMA.ORG

Schema.org²⁵ no és cap repositori orientat a les bibliografies sinó que és una iniciativa impulsada en 2011 per Google, Yahoo i Microsoft Bing per proporcionar als webmasters una forma d'afegir etiquetes de metadades a les seves pàgines amb l'objectiu d'ajudar als motors de recerca a comprendre millors els llocs web amb l'objectiu que els grans cercadors arribessin a un consens entorn d'un d'ells per estalviar treball als webmasters i millorar la consistència de dades entre els diferents motors.

L'explicació d'aquesta iniciativa és bàsica per explicar altres repositoris bibliogràfics com, per exemple, OCLC/Worldcat i és que moltes llibreries o ens amb repositoris bibliogràfics utilitzen Schema.org per a representar o etiquetar les obres catalogades.

Aquest marcat estructurat de dades pot realitzar-se a través de diverses tecnologies: microdades, microformats i RDFa perquè, en paraules de Google, "*qualsevol dels formats de marcat té els seus avantatges, però s'ha triat el marcat amb microdades a causa que comparteix la senzillesa dels microformats i la capacitat d'ampliació de RDFa*".

²⁴ Library Linked Data Incubator Group: <https://www.w3.org/2005/Incubator/ld/>

²⁵ Schema.org: <http://schema.org/>

Schema.org proporciona un centenar d'esquemes diferents d'etiquetatge semàntic, que permet utilitzar microdades específiques per a cada tipus de contingut (pel·lícules, llibres, persones, restaurants, esdeveniments, llocs, reviews, etc.) i qualsevol pot crear les seves microdades fent ús del Schema Creator, i verificar la seva validesa amb l'eina de Google Structured Data Testing Tool²⁶.

L'ús de les microdades de Schema.org genera una sèrie d'avantatges com són la millora la indexació del lloc web o permetre que cercadors i algunes xarxes socials ressaltant el seu contingut mitjançant fragments enriquits i encara que les perspectives de negoci de les tres grans potències en matèria de recerca són uns altres, l'impacta potencial d'utilitzar aquestes microdades per a qualsevol projecte de catalogació bibliogràfica. Tota la informació referent a Schema.org es pot trobar a la seva web²⁷.

3.4.1.2. FUNCTIONAL REQUERIMENTS FOR BIBLIOGRAPHIC RESOURCES (FRBR)

El FRBR²⁸ és un model entitat-relació conceptual desenvolupat per la International Federation of Library Associations and Institutions (IFLA²⁹) que s'enfoca en la relació de les tasques dels usuaris amb les dades dels registres bibliogràfics i la recuperació i accés aquestes.

Representa una aproximació a la recuperació i l'accés derivat de les relacions entre les entitats formen enllaços entre elles que permeten a navegar dins una jerarquia de relacions. El model és significatiu perquè és independent dels estàndards catalogràfics específics, com ara les AACR2 o les ISBD³⁰ i s'enfoca en 4 tasques bàsiques que desenvolupa l'usuari i les hi dona suport:

- Cercar (Per trobar entitats segons els criteris de recerca declarats per l'usuari).
- Identificar (Per identificar entitats).
- Selecció (Per seleccionar una entitat que sigui apropiada a les necessitats de l'usuari).
- Obtenir (Per adquirir o obtenir l'accés a l'entitat descrita).

Aquest model entitat-relació, com qualsevol altre consta d'entitats, atributs i relacions entre entitats. Les entitats definides als FRBR representen els objectes d'interès per als usuaris de les dades bibliogràfiques diferenciades en tres grups:

- Grup 1: Obra, l'expressió, la manifestació i l'ítem representen els productes de activitat intel·lectual o artística.

En aquest model, qualsevol tipus de producció escrita o audiovisual, pot entendre's com una treball (work), en forma genèrica, realitzat mitjançant una expressió concreta, continguda en una manifestació que es manifesta físicament mitjançant un ítem. Explicat amb exemple audiovisual, un exemple d'obra podria ser una sèrie de televisió, el seu títol (expressió) és "Breaking Bad", continguda en capítols (cadascun d'ells seria una manifestació) que estan compostos per un o més fragments (cadascun dels quals seria un ítem).

- Grup 2: Persona i l'entitat corporativa, són responsables de la custòdia de l'activitat intel·lectual o artística del grup 1.
- Grup 3: Les matèries de l'activitat intel·lectual o artística del grup 1 o del grup 2, i inclouen conceptes, objectes, esdeveniments, i llocs.

26 Google Structured Data Testing Tool: <https://search.google.com/structured-data/testing-tool>

27 Documentació de Schema.org: <http://schema.org/docs/documents.html>

28 FRBR: https://en.wikipedia.org/wiki/Functional_Requirements_for_Bibliographic_Records

29 IFLA: <http://ifla.org>

30 ISBD (International Standard Bibliographic Description): <http://www.ifla.org/publications/international-standard-bibliographic-description>

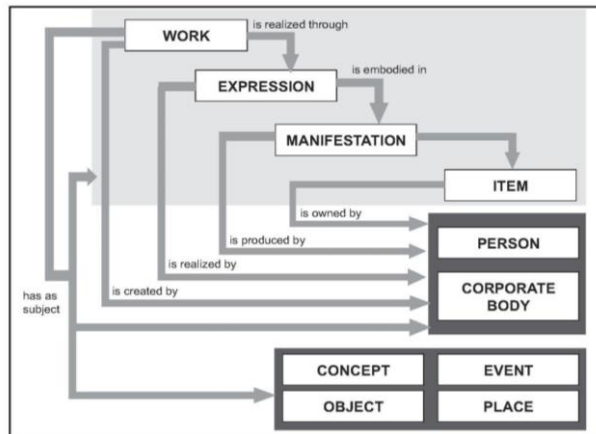


Figura 5 Esquema del model FRBR

Com es pot veure, la síntesi del model utilitzar-se per pràcticament qualsevol obra estigui en el format que estigui. A més FRBR és un model conceptual i no defineix una sintaxi de transferència particular. Sí que existeixen diversos esforços per definir i proposar XML-Schemas per FRBR incloent-hi el Variations/FRBR Project³¹ de la Indiana University

3.4.1.3. RESOURCE DESCRIPTION ACCESS (RDA)

RDA és un estàndard de catalogació que proporciona les directrius sobre la formulació de registres bibliogràfics. Es va publicar conjuntament per la *American Library Association*³²(ALA), la *Canadian Library Association*³³(CLA) i el *Chartered Institute of Library and Information Professionals*³⁴ (CILIP) del Regne Unit i el seu manteniment del *Joint Steering Committee for Development of RDA*³⁵ (JSC) que està compost per representants de ALA, CILIP, *Deutsche Nationalbibliothek* (DN), el *Canadian Committee of Cataloguing*³⁶i la *Library of Congress* (LC).

RDA va ser llançat al juny 2010 però no va ser fins al 2012 que el *U.S Test Coordinating Committee* va recomanar la seva adopció per a les 3 grans biblioteques dels U.S.A. (LC, NAL i NLM) sent implementat completament a finals del març del 2013 per la LC i després adoptat per moltes altres biblioteques nacionals incloent-hi les esmentades, més la *British Library*(BL), *Library and Archives Canada*³⁷, *National Library of Australia*³⁸, *National Library of New Zealand*³⁹ i la DN.

El model de dades de RDA és construït sobre el FRBR però sense prescriure un mètode particular de codificar o mostrar les dades i apareix també com un repositori de dades al Web, a on cada element o valor associat a un element està vinculat a un identificador permanent, la qual cosa permet l'expressió en dades RDF creats amb RDA.

Els vocabularis de RDA (elements i valors) es van publicar en el *Open Metadata Registry*⁴⁰ a l'agost del 2011. La creació i publicació de vocabularis controlats de RDA permeten als dissenyadors d'aplicacions accedir a les dades obertes enllaçades.

31 Variations/FRBR Project: <http://www.dlib.indiana.edu/projects/vfrbr/>

32 American Library Association: <http://www.ala.org/>

33 Canadian Library Association: <http://www.cla.ca/>

34 Chartered Institute of Library and Information Professionals: www.cilip.org.uk/

35 Joint Steering Committee for Development of RDA: www.rda-jsc.org/rda.html

36 Canadian Committee of Cataloguing: <http://www.bac-lac.gc.ca/eng/services/cataloguing-metadata/Pages/canadian-committee-cataloguing.aspx>

37 Library and Archives Canada: <http://www.bac-lac.gc.ca/ENG>

38 National Library of Australia: <https://www.nla.gov.au/>

39 National Library of New Zealand: <http://www.natlib.govt.nz/>

40 Open Metadata Registry: <http://onlinelibrary.wiley.com/doi/10.1002/bult.2011.1720370410/pdf>

El JSC s'ha compromès a publicar i mantenir el contingut dels vocabularis RDA, la sincronització amb el text de RDA, amb la finalitat de recolzar el seu ús pels desenvolupadors d'aplicacions de Web Semàntica.

3.4.2.4. ONLINE INFORMATION EXCHANGE (ONIX)

ONIX⁴¹ és una família d'estàndards que, encara que al principi era només un de sol per compartir les dades bibliogràfiques d'un llibre, actualment comprèn un grup de 3 formats de XML que s'utilitzen per a la comunicació detallada de metadades pel comerç electrònic de llibres, e-books, revistes i altres publicacions media i pretenen donar suport a la comunicació ordinador–ordinador entre les diferents parts involucrades en crear, distribuir, llicenciar propietat intel·lectual disponible publicada físicament o digitalment.

L'original es va expandir per abastir també els e-books (ONIX for books), el segon estàndard s'utilitza per publicacions serialitzades (ONIX for Serials) i el tercer estàndard és per llicències de publicacions (ONIX-PL) i es va dissenyar per manejar les llicències sota les quals les utilitzen les biblioteques i altres institucions els recursos digitals.

ONIX desenvolupat i mantingut per EDItEUR⁴² juntament amb la *Book Industry Communication*⁴³(BIC) del Regne Unit i la *Book Industry Study*⁴⁴dels Estats Units. Existeixen grups d'usuaris a molts països del món.

El model, basat en XML, consta d'un esquema que pot estar definit en un DTD, en XSD o en l'estàndard ISO RelaxNG⁴⁵(RNG) associats a un document XML que especifica les dades. Encara que la versió 3.0 és última revisió major de les especificacions d'ONIX es va llançar a l'abril del 2009, la darrera revisió de l'especificació és la 3.0.3, llançada a abril del 2016. Les especificacions completes del model i els esquemes es poden trobar a la pàgina de EDItEUR⁴⁶

3.4.2.5. DUBLIN CORE METADATA INITIATIVE (DCMI)

La Dublin Core Metadata Initiative⁴⁷ és una organització de caràcter privat dedicada, des de finals del segle XX, fonamentalment a la creació d'elements de descripció i metadades aplicables sobretot a continguts digitals i, particularment, accessibles a través d'Internet (encara que no només). D'entre el conjunt d'elements de descripció creats per la DCMI, el que potser és el seu resultat més conegut és el conjunt d'elements de metadades Dublin Core que ha estat adoptada com a norma ISO⁴⁸(ISO 15836:2003) i NISO⁴⁹(NISO Z3985).

En essència, consta d'un conjunt de 15 elements, que el seu objectiu és el de permetre la descripció bàsica de qualsevol recurs (que podria fer-se equivaler a un objecte físic). Aquests 15 elements bàsics (i les seves etiquetes) són:

1. **Títol:** el nom donat a un recurs, habitualment per l'autor. *Etiqueta: DC.Title*
2. **Claus:** els temes del recurs. Típicament, Subject expressarà les claus o frases que descriuen el títol o el contingut del recurs. Es fomentarà l'ús de vocabularis controlats i de sistemes de classificació formals. *Etiqueta: DC.Subject*
3. **Descripció:** una descripció textual del recurs. Pot ser un resum en el cas d'un document o una descripció del contingut en el cas d'un document visual. *Etiqueta: DC.Description*

41 ONIX: <http://www.editeur.org/83/Overview/>

42 EDItEUR: www.editeur.org/

43 Book Industry Communication: <http://www.bic.org.uk/>

44 Book Industry Study: bisg.org/

45 RNG: <http://relaxng.org/>

46 Model i esquemes ONIX: <http://www.editeur.org/93/release-3.0-downloads/#Specification>

47 Dublin Core Metadata Initiative: <http://dublincore.org/>

48 ISO 15836:2003: http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?csnumber=37629

49 NISO Z3985: http://www.niso.org/kst/reports/standards/kfile_download?id%3Astring%3Aiso-8859-1=Z3985-2001.pdf

4. **Font:** seqüència de caràcters usats per identificar unívocament un treball a partir del com prové el recurs actual. *Etiqueta: DC.Source*
5. **Tipus del Recurs:** la categoria del recurs. Per exemple, pàgina personal, romanç, poema, diccionari, etc. *Etiqueta: DC.Type*
6. **Relació:** és un identificador d'un segon recurs i la seva relació amb el recurs actual. Aquest element permet enllaçar els recursos relacionats i les descripcions dels recursos. *Etiqueta: DC.Relation*
7. **Cobertura:** és la característica de cobertura espacial i/o temporal del contingut intel·lectual del recurs. La cobertura espacial es refereix a una regió física, utilitzant per exemple coordenades. La cobertura temporal es refereix al contingut del recurs, no a quan va ser creat (que ja ho trobem en l'element Data). *Etiqueta: DC.Coverage*
8. **Autor o Creador:** la persona o organització responsable de la creació del contingut intel·lectual del recurs. Per exemple, els autors en el cas de documents escrits; artistes, fotògrafs i il·lustradors en el cas de recursos visuals. *Etiqueta: DC.Creator*
9. **Editor:** l'entitat responsable de fer que el recurs es trobi disponible a la xarxa en el seu format actual. *Etiqueta: DC.Publisher*
10. **Altres Col·laboradors:** una persona o organització que hagi tingut una contribució intel·lectual significativa, però que aquesta sigui secundària en comparació de les persones o organitzacions especificades en l'element Creator. (per exemple: editor, il·lustrador i traductor). *Etiqueta: DC.Contributor*
11. **Drets:** són una referència (per exemple, una URL) per a una nota sobre drets d'autor, per a un servei de gestió de drets o per a un servei que donarà informació sobre termes i condicions d'accés a un recurs. *Etiqueta: DC.Rights*
12. **Data:** una data en la qual el recurs es va posar a la disposició de l'usuari en la seva forma actual. Aquesta data no s'ha de confondre amb la qual pertany a l'element Coverage, que estaria associada amb el recurs en la mesura que el contingut intel·lectual està d'alguna manera relacionat amb aquella data. *Etiqueta: DC.Date*
13. **Format:** és el format de dades d'un recurs, usat per identificar el programari i, possiblement, el maquinari que es necessitaria per mostrar el recurs. *Etiqueta: DC.Format*
14. **Identificador del Recurs:** seqüència de caràcters utilitzats per identificar unívocament un recurs. Exemples per a recursos en línia poden ser URLs i URNs. Per a altres recursos poden ser usats altres formats d'identificadors, com per exemple ISBN ("International Standard Book Number"). *Etiqueta: DC.Identifier*
15. **Idioma:** llengua/s del contingut intel·lectual del recurs. *Etiqueta: DC.Language*

Com pot veure's, aquest conjunt de metadades és molt bàsic i molt genèric i poden ser utilitzats per qualsevol persona per a descriure els recursos de qualsevol sistema d'informació. No obstant això, aquesta simplicitat és la que ha afavorit la seva implantació en multitud d'institucions i organitzacions en sector públic i privat. DCMI ha promogut posteriorment ampliacions d'aquest conjunt però dedicades específicament a temes concrets i de vocabularis derivats.

3.4.2. REPOSITORIS LOD BIBLIOGRÀFICS

3.4.2.1. BRITISH LIBRARY (BL)

3.4.2.1.1 Introducció

La British Library⁵⁰ (BL), Biblioteca Nacional Britànica, està desenvolupant des de ja fa temps la British National Bibliography (BNB), la seva bibliografia en format LOD. En aquests moments, només inclou llibres publicats però preveuen estendre el projecte per ampliar la cobertura a publicacions en sèrie, treballs amb múltiples parts, articles o pròximes publicacions. Tenen exemples de com fer recerques en SPARQL i tota la documentació sobre el projecte i el seu model RDF penjats a la web de la BNB⁵¹

3.4.2.1.2. Estat

Com s'ha explicat, actualment és un projecte encara en desenvolupament i, encara que la BL té publicat un repositori amb un Endpoint disponible per poder-li fer consultes però es garanteixin que el repositori sigui estable ni que els seus continguts canviïn sense avis com a part del seu procés de desenvolupament. El contingut es va actualitzant regularment.

⁵⁰ British Library (BL): <http://www.bl.uk/>

⁵¹ British National Bibliography (BNB): <http://bnb.bl.uk/>

3.4.2.1.3. Model de dades

A la British National Bibliography s'inclouen els més de 3,3 milions de títols que els hi van servir de base per dissenyar i implementar aquest repositori. La transcripció gràfica del seu model de dades la tenen enllaçada a la seva web⁵².

Seguint les bones pràctiques de Dades Enllaçades utilitzades per governs, o els negocis modelen les dades reutilitzant tot el esquema descriptiu que sigui possible i es pugui aprofitar. Per a les parts que no es pot cobrir el reaprofitament d'altres esquemes han desenvolupat el British Library Terms que també tenen publicat a l'igual que les dades.

Altres vocabularis utilitzats per a descriure recursos de la BNB són:

- *Bibliographic Ontology (Bibo)*
- *Bio: Un vocabulari de Informació Bibliogràfica.*
- *Dublin Core*
- *International Standard Bibliographic Description (ISBD)*
- *Org. Organization Ontology*
- *SKOS*
- *RDF Schema*
- *OWL*
- *FOAF*
- *The Event Ontology*
- *WGS84 Geo Positioning and the preliminary version of RDA.*

Les dades estan interconnectades amb altres fonts LOD, en particular: VIAF, LCSH, Lexvo, GeoNames (per assenyalar el país de publicació), MARC per països i codis de llengua, Dewey.info i RDF Book Mashup.

Tot el conjunt de la BNB que cobreix tot el material publicat o distribuït en el Regne Unit des de 1950, està actualment disponible en llicència Creative Commons CC0 1.0 Universal Public Domain Dedication license i es troba constituït per:

- Un SPARQL Endpoint⁵³.
- Un DESCRIBE Endpoint⁵⁴.
- El servei de Recerca⁵⁵.
- El repositori complet⁵⁶ amb l'índex de recursos (arxiu de ternes plenes)

3.4.2.2. DEUTSCHE NATIONALBIBLIOTHEK (DNB)

3.4.2.2.1. Introducció

La Deutsche Nationalbibliothek (DNB), Biblioteca Nacional Alemanya, està desenvolupant un servei LOD que, a la llarga, permetrà a la comunitat de web semàntica utilitzar completament totes les dades de la seva bibliografia nacional. El seu objectiu declarat és fer una important contribució a la informació global amb un servei que posa a disposició del públic totes les dades en RDF de la DNB i de la German Union Catalogue of Serials⁵⁷, per el seu possible aprofitament o per a la construcció de qualsevol nou projecte comercial (o no). També inclou les dades del Integrated Authority File (GND) que inclou les dades d'autoritats.

52 Model de dades BNB: <http://www.bl.uk/bibliografic/pdfs/bldatamodelbook.Pdf>

53 SPARQL Endpoint BNB: <http://bnb.data.bl.uk/sparql>

54 DESCRIBE Endpoint BNB: <http://bnb.data.bl.uk/describe>

55 Cercador BNB: <http://bnb.data.bl.uk/search>

56 Repositori complet BNB: <http://thedatahub.org/dataset/bluk-bnb>

57 German Union Catalogue of Serials (ZDB): German Union Catalogue of Serials (ZDB).

Tota la documentació sobre el projecte, el seu model RDF, així com uns exemples i l'explicació de tot el procés de serialització i dels canvis que s'han anat fent estan penjats a la web del projecte⁵⁸.

3.4.2.2.2. Estat

El DNB va començar a publicar les seves dades com LOAD al 2010. El seu servei LOD s'ha anat expandint fins a incloure la bibliogràfica des de gener del 2012. Com a primer pas, les dades bibliogràfiques corresponents a la col·lecció principal del DNB i totes les publicacions en sèrie (revistes, diaris, i sèries) del ZDB ha estat convertits a un format de dades enllaçades.

La darrera actualització en la que el que s'agrupen tots els canvis i dumps de dades són de maig de 2016 i estan llistat a un document⁵⁹ de la seva web. Consta de més de 3,5 milions de triples

3.4.2.2.3. Model de dades

Les dades disponible a través del servei LOD de DNB estan serialitzades en RDF/XML i RDF/Turtle. Les entrades bibliogràfiques estan modelades amb RDF en les que s'utilitzen propietats específiques de cada tipus de recurs modelat (diaris, llibres, col·leccions, articles o sèries). El modelat està basat en els formats interns específics de la DNB i de la ZDB a més d'afegir-hi les propostes del DINI WG KIM, Bibliographic Data group.

L'elecció de vocabularis utilitza per descriure recursos DNB, s'ha fet entre una quantitat amplia dels vocabularis disponibles i reutilització d'altres. Com als vocabularis de les ontologies no sempre es poden trobar els requeriments necessaris per representar profundament les d'estructures de dades que es volen representar. S'han reutilitzat algunes ontologies existents per facilitar conversió ràpida i pragmàtica de les dades bibliogràfica.

Altres vocabularis utilitzats per aconseguir aquest objectiu són les següents:

- *Dublin Core terms*
- *Dublin Core elements*
- *Bibliographic Ontology (Bibo)*
- *Resource Description and Access Vocabularies (RDA)*
- *International Standard Bibliographic Description (ISBD)*
- *RDF Schema*
- *Upper Mapping and Binding Exchange Layer (Umbel)*
- *FOAF*
- *Schema.org: Library extension terms*

Tant el model en qüestió com la forma en que està creat, es poden consultar directament a la seva pàgina web⁶⁰. També tenen disponible la relació completa de totes les classes utilitzades a una llista⁶¹ actualitzada a gener de 2016.

Tot el conjunt de dades de la DNB està actualment disponible en llicència Creative Commons Zero (CC0) i però no compten, com s'explica a la seva web a l'apartat de FAQ, amb un SPARQL Endpoint (ni tenen planificat crear-ne cap) però es troben enllaços descarregables de les dades del repositori de dades de la DNB⁶² i del repositori de dades de la Gemeinsame Normdatei⁶³ (Integrated Authority File).

58 Projecte DNB: http://www.dnb.de/EN/Service/DigitaleDienste/LinkedData/linkedata_node.html

59 Darrera actualització DNB:

http://www.dnb.de/SharedDocs/Downloads/EN/DNB/service/rundschreiben20160405AenderungenRdfXmlExportRelease20162En.pdf?__blob=publicationFile

60 Model de dades DNB: http://www.dnb.de/SharedDocs/Downloads/EN/DNB/service/linkedDataModellierungTitelraten.pdf?__blob=publicationFile

61 Llistat de classes DNB: <http://d-nb.info/standards/elementset/gnd2016-01-12>

62 Dades del repositori DNB: <http://thedatahub.org/dataset/deutsche-nationalbibliografie-dnb>

63 Dades del repositori Gemeinsame Normdatei: <http://thedatahub.org/dataset/dnb-gemeinsame-normdatei>.

3.4.2.3. DATOS.BNE.ES

3.4.2.3.1. Introducció

Datos.bne.es és un projecte experimental que va començar en 2014 de publicació de les dades de la Biblioteca Nacional de España⁶⁴ com a LOD, basat en les tecnologies i estàndards de la Web Semàntica i que ha estat desenvolupat amb el suport tecnològic del *Ontology Engineering Group*⁶⁵ de la Universidad Politécnica de Madrid.

Amb les dades del 2014 en els que ja s'inclouen la majoria de registres bibliogràfics i d'autoritat de la Biblioteca Nacional de España (més de 2,4 milions de registres) a la que es va afegir també una interfície gràfica amb un cercador a de la seva web a on es poden trobar gran part de les dades que un usuari corrent podria trobar en el catàleg tradicional seguint i explorant les relacions sobre les dades.

Actualment s'inclouen 4 grans classes sobre les quals es poden efectuar les recerques:

- Persones: Autors o personatges que apareixen a qualsevol publicació (Quijote, Arthur Conan Doyle, etc...).
- Entitats. Ens corporatius, congressos professionals d'autors de recursos de la biblioteca o de qualsevol estudi (Per exemple: Agencia Española del Medicamento, The Rolling Stones, Convento del Beato Orozco de Madrid, Congreso Interamericano de Filosofía, etc..).
- Obres: Títols d'obres o recursos de qualsevol tipus.
- Temes: Materials sobre les quals tracten els diferents recursos (Guerra Civil, etc...).

La informació presentada, de totes maneres, encara no es completa i presenta algunes excepcions com:

- Dades del fons de premsa i revistes. Sí que hi són els: llibres nous i antics, fotografies, mapes, dibuixos, gravats, partitures o registres sonors.
- No s'inclouen les localitzacions dels objectes encara que a les recerques que possibiliten tenen un enllaç al seu catàleg que si les indica.
- Les recerques permeses són bastant simples, ja que només hi ha recerques per persona, obra o tema. No està encara implementades recerques més complexes com serien per data, lloc d'impressió o editor.

També s'han extret més de 4 milions de registres d'autoritat de persones, entitats, congressos, expressions, noms geogràfics, etc... que han estat codificats d'acord amb el format MARC21 i també s'han agafat dades de les digitalitzacions existents a la Biblioteca Digital Hispànica⁶⁶ i s'enriqueixen aquestes dades amb enllaços a les seves entrades corresponents, si existeixen a altres repositoris com la LC, DN, Biblioteca Nacional de França, Sudoc, Biblioteca Nacional de Suècia, VIAF o ISNI.

Tota la informació referent al projecte es pot consultar la tenen recopilada en un document⁶⁷ penjat a la seva web.

3.4.3.3.2. Estat

Es tracta encara d'una versió beta d'aquesta nova plataforma sobre la qual s'aniran fent millores. Sol·liciten i conviden als usuaris a fer les seves propostes de millora de la plataforma per tal de seguir desenvolupant el projecte que segueix el model de les biblioteques nacionals ja presentades als punts anteriors d'aquest capítol.

64 Biblioteca Nacional de España: <http://www.bne.es/es/Inicio/index.html>

65 Ontology Engineering Group: <http://www.oeg-upm.net/>

66 Biblioteca Digital Hispànica: <http://www.bne.es/es/Catalogos/BibliotecaDigitalHispanica/Inicio/index.html>

67 Projecte datos.bne.es: <http://www.bne.es/webdocs/Inicio/Perfiles/Bibliotecarios/ayudadatosenlazados.pdf>

A part de les restriccions i implementacions no desenvolupades abans presentades, s'ha de tenir en compte que, com en molts d'aquests repositoris, les dades no estan sincronitzades amb el catàleg tradicional sinó que es fan bolcats cada cert temps i les dades poden no estar actualitzades.

Els enllaços cap a altres repositoris s'esperen ampliar també incloent-ne contra, per exemple, la DBpedia.

3.4.2.3.3. Model de dades

Al principi del projecte, per representar els recursos en RDF es van utilitzar ontologies o models consensuats per la comunitat de biblioteques i proposats per la IFLA⁶⁸. Les dades s'han modelat seguint el model FRBR publicat per IFLA al 1999, que s'ha utilitzat com a model conceptual de referència i com a model de dades ja que proveeix d'una descripció completa i ordenada del món bibliogràfic.

Compte amb més de 58 milions de triples RDF i les entitats, relacions i propietats d'aquestes ontologies han estat publicats per IFLA en RDF en el OMR (Open Metadata Registry) sent reconegudes per W3C com a estàndards vàlids de Web Semàntica.

Paral·lelament, i en col·laboració amb IFLA, van traduir a l'espanyol les etiquetes d'aquestes classes i propietats⁶⁹. Per a les descripcions bibliogràfiques pròpiament dites, van seguir el model i atributs propis de ISBD (International Standard for Bibliographic Description), declarats en RDF en http://metadataregistry.org/schemaprop/list/schema_id/25.html.

No obstant això, per a un major control del vocabulari en la descripció dels objectes, per a la nova versió del 2014 de datos.bne.es es va implementar una ontologia pròpia de la BNE connectada amb tots aquests vocabularis. A més de propietats de FRBR i ISBD, s'han utilitzat com a referència d'altres vocabularis, com RDA, Dublin Core, o BIBO, i, en algun cas, s'han descrit propietats úniques dins de l'ontologia. Tota la informació sobre el model de dades⁷⁰ es pot consultar a la seva web.

Totes les seves dades estan presentades amb llicència Creative Commons CCZero a la seva web⁷¹. La resta de recursos i exemples de BNE estan al seu repositori de Datahub.io⁷². També, es poden realitzar consultes a través del seu SPARQL Endpoint⁷³

3.4.2.4. OCLC/WorldCat

3.4.2.4.1. Introducció

WorldCat⁷⁴ és la xarxa més gran del món de continguts i de serveis bibliogràfics. Les llibreries de WorldCat es dediquen a proporcionar accés als seus recursos a través de la Web per a proporcionar a la gent informació per a les seves recerques.

WorldCat.org permet les recerques en col·leccions de milers de biblioteques a voltant del món i creix cada dia gràcies als esforços dels bibliotecaris i usuaris.

Permet, per exemple, la recerca de llibres (llibres impresos, e-books, articles de recerca) a més de vídeos, fotografies, música o, inclús audiobooks d'un autor o cercar pels més populars de

68 IFLA((Federación Internacional de Asociaciones de Bibliotecarios y Bibliotecas): <http://www.ifla.org/ES>

69 Lista de clases i propietats: http://metadataregistry.org/schemaprop/list/schema_id/5.html

70 Model de dades BNE: <http://www.bne.es/es/Inicio/Perfiles/Bibliotecarios/DatosEnlazados/Modelos/>

71 Dades BNE: <http://www.bne.es/es/Inicio/Perfiles/Bibliotecarios/DatosEnlazados/DescargaFicheros/>

72 Repositori Datahub de BNE: <https://datahub.io/es/dataset/datos-bne-es>

73 BNE SPARQL Endpoint: <http://datos.bne.es/sparql>

74 WorldCat: <https://www.worldcat.org/>

cada secció. Tot en diversos llenguatges. Baix autorització justificada també es posen a disposició documents o fotografies rares que normalment no estan obertes al públic.

Deponent de si l'usuari està afiliat a alguna de les biblioteques que incloses a la xarxa es podria consultar també si el material que es vol està prestat i, en cas que sigui possible, reservar-ho o comprar-ho.

Al Juny del 2012, OCLC va incloure a WorldCat un sistema de descriptiu basat en Schema.org de metadades que permet que estiguin disponibles tots els seus materials en format de dades enllaçades. Es sol referenciar aquest projecte com a WorldCat Linked Data⁷⁵, conegut també com OCLC/WorldCat, que consisteix en una relació 1 a 1 entre cada entrada de Worldcat y cadascuna d'aquestes descripcions en dades enllaçades.

Com ja s'ha explicat, els principis dels Linked Data defensa utilitzar URIs com noms per coses - referències a objectes mundials reals i conceptes abstractes, en comptes de documents. Aquests poden incloure coses tangibles com persones i llocs, o coses que són més abstracte, com temes i esdeveniments. WorldCat Linked Data fa servir entitats (Triple RDF) per representar un tipus (o classe) dels objectes.

Les entrades bibliogràfiques que es creen a WorldCat representen ítems trobats en col·leccions al voltant del món i les seves metadades inclouran referències a un nombre vast d'entitats (feines, persones, organitzacions, temes, llocs, i esdeveniments). Es dediquen a extreure les metadades d'aquestes referències i crear descripcions riques dels objectes que representen.

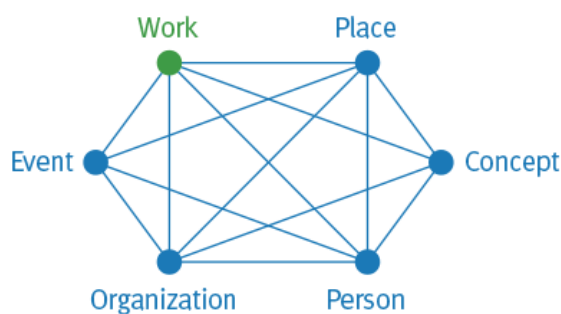


Figura 6. Entitats OCLC/Worldcat

3.4.2.4.2. Estat

OCLC es compromet en què el sistema sigui estable però continua millorant les seves funcionalitats LOD. El seu esforç inicial s'enfoca voltant de la millora, quan sigui possible, del vocabulari de schema.org estenent-ho només en cas de necessitat. Aquesta visió inicial està descrita com experimental i subjecte al canvi si no anés bé.

OCLC només han fet accessibles les dades sobre els treballs i registres bibliogràfiques de Worldcat però, en un futur, tenen previst alliberar les dades d'altres entitats de WorldCat com Persones, Llocs, Organitzacions o Esdeveniments.

3.4.2.4.3. Model de dades

El model dels LOD de OCLC segueix Schema.org amb algunes ampliacions i extensions específiques segons van necessitant. Cada ítem està definit amb RDFa i microformats de dades. A la seva web es pot trobar la llista completa⁷⁶. dels termes utilitzats per WorldCat Linked Data.

Les distribucions de dades enllaçades de WorldCat.org estan disponibles a través de OCLC baix la llicència de Open Data Commons Attribution License (ODC-BY) i es pot accedir a la URI de

⁷⁵ World Linked Data <https://www.oclc.org/developer/develop/linked-data.en.html>

⁷⁶ Llista de termes WorldCat Linked Data: <https://www.oclc.org/developer/develop/linked-data/worldcat-vocabulary.en.html>

qualsevol registre directament. Per defecte la URI es presenta en format HTML però es pot, afegint l'extensió entre parèntesis a la direcció HTML obtenir la representació del objecte en els següents formats:

- Turtle (.ttl)
- N-Triples (.nt)
- JSON-LD (.jsonld)
- RDF/XML (.rdf)
- HTML (per defecte or .html)

Tota la informació de com fer servir i recuperar les dades es troba es pot trobar penjada a la seva web aquí⁷⁷.

3.4.2.5. DIGITAL BIBLIOGRAPHY & LIBRARY PROJECT (DBLP)

3.4.2.5.1. Introducció

Ja en un àmbit científic més especialitat, es troba DBLP és un lloc web d'un enorme repositori bibliogràfic d'articles relacionats amb les ciències de la computació i està allotjat a la Universitat de Trier, Alemanya i és un servei proporcionat actualment de forma conjunta per la pròpia universitat i per *Schloss Dagstuhl Leibniz Center for Informatics*⁷⁸, una ONG alemanya fundada amb l'objectiu de promoure la ciència al món.

Les sigles *DBLP* originalment varen ser creades per a referir-se a *DataBase Systems and Logic Programming* (Sistemes de bases de dades i programació lògica), un servei oferint per un grup de recerca de la universitat de Trier, Alemanya però, però actualment se'l refereix com a "*Digital Bibliography & Library Project*" i, encara que el projecte va començar abastint només l'àrea de sistemes de bases de dades i la lògica de programació, va anar ampliant-se el projecte per tal d'abastir totes les disciplines de les ciències de la computació.

La DBLP publica les seves dades segons els principis LOD i conté les metadades de més 3,3 milions de publicacions escrites per més d'1,7 milió d'autors en milers de publicacions o conferències. La llibreria es nodreix, bàsicament del arxius de VLDB⁷⁹, una revista tracta sobre bases de dades, articles de la IEEE⁸⁰ i de la ACM⁸¹, així com d'articles científics de diferents conferències i augmenta la seva qualitat d'informació en uns 300.000 registres/any.

3.4.2.5.2. Estat

La DBLP és un repositori en constant desenvolupament i modificació tant de les seves dades (Es fa un dump diari de les dades⁸²) com del seu model de dades (DTD amb Darrera actualització 1 Octubre 2016) o de l'API i les funcionalitats que proporciona a l'usuari (humà o digital) per a poder fer recerques sobre el seu repositori (Des de final d'Agost permet fer recerques de forma experimental per full text gràcies a *Semantic Scholar*⁸³ que milloren molt les seves prestacions a les cerques d'informació. Cada cop que es fa un canvi, així com les versions antigues es poden trobar a la seva web.

Actualment existeixen, en 3 mirrors sincronitzats situats a llocs diferents que es sincronitzen entre ells cada pic que les dades canvien. Abans n'existien més d'externs però ara mateix no estan sincronitzats amb el repositori original.

Mirrors:

- Mirror 1 (Trier 1): <http://www.informatik.uni-trier.de/~ley/db/>

77 Informació recuperació de dades WorldCar Linked Data: <https://www.oclc.org/developer/develop/linked-data/linked-data-exploration.en.html>

78 Schloss Dagstuhl Leibniz Center for Informatics: <https://www.dagstuhl.de/en/>

79 VLDB (Very Large Data Bases): <http://www.vldb.org/archives/website.html>

80 IEEE (Institute of Electrical and Electronics Engineers): <https://www.ieee.org/index.html>

81 ACM (Association for Computing Machinery): <https://www.acm.org/>

82 Dump diari: <http://dblp.org/xml/>

83 Semantic Scholar: <https://www.semanticscholar.org/>

- Mirror 2 (Trier 2): <http://dblp.uni-trier.de/db/>
- Mirror 3 (Dagstuhl): <http://dblp.dagstuhl.de/db/>

3.4.2.5.3. Model de dades

El format de les dades en XML ve definit per un DTD i es compon de les següents entitats:

- Article
- Inproceedings
- Proceedings
- Book
- Incollection
- Phdthesis
- Mastersthesis
- www
- person
- data

L'explicació completa del model de dades⁸⁴ està penjat a la seva web, al igual que el document DTD⁸⁵.

3.4.2.5.4. Accés a les dades

Les distribucions de dades enllaçades de DBLP estan disponibles amb llicència de Open Data Commons Attribution License (ODC-BY) i es poden accedir i descarregar les dades actualitzades a la seva web en un fitxer XML incloent exemples i una explicació de com parsetjar les seves dades i com realitzar-hi consultes contra ell. Es pot descarregar a la seva web⁸⁶.

Disposen d'una API amb exemples sobre la qual es poden construir consultes contra la seva web explicada aquí⁸⁷, i d'un SPARQL Endpoint⁸⁸ per a poder llançar consultes. Tots els recursos estan allotjats a Datahub⁸⁹.

3.4.2.6. ALTRES

A través de tots aquests repositoris i iniciatives presentades, i amb un nombre que va creixent cada any, es pot adquirir una visió global i altament precisa de l'estat actual de les dades obertes enllaçades d'àmbit bibliogràfic i de la importància que aquests tipus de repositoris de dades enllaçades tindran en un futur. Sal dir que n'hi ha molts d'interessants a més dels presentats en aquest treball que, degut al seu límit d'extensió, no s'han plasmat ni analitzat en profunditat en aquest treball però que seria important tenir-los en compte de cara a una possible futura ampliació d'aquest estudi.

Alguns exemples de repositoris LIBRIS⁹⁰, DPLA/Europeana Quero⁹¹, bibsonomy.org⁹², Open University⁹³, CORE⁹⁴, ODISEA⁹⁵, E-LIS⁹⁶ o els de moltes de les biblioteques nacionals d'altres

84 Explicació model de dades de DBLP: <http://dblp.uni-trier.de/xml/docu/dblpxml.pdf>

85 Document DTD del model DBLP: <http://dblp.uni-trier.de/xml/>

86 Descàrrega dades DBLP: <http://dblp.uni-trier.de/xml/> o <https://datahub.io/es/dataset/dblp>

87 Explicació API DBLP: <http://dblp.uni-trier.de/xml/docu/dblpxmlreq.pdf>

88 DBLP SPARQL Endpoint: <http://dblp.l3s.de/d2r/sparql>

89 Recursos DBLP: <https://datahub.io/es/dataset/dblp>

90 LIBRIS (Swedish Union Catalog): <http://libris.kb.se/>

91 Europeana: <http://labs.europeana.eu/api/linked-open-data-introduction>

92 Bibsonomy.org (The blue social bookmark and publication sharing system): <https://www.bibsonomy.org/> (Consultada 10/2016)

93 Open University: <https://data.open.ac.uk/>

94 CORE: <https://core.ac.uk/>

95 ODISEA: International Registry on Research Data: <https://core.ac.uk/download/pdf/19331969.pdf>

96 E-LIS: <http://eprints.rclis.org/>

països que han estat anomenades al llarg d'aquests capítols. Tampoc s'han pogut incloure altres vocabularis o iniciatives com SKOS⁹⁷, SIOC⁹⁸, ISBD⁹⁹, BIBO¹⁰⁰ o BIBFRAME¹⁰¹.

4. SPARQL

Existeixen diversos llenguatges que es poden utilitzar per interrogar ontologies per intentar obtenir llistats resultats que compleixin les restriccions plantejades a través de les preguntes que es realitzin. Encara que existeixen diversos llenguatges per a realitzar aquestes consultes essent més adient la utilització d'un o d'un altre depenent també de les característiques del model definit, en aquest capítol només s'explicarà el SPARQL, que s'utilitza per interrogar models basats en RDF i que és el que s'ha de fer servir per realitzar aquest projecte.

Durant aquest capítol es repassaran les característiques d'aquest llenguatge de consultes així com el concepte de SPARQL endpoint a través del qual qüestionar un RDFS.

4.1. INTRODUCCIÓ

SPARQL (*SPARQL Protocol and RDF Query Language*) és un llenguatge estandarditzat per a les consultes sobre models de dades en RDF. Va ser normalitzat pel *RDF Data Access Working Group* (Dawg) del W3C i és una tecnologia clau en el desenvolupament de la Web Semàntica que es va constituir com recomanació oficial del W3C al 15 de gener de 2008.

Al igual que succeeix amb SQL, s'ha de distingir entre el llenguatge de consulta i el motor per a l'emmagatzematge i recuperació de les dades. Per aquest motiu, hi ha múltiples implementacions de SPARQL, generalment lligats a entorns de desenvolupament i plataforma tecnològiques permetent obtenir informació dels grafs RDF i proporcionant la possibilitat d'extreure informació en forma de URIs, nodes buits i literals, extreure subgrafs o construir nous grafs RDF basats en els grafs de dades.

Per una part tenim que un IRI (*Internationalized Resource Identifiers*) és una generalització d'un URI que es representa delimitat entre < > o mitjançant un prefix més una etiqueta. Per exemple <<http://xmlns.com/foaf/0.1/>> o *foaf:name*. (Un node buit s'escriu com *_:node* on *node* és el nom que se li vulgui donar).

Com s'ha vist a l'apartat sobre RDF, un graf RDF és un conjunt de triples que consisteix en un subjecte (IRI o Node en blanc), un predicat (IRI) i un objecte (IRI o Node en blanc o Literal) i cada triple pot provenir des de diverses fonts (d'un document RDF, inferides des d'una altra triple RDF o poden ser expressions d'RDF de dades emmagatzemades en altres formats, com XML o base de dades relacionals).

Un literal RDF pot contenir una etiqueta per definir-li una característica. Per exemple, amb @ es podria definir l'idioma o ^^ per un tipus. Per exemple "chat"@cat per definir l'idioma "cat" o 1^^xsd:integer que definiria que 1 és de tipus integer.

Una variable de consulta és un identificador per combinar amb expressions RDF i es representa per ? o \$ seguit pel nom que li volem donar, per exemple ?titol o \$autor.

97 SKOS (Simple Knowledge Organization System): <http://skos.um.es/>

98 SIOC (Content Exchange and Semantic Interoperability Between Social Networks): <https://www.w3.org/2008/09/msnws/papers/sioc.html>

99 ISBD (International Standard Bibliographic Description): <http://www.ifla.org/node/1795>

100 BIBO (The Bibliographic Ontology): <http://bibliontology.com/>

101 BIBFRAME (Bibliographic Framework Initiative): <https://www.loc.gov/bibframe/>

Una tripleta patró és una tripleta que pot contenir variables com a subjecte, predicat i/o objecte separades per un espai i amb un punt al final. Per xemple: ?x_foaf:name ?name . (amb el punt final inclòs).

Un graf patró és un conjunt de tripletes patró escrites entre { } i separades amb un punt. N'hi ha de 5 tipus:

Bàsic	És un graf patró que conté tripletes patró.	{ <http://exemple.com/abc ?y "Hola" . ?subjecte \$objecte "literal" }
De grup	És un graf patró que conté altres grafs patrons que coincideixen.	{{ ?persona rdf:type foaf:Person { ?persona foaf:name "Aleix" }}
Opcional	És un graf patró que falla al coincidir i proporciona unions sense causar la fallida de la consulta. S'escriu amb la paraula clau OPTIONAL.	OPTIONAL { ?persona foaf:nick ?nick }
Unió	Són un parell de grafs patró que han de coincidir i unir les mateixes variables. S'escriu amb la paraula clau UNION.	{ ?node ex:name ?name } UNION { ?node vcard:FN ?name }
Graf	És una paraula clau per especificar el nom s'un graf que s'utilitza per retornar el nom del graf com una unió. S'escriu amb la paraula clau GRAPH abans del graf patró.	GRAPH {?persona foaf:name ?nom} GRAPH ?graph {?persona foaf:name ?nom}

Taula 12. Tipus de grafs patró

4.2. CONSULTES

SPARQL és, a més d'un llenguatge de consultes que especifica la sintaxi per la composició, concordança i experimentació, un protocol que descriu l'accés remot a dades i la transmissió de consultes dels clients als servidors. El resultat de les consultes es generen en XML utilitzant l'SPARQL Query Results XML Format i poden ser ordenats, limitats o presentats de diverses maneres i, encara que al principi SPARQL únicament incorpora funcions de recuperació sentències, algunes propostes també inclouen operacions per al manteniment (creació, modificació i esborrat) de dades. S'utilitzen quatre variacions de consultes:

SELECT	S'utilitza per extreure els valors en brut d'un extrem de SPARQL, els resultats es retornen en un format de taula. Serien les consultes similar a les que es fan amb SQL.
CONSTRUCT	S'utilitza per extreure la informació des del punt final SPARQL i transformar els resultats en RDF vàlids.
ASK	S'utilitza per a proporcionar un simple Verdader / Fals resultat d'una consulta en un extrem de SPARQL.
DESCRIBE	S'utilitza per extreure un graf RDF des del punt final SPARQL, el contingut es deixa al criteri de valoració per decidir sobre la base del que el responsable consideri com informació útil.

Taula 13. Tipus de consultes SPARQL

Cadascuna d'aquestes formes de consulta segueix el següent patró:

Pròleg	Opcional	BASE <IRI> PREFIX prefix: <IRI>
Format del Resultat	Obligatori	SELECT (DISTINCT) ?var1 ... ?varN SELECT (DISTINCT) DESCRIBE ?var1 ... ?varN DESCRIBE * CONSTRUCT (graf patró) ASK
Conjunt de dades	Opcional	FROM <IRI> FROM NAMED <URI>
Graf patró	Obligatori per ASK	WHERE (graf patró [FILTER expressió])
Ordenació	Opcional	DISTINCT ORDER BY ASC/DESC
Selecció	Opcional	LIMIT n, OFFSET m

Taula 14. Estructura d'una consulta SPARQL

El desenvolupament d'aquest projecte es basarà en consultes que es podran construir mitjançant una clàusula SELECT per la qual cosa no s'entrarà en detall en els altres 3 tipus de consultes.

4.2.1. PRÒLEG

La part del pròleg, com s'ha explicat és opcional i el que el format de les dades estigui en llenguatge Turtle dona la possibilitat d'expressar diverses sentències per a un recurs o valors per a una propietat amb la qual cosa podem abreujar URIs mitjançant la instrucció *@prefix* per a no haver d'anar utilitzant sempre la URI a la resta de la consulta. Al pròleg de la consulta, també es pot fer referència a aquestes URIs abreujades a les dades amb PREFIX.

Algunes de les més habituals són:

Namespaces	Prefixos	URIs
RDF	rdf:	http://www.w3.org/1999/02/22-rdf-syntax-ns#
RDFS	rdfs:	http://www.w3.org/2000/01/rdf-schema#
XML Schema Datatypes	xsd:	http://www.w3.org/2001/XMLSchema#
FOAF	foaf:	http://xmlns.com/foaf/0.1/
Dublin Core	dc:	http://purl.org/dc/elements/1.1/
OWL	Owl:	http://www.w3.org/2002/07/owl#

Taula 15. URIs més habituals abreujades a SPARQL

4.2.2. CONSULTES SIMPLS

Al estar SPARQL basat en el model de grafs i ser-ne el model triple (una triple RDF però amb la possibilitat d'una variable en lloc d'un terme RDF en la posició del subjecte, predicat i objecte) el més senzill dels models, es podria definir una consulta simple com la consistent en dues parts diferenciades:

1. Una clàusula SELECT per identificar la variable que apareixerà en el resultat de la consulta.
2. Una clàusula WHERE, que té una triple.

Com a resultat s'obté una o un conjunt de parelles variable – valor depenent de la consulta formulada i quantes de les dades incloses al model de dades compleixen els seus requeriments o cap si o els compleixen.

S'utilitzaran com exemples de consultes llançades sobre un repositori imaginari de llibres utilitzant l'ontologia Dublin Core(DC) explicada anteriorment. Els conceptes més difícils d'explicar mitjançant els exemples, es complementaran amb les mateixes consultes adaptades i executades sobre el SPARQL (Endpoint de la Dbpedia) i que poden ser comprovades, utilitzades i modificades pel lector per afermar coneixements.

4.2.2.1. SELECT i WHERE

Dades triples amb *@prefix* dc: <http://purl.org/dc/elements/1.1/> i *@prefix* ns: <http://example.org/ns#> a on cada fila és un subjecte(:book1, :book2...:bookn), cada columna el predicat i el valor a on s'encreuen l'objecte.

dc:Title	dc:Creator	dc:Publisher	dc:Identifier	dc:Date	ns:price
EL TEMOR DE UN HOMBRE SABIO	<PATRICK ROTHFUSS>	DEBOLSILLO	9788499899619	2013	11
EL NOMBRE DEL VIENTO	<PATRICK ROTHFUSS>	DEBOLSILLO	9788499082479	2011	19
LAS PUERTAS DE PIEDRA	<PATRICK ROTHFUSS>	DEBOLSILLO	9788499082888		30

Taula 16. Taula de dades exemple SPARQL

Consulta sobre per a saber els títols dels llibres i els seus autora:

```
PREFIX dc: <http://purl.org/dc/elements/1.1/>
SELECT ?titol ?autor
WHERE {
  ?x dc:title ?titol.
  ?x dc:creator ?autor.
```

```
}
ORDER BY DESC(?autor) ASC(?titol)
```

En una consulta, les variables del SELECT es separen per 1 espai en blanc en lloc d'amb comes com al SQL i marquen els objectes dels subjectes que compleixen les condicions de la consulta marcades a la clàusula WHERE.

En aquest cas, el resultat de la consulta formulada són dues parelles que corresponen als dos llibres:

titol	autor
EL NOMBRE DEL VIENTO	<PATRICK ROTHFUSS>
EL TEMOR DE UN HOMBRE SABIO	<PATRICK ROTHFUSS>
LAS PUERTAS DE PIEDRA	<PATRICK ROTHFUSS>

4.2.2.2. ORDER BY

Els resultats de les consultes, a l'igual que en SQL, es poden ordenar amb l'operador ORDER BY més el nom de la variable. Per defecte l'ordenació és ascendent però es pot seleccionar la que es vulgui amb ASC (ascendent) o DESC (descendent) ficant la variable darrera d'un dels dos entre parèntesis com s'ha fet a l'exemple anterior. Com es veu a l'exemple anterior, es pot ordenar per més d'una variable ordenant primer per la variable més a l'esquerra.

4.2.2.3. LIMIT i OFFSET

També es poden limitar el nombre màxim de resultats que retorni una consulta amb l'operador LIMIT n, a on n és un nombre sencer. En el cas de l'exemple només hi ha 3 llibres però es podria haver limitat el resultat de la consulta a un màxim de dos llibres afegint la clàusula LIMIT 2 a la línia següent del ORDER BY.

Un altre operador que es sol utilitzar amb LIMIT és OFFSET n, a on n és un nombre sencer que ens indica a partir de quin subjecte començar a mostrar resultats. Per exemple, si s'afegeix OFFSET 1, es mostraria el resultat de la consulta a partir del segon objecte que compleixi les condicions. La combinació de LIMIT i OFFSET pot ser molt útil per paginar resultats.

4.2.2.4. DISTINCT

Es poden eliminar possibles resultats duplicats amb l'operador DISTINCT:

```
PREFIX dc: <http://purl.org/dc/elements/1.1/>
SELECT DISTINCT ?autor
WHERE {
  ?x dc:title ?titol.
  ?x dc:creator ?autor.
}
```

En una consulta, les variables del SELECT es separen per 1 espai en blanc en lloc d'amb comes com al SQL. En aquest cas, el resultat de la consulta formulada són dues parelles que corresponen als dos llibres:

4.2.2.5. CONSULTES NIADES

Una possibilitat que cal destacar de SPARSQL és que es poden afegir consultes niades per tal de restringir les cerques o els subjectes sobre els quals fer-les. Aquestes subconsultes o consultes niades restringeixen l'àmbit de recerca de la consulta principal al només tenir en compte els subjectes resultants de la subconsulta com a susceptibles de ser elegibles a la consulta principal en lloc de tots els del domini. Per exemple:

```

PREFIX dc: <http://purl.org/dc/elements/1.1/>
SELECT ?titol ?autor
WHERE {
  ?x dc:title ?titol.
  ?x dc:creator ?autor.
  {
    SELECT ?x
    WHERE {
      ?pub dc:creator ?autor2 .
      ?autor2 foaf:name ?name .
      FILTER regex(?name, 'Patrick', 'i') .
    }
  }
}
ORDER BY DESC(?autor) ASC(?titol)

```

En aquesta consulta niada es restringiria la mostra de publicacions i autor resultat a les publicacions d'un autor que contingués la cadena 'Patrick' en el seu nom (Totes les entrades en el nostre exemple). Es veurà com funciona exactament l'operador FILTER en el pròxim apartat.

L'opció de niar consultes serà molt útil a l'hora d'extreure tots els autors d'una publicació en la qual un d'ells té un nombre concret, tal i com es requerirà al cas pràctic.

4.2.3. OPERADORS SPARQL

Un cop explicades les bases de com fer una consulta simple en SPARQL, els operadors bàsics per a realitzar-les i com fer altres tipus de consultes diferents menys habituals es presentaran una sèrie d'operadors que es poden fer servir i combinar per a construir unes consultes més complexes amb l'objectiu de filtrar i afitar millor la informació que es pretén esbrinar.

4.2.3.1. FILTER

Es poden realitzar restriccions a les consultes utilitzant FILTER. Per comparar un *string* amb una expressió regular s'inclouria també l'operador *regex(var, string)*:

```

PREFIX dc: <http://purl.org/dc/elements/1.1/>
SELECT ?titol
WHERE {
  ?x dc:title ?titol.
  FILTER regex ( ?titol, "VIENTO" ).
}

```

En aquest cas, el resultat seria el títol del subjecte :book2

titol
EL NOMBRE DEL VIENTO

Per no tenir en compte la diferenciació majúscules/minúscules s'ha d'afegir "i" com a tercer operador a l'operador regex: *regex(var, string, "i")*.

Es poden restringir també valors numèrics. Per exemple, si es vol obtenir un llistat dels llibres que valen més de 19€:

```

PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX ns: <http://example.org/ns#>
SELECT ?titol ?preu
WHERE {
  ?x ns:price ?preu.
  FILTER (?preu > 19 ).
  ?x dc:title ?titol.
}

```

El resultat, amb les dades anteriors seria el títol i el preu del subjecte :book3:

titol	preu
LAS PUERTAS DE PIEDRA	30

El resultat seria:

titol	preu
EL NOMBRE DEL VIENTO	19
LAS PUERTAS DE PIEDRA	30

A part de regex existeixen moltes més formes a través d'altres operadors, funcions o especificant el tipus concret de la dada que s'avalua al restringir-ne o comparar una amb FILTER.

Alguns, són els operadors habituals que també apareixen a SQL:

- Els numèrics ($A > B$, $A < B$, $A \geq B$, $A \leq B$, $A \neq B$, i $A = B$)
- Aritmètics ($A + B$, $A - B$, $A * B$ i A / B)
- Booleans ($A \& B$ o $A || B$ per indicar AND i OR)

I d'altres més específics:

- **sameTERM(A,B)**: Retorna Vertader o Fals segons dos termes RDF siguin iguals o diferents.
- **langMATCHES (A, B)**: Retorna Vertader o Fals depenent si l'idioma dels 2 literals és el mateix.
- **isIRI(A)** o **isURI(A)**: Retorna Vertader A és un RDF term és una IRI o una URI respectivament.
- **isBlank(A)**: Retorna Vertader si A està buit.
- **isLiteral(A)**): Retorna Vertader si A està és un literal.
- **Str(A)**: Retorna A en un string de A. (Per examinar, per exemple parts d'un IRI)
- **Lang(A)**: Retorna el literal en l'idioma A. (Útil si es volen extreure només els noms d'alguna cosa només en un idioma).
- **Datatype(A)**: Retorna el tipus de dada de A.

4.2.3.2. UNION

Aquest operador permetrà concatenar dues consultes sobre condicions alternatives i unir-ne el seu resultat. Per exemple, si volen consultar el títol dels llibres que valen més de 15€ i menys de 25€.

```
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX ns: <http://example.org/ns#>
SELECT ?titol
WHERE {
  {
    ?x ns:price ?preu.
    FILTER (?preu > 15).
    ?x dc:title ?titol.
  }UNION{
    ?x ns:price ?preu.
    FILTER (?preu < 25).
    ?x dc:title ?titol.
  }
}
```

El resultat seria:

titol	preu
EL NOMBRE DEL VIENTO	19
LAS PUERTAS DE PIEDRA	30

4.2.3.3. OPTIONALs

Aquest operador permet que, en SPARQL, es pugui ser capaç de consultar dades i no suspengui la consulta quan aquestes dades no existeixen definint una part opcional que amplia la solució.

Per exemple, si es volen extreure els llibres amb els anys en que han estat publicats en el cas de que aquesta data estigui registrada:

```
PREFIX dc: <http://purl.org/dc/elements/1.1/>
SELECT ?titol ?any
WHERE {
  {
    ?x dc:title ?titol.
    OPTIONAL { ?x dc:published ?any }
  }
}
```

El resultat seria, com es pot veure, serien tots el títols afegint els anys si estan referenciats:

titol	any
EL NOMBRE DEL VIENTO	2011
LAS PUERTAS DE PIEDRA	
EL TEMOR DE UN HOMBRE SABIO	2013

En SPARQL també es poden definir funcions però amb les bases presentades en aquest capítol ja es presenten els coneixements suficients per a dur a terme el projecte.

4.2.4. ALTRES TIPUS DE CONSULTES

Un cop s'han explicat les consultes simples amb SELECT es profunditzarà en els altres tipus de consultes possibles. Per les més complicades d'entendre, s'ha inclòs un exemple del seu ús sobre la DBpedia.

4.2.4.1. CONSTRUCT

CONSTRUCT s'utilitza per extreure la informació parcial o completa d'un Endpoint SPARQL i transformar els resultats obtinguts en RDF vàlids. És a dir, es podria, per exemple, invertir la relació "autor" "un llibre per "escriu" i crear un RDF resultat que presentés que escriu cada autor.

Utilitzant les dades dels exemples anteriors es podria invertir la propietat d'autoria dc:creator

```
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX ns: <http://example.org/ns#>
CONSTRUCT (?y dc:write ?x)
FROM http://exemple.es/lilibre
WHERE (
  ?x dc:creator ?y
)
```

Els triples resultants d'aquesta consulta serien triples a on cada a on cada fila és un subjecte autor(:autor1, :autor2 i autor3), cada columna el predicat construït dc:write i el valor a on s'encreuen l'objecte(de tipus llibre).

Per entendre millor aquest concepte, es pot traslladar la mateixa consulta sobre qui escriu cada llibre a la DBPEDIA. Quedaria una consulta com aquesta:

```
PREFIX res: <http://dbpedia.org/resource/>
PREFIX dbpedia-owl: <http://dbpedia.org/ontology/>
PREFIX dbpprop: <http://dbpedia.org/property/>
CONSTRUCT {?author dbpprop:write ?book}
WHERE { ?book a dbpedia-owl:Book .
  ?book dbpprop:author ?author .
  FILTER isIRI(?author)
```

} LIMIT 5

S'ha posat un operador LIMIT 5 per limitar una mica els resultats visualitzar-los més clarament:

```
@prefix dbp:      <http://dbpedia.org/property/> .
@prefix dbr:      <http://dbpedia.org/resource/> .
dbr:Henry_Watson_Fowler dbp:write dbr:A_Dictionary_of_Modern_English_Usage .
dbr:Basil_Naushad      dbp:write dbr:A_Nature_Conservation_Review .
dbr:Sylvia_Nasar       dbp:write <http://dbpedia.org/resource/A_Beautiful_Mind_(book)> .
dbr:Troy_Denning       dbp:write <http://dbpedia.org/resource/Abyss_(Star_Wars_novel)> .
dbr:Charles_Stross     dbp:write dbr:Accelerando .
```

4.2.4.2. ASK

Les consultes amb ASK, retornen, en lloc d'un nombre determinant de resultats, un simple vertader o fals segons si existeix qualche resultat que compleixi les condicions marcades al WHERE.

Així doncs, no té més complicació aquest operador i, per exemple, si es vol saber si existeix qualche llibre que costi més de 19€ i no interessa saber si n'hi ha un o n'hi ha més, ni els títols dels llibres, s'utilitzaria ASK:

```
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX ns: <http://example.org/ns#>
ASK
WHERE{
  ?x ns:price ?preu.
  FILTER (?preu > 19 ).
}
```

El resultat de la consulta utilitzant les dades del primer exemple seria Vertader, perquè, tal com s'ha vist a l'exemple presentat a FILTER, hi ha 2 llibres que compleixen aquestes condicions.

4.2.4.3. DESCRIBE

Les consultes en DESCRIBE retorna una descripció d'una URI, de la identificació d'un recurs, o pròpiament d'un recurs segons contra què s'invoqui aquesta instrucció donant com a resultat un graf RDF. Si la consulta retorna més d'un graf, la resposta és la unió de tots els grafs.

A continuació, un exemple dels 3 casos sobre l'exemple presentat i com es traslladarien les consultes. S'ha afegit també una restricció amb el LIMIT 3 simplificar les resultats.

EXEMPLE 1

Consulta:

```
DESCRIBE http://exemple.es/livre
```

RDF Resultat:

Descripció de l'esquema de Book

Consulta DBPedia:

```
PREFIX dbpedia-owl: <http://dbpedia.org/ontology/>
```

```
PREFIX dbpprop: <http://dbpedia.org/property/>
```

```
DESCRIBE dbpedia-owl:Book
```

EXEMPLE 2

Consulta:

```
DESCRIBE ?llibre
```

```
WHERE ?llibre a <http://exemple.es/livre>
```

RDF Resultat:

Descripció de les llista del 3 primers llibres

Consulta DBPedia:

```
PREFIX dbpedia-owl: <http://dbpedia.org/ontology/>
```

```
PREFIX dbpprop: <http://dbpedia.org/property/>
DESCRIBE ?book
WHERE { ?book a dbpedia-owl:Book .
} LIMIT 3
```

EXEMPLE 3

Consulta:

```
PREFIX dc: http://purl.org/dc/elements/1.1/
DESCRIBE ?autor
WHERE {
  ?x dc:title ?titol.
  ?x dc:creator ?autor.
}
```

RDF Resultat:

Descripció de les llista del autors dels 3 primers llibres

Consulta DBPedia:

```
PREFIX res: <http://dbpedia.org/resource/>
PREFIX dbpedia-owl: <http://dbpedia.org/ontology/>
PREFIX dbpprop: <http://dbpedia.org/property/>
DESCRIBE ?autor
WHERE { ?book a dbpedia-owl:Book .
?book dbpprop:author ?autor
} LIMIT 5
```

4.3. SPARQL ENDPOINTS

Un SPARQL Endpoint és una URL d'HTTP que permet realitzar consultes SPARQL contra un graf d'entrada compost per triples RDF que pot retornar el resultats en multitud de formats com Turtle, XML, Triples, etc...

Tècnicament, es pot dir que un Endpoint implementa una interfície descrita en l'especificació de W3C SPROT¹⁰² de 2008 i defineix una operació, un missatge d'entrada i dos possibles missatges de sortida. El missatge d'entrada correspon a la consulta SPARQL que es desitja executar, a més del graf (que és opcional) sobre el qual s'executa la consulta.

L'operació permet executar una consulta SPARQL sobre la lògica d'aplicació en la qual existeix el Endpoint. El primer missatge de sortida correspon als resultats obtinguts a partir de la consulta en el cas que no existeix cap error en la consulta realitzada). El segon missatge de sortida correspon a missatges d'error en el cas de falla de la consulta (que pot estar causada per errors de sintaxis, semàntics, excepcions en temps d'execució, o uns altres).

D'acord amb l'informe (State of the LOD CLOUD, 2011), la majoria de repositoris són accessibles, entre altres mètodes, mitjançant SPARQL Endpoints (68.14%) per un percentatge menor de dumps (39.66%). Aquestes possibilitats d'accés no són excloents.

En els punts anteriors, com ja s'ha explicat, s'han adaptat algunes de les consultes exemple a com es farien contra la Dbpedia i s'han executat contra el seu (Enpoint)

¹⁰² <https://www.w3.org/TR/rdf-sparql-query/>

5. CAS PRÀCTIC

5.1. SELECCIÓ DE REPOSITORI DE DADES

Per a la realització del cas pràctic s'ha decidit utilitzar uns dels repositoris que s'han analitzat a un dels capítols anteriors: DBLP.

Encara que existeixen altres repositoris més genèrics sobre els quals s'han fet alguns exemples als capítols anteriors com la DBPedia, a final s'ha seleccionat aquest a suggeriment del consultor per estar més relacionat amb l'àmbit científic.

Si bé existeixen problemes en el llançament de les peticions de l'API cap el repositoris si s'executen des de alguns navegadors web per problemes de Cross Domain que limitaran l'ús de l'aplicació a una família de navegadors web específica (Internet Explorer), el tipus de dades contingudes en ell decanten la balança en la selecció contra altres de més genèrics com la DBPedia, que no presenten aquests problemes d'accés.

El problema de Cross Domain, es basa en una limitació que s'implementa alguns navegadors web. Aquests navegadors web no permeten les peticions Ajax entre diferents dominis a no ser que estigui específicament indicat al servidor de destí o que el servidor doni la possibilitat de retornar la resposta en JSONP. El problema és que el SPARQL Endpoint de DBLP ni té aquest paràmetre activat, ni dona la opció de retornar la resposta en JSONP i, per això, no es poden realitzar peticions des de Chrome o Firefox.

Els SPARQL Endpoint d'altres repositoris com la DBPedia si que ho tenen activat, amb ells no existiria aquesta limitació però, com ja s'ha dit, el caràcter més específic de les dades incloses al repositori DBLP han fet que fos el seleccionat. L'aplicació s'ha desenvolupat perquè sigui fàcilment ampliable i es puguin incorporar recerques a altres repositoris de forma ràpida.

5.2. ANÀLIS D'ESTRUCTURA DE REPOSITORI

Després de estudiar l'estructura explicada de DBLP, el DTD i un dump de les darreres dades actuals dels enllaços proporcionats a l'anàlisi d'aquest repositori, s'han establert que la relació existent entre qualsevol tipus de publicació (Super classe Publication) que hagi fet cap persona, ve relacionada en l'autor de la mateixa per dc:creator (de Dublin Core) que vincula l'obra amb una o més d'un autor (Author, subclasse de Person).

Si l'objectiu de l'aplicació és el de llistar totes les publicacions d'un autor, serà aquesta darrera la relació a explotar per la query.

Per un altra banda, es pot extreure el títol i any de la publicació mitjançant les dc:title i dcterms:issued respectivament i el nom del autor s'extreu de la classe Author amb foaf:name.

Un cosa que caldrà tenir molt en compte a l'hora de crear la consulta és que, encara que com s'ha vist d'una manera fàcil es poden extreure totes les publicacions d'un autor, és més difícil extreure després altres dades com els coautors de la publicació lo que obliga a fer una consulta niada prèvia amb la relació presentada i un posterior tractament sobre el resultat de la consulta.

5.3. CREACIÓ DE LES CONSULTES SPAQRL

Després d'analitzar l'estructura del repositori, i tenint en que el cas pràctic tracta de contestar a la pregunta de quines publicacions ha escrit un autor determinat, en base als principis de SPARQL presentats en el capítol 3 d'aquest document s'ha creat la següent consulta.

```

PREFIX swrc: <http://swrc.ontoware.org/ontology#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX d2r: <http://sites.wiwiwiss.fu-berlin.de/suhl/bizer/d2r-server/config.rdf#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX map: <file:///home/diederich/d2r-server-0.3.2/dblp-mapping.n3#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>

```

```

SELECT (Str(?name2) as ?nomautor) (Str(?pubname) as ?nompub) (Str(?date) as ?datapub) ?pub
WHERE{
  ?autor2 foaf:name ?name2 .
  ?pub dc:creator ?autor2 .
  ?pub dcterms:issued ?date .
  ?pub dc:title ?pubname .
  {
    SELECT DISTINCT ?pub ,
    WHERE {
      ?pub dc:creator ?autor .,
      ?autor foaf:name ?name .
      FILTER regex(str(?name), 'NOM_AUTOR_RECERCA', 'i') .
    }ORDER BY ?pub
    OFFSET X
    LIMIT Y
  }
}ORDER BY ?pub ?autor2

```

La consulta principal retornarà el nom del autor (com a ?nomautor), el títol de la publicació (com a ?pubname), la data de publicació (com a ?datapub) i la publicació (?pub) sobre la mostra de publicacions resultat de la consulta niada.

La consulta niada tindrà com a variables NOM_AUTOR_RECERCA que serà directament el nom de l'autor de la recerca realitzada per l'usuari de la recerca que faci l'usuari i es limiten els resultats amb LIMIT Y en Y publicacions retornades, per no carregar massa de resultats a la consulta. Aquest valor Y se li passarà per paràmetre a la consulta i, a l'API, Y serà per defecte 15. Si es volguessin retornar més (o menys) resultats, es podria canviar el valor de Y sense cap problema.

L'ús de Y deixa també preparada l'aplicació per implementar una paginació utilitzant aquest paràmetre Y a mode de nombre de resultats per pàgina i OFFSET X que retorna el resultat a partir del nombre X també passat per paràmetre. L'OFFSET servirà per simular un resultat paginat o anar afegint més resultats al final del anterior si s'implementa aquesta .

Si, en un futur es volguessin fer recerques per varis autors separats per un espai, s'hauria de tractar aquesta variable, dividir l'entrada per els espais i afegir un OPTIONAL per cadascun dels autors sobre els quals es consulta:

```

OPTIONAL { FILTER regex(?name, NOM_AUTOR_RECERCA_1, "i") .
OPTIONAL { FILTER regex(?name, NOM_AUTOR_RECERCA_2, "i") .

```

El resultat d'aquesta consulta retornarà una llista d'autors-publicació ordenada per publicació i després per autor lo que farà que els coautors de la mateixa publicació apareixen seguits i en ordre alfabètic.

5.4. CREACIÓ DE L'API

L'API creada són una sèrie de funcions javascript agrupades en una llibreria (API.js) que s'invocarien amb l'autor, tipus de consulta i l'endpoint com a paràmetres. Les funcions estan encadenades sent la primera que es crida la funció Consultar i, a través d'ells, les altres.

Les funcions serien:

Consultar(offset)	<p>Segons el paràmetre contingut als camps dels formularis de la web tipus de consulta i repositori, crida a construeix_query per montra la query SPARQL y amb ells, llança consulta_SPARQL_JSON per fer la consulta.</p> <p>El paràmetre offset, en la primera crida hauria de ser 0 i es podria anar variant si s'implementa la paginació o els resultats incrementals.</p>
construeix_query (tipus_consulta, offset, urlrepo)	<p>Segons el tipus de consulta(tipus_consulta) el repositori urlrepo, l'offset i la variable global tamany_pagina construeix la query i la retorna amb format string. Retorna Fals si el tipus de consulta passat per paràmetre no està tipificat.</p> <p>La funció, al construir la query depenent del valor de tipus_consulta i del repositori està preparada per a poder abastir més tipus de consulta en futurs desenvolupaments i poder construir les noves querys que es defineixin contra aquests o altres repositoris.</p>
consulta_SPARQL_JSON (queryStr, tipus_consulta, urlrepo)	<p>Llança la consulta passada per paràmetre a query per <u>Ajax</u> al SPARQL Endpoint seleccionat i indicat per la variable urlrepo i, segons el resultat de la consulta, llança tractaResposta si tot va bé, o mostraErrors, si succeeix qualche problema en la consulta.</p>
tracta Resposta (jsonObj,tipus_consulta)	<p>Segons el tipus de consulta tipus_consulta, tracta les dades JSON jsonObj passades per paràmetre per presentar-les al div de la web com una taula.</p> <p>Les dades es tracten de forma que quan arriben ternes resultat autor-publicació de forma ordenada, es van acumulant els coautors i quan canvia la publicació, es crea la línia amb totes les característiques de la publicació anterior incloent els seus coautors.</p> <p>La funció està preparada per a poder abastir més tipus de consulta en futurs desenvolupaments i poder tractar-los o presentar-los segons el tipus.</p>
mostraErrors (xhr,textStatus, errorThrown)	<p>Mostra en un popup els possibles errors que puguin succeir en la consulta ajax.</p>

Taula 17. Funcions de l'API

El paràmetre tamany_pagina (Nombre de resultats màxim de la consulta), s'ha introduït a la mateixa llibreria com variable global (valor:15). És modificable si es vol ampliar el nombre de resultats en aquesta aplicació exemple i útil si es vol crear una paginació.

5.5. CREACIÓ DE L'APLICACIÓ

L'aplicació d'exemple simplement consta d'una pàgina web escrita en Html (APP.html) que conté un formulari de cerca que consta d'un camp per introduir el nom de l'autor, un selector del tipus de consulta que es vol realitzar, un selector de contra quin repositori llançar la consulta (SPARQL Endpoint) i un botó de Cerca per llançar la consulta, i d'un API escrit en Javascript que es poden executar des de qualsevol PC o allotjar a un servidor web.

L'aplicació té enllaçada la llibreria de l'API (API.js) que ha d'estar situada en el mateix directori que la pàgina web (En el PC o al servidor). La resposta a la consulta, com es veurà en l'apartat de funcionament de l'aplicació, es presentarà en una taula sota aquest formulari i mentre arriba la resposta, s'inclou un missatge del fet que s'està realitzant la consulta.

Encara que només hi ha implementada una consulta a l'API, si en posteriors desenvolupaments s'ampliés el nombre de consultes, es podrien afegir al camp selector del formulari sense problemes. El mateix passa amb el repositori, es podria incorporar un altre a la llista però també s'haurien de crear les consultes específiques per aquest repositori a l'API, ja que els repositoris probablement tindran diferents estructures i les consultes d'un o serviran a l'altre.

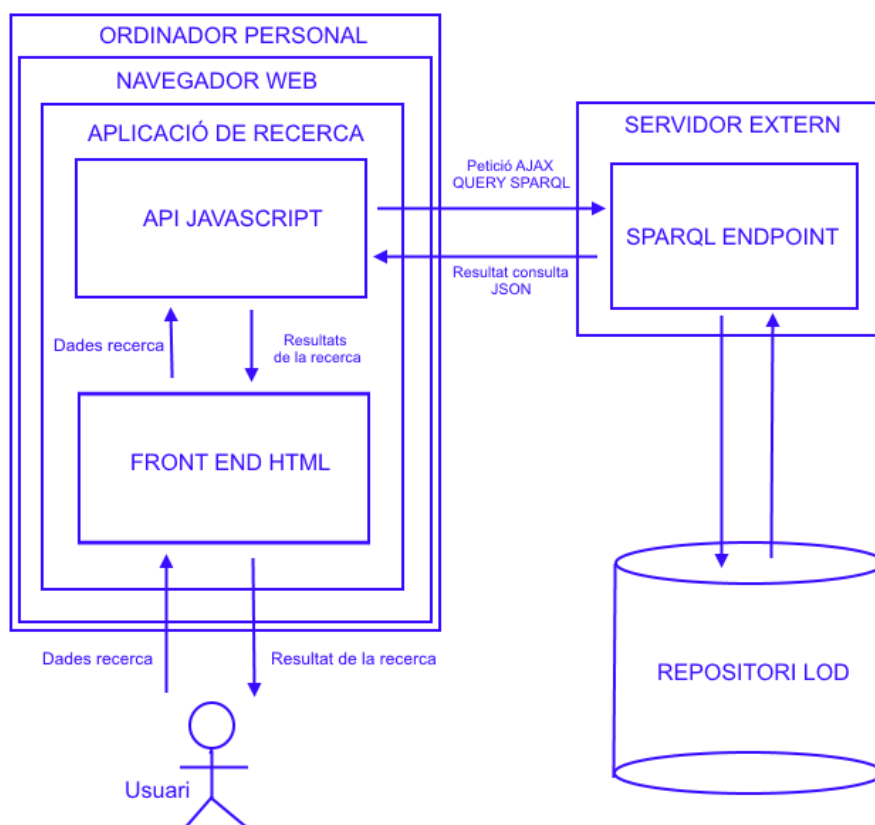


Figura 7. Estructura de funcionament de l'aplicació instal·lada en el mateix PC.

S'ha utilitzat Bootstrap¹⁰³, un framework de codi obert per millorar una mica l'aspecte i el disseny de l'aplicació i JQuery¹⁰⁴, una llibreria javascript, perquè és necessària per utilitzar Bootstrap i perquè facilita les consultes Ajax contra SPARQL Endpoints. En lloc de descarregar-les i tenir que allotjar-les al servidor, he referenciat les llibreries i fulles d'estil al CDN ¹⁰⁵de Google (jquery) i MaXCDN¹⁰⁶ (Bootstrap).

Tal i com j'ha s'explicat, per problemes de Cross-domain amb el servidor del SPARQL Endpoint, la pàgina web ha de ser executada en Internet Explorer.

5.6. FUNCIONAMENT DE L'APLICACIÓ

Si l'aplicació està allotjada a un servidor simplement cridarem a web allotjada, sinó bastarà amb arrossegar l'arxiu app.html damunt l'Internet Explorer per executar-la.

En carregar la pàgina web, la primera cosa que es visualitza serà un formulari amb un camp d'autor, a on s'haurà d'introduir la recerca, i dos camps selector, tipus de consulta i el repositori,

103 Bootstrap: <http://getbootstrap.com/>

104 JQuery: <https://jquery.com/>

105 CDN (Content Delivery Network):

106 MaxCDN: <https://www.maxcdn.com/>

que només tenen una opció possible perquè només s'ha implementat una possible consulta i un possible repositori destí contra el qual llançar-la.

Recerca Bibliogràfica

Autor:

Tipus de Recerca: Bibliografia d'un autor ▼

Repositori: DBLP ▼

Cercar

Figura 8. Formulari de cerca

Un cop introduït l'autor, es pitja al botó Cercar per llançar la consulta i apareix un missatge a la part inferior de que ja s'està realitzant la consulta. S'ha d'esperar a que aquest missatge desaparegui.

Recerca Bibliogràfica

Autor:

Tipus de Recerca: Bibliografia d'un autor ▼

Repositori: DBLP ▼

Cercar

En uns instants apareixeran els resultats de la consulta



Figura 9. Missatge d'espera

Si no es troben resultats que satisfacin la consulta, s'indicarà amb un missatge explícit.

Recerca Bibliogràfica

Autor:

Tipus de Recerca: Bibliografia d'un autor ▼

Repositori: DBLP ▼

Cercar

No existeix camp resultat per a la consulta realitzada.

Figura 10. Execució sense resultat

Si, en canvi, si que es troben resultats bibliogràfics sobre l'autor consultat, es mostraran en una taula:

Recerca Bibliogràfica

Autor: Tipus de Recerca:

Repositori:

Publicacions	
1	André van der Hoek,Ivan Mistrik,Jim Whitehead,John Grundy.(2010). Collaborative Software Engineering.
2	Ivan Mistrik,John Grundy,Jon G. Hall,Paris Avgeriou,Patricia Lago.(2011). Relating Software Requirements and Architectures..
3	Elena Simperl,John Davies,Paul Warren.(2011). Context and Semantics for Knowledge Management - Technologies for Personal Productivity..
4	John P. Collomosse,Paul L. Rosin.(2013). Image and Video-Based Artistic Stylisation.
5	John Cavazos,Keita Teranishi,Ken Naono,Reiji Suda.(2010). Software Automatic Tuning, From Concepts to State-of-the-Art Results.
6	Arne Sølvberg,Barbara Pernici,Colette Rolland,Janis A. Bubenko Jr.,John Krogstie,Oscar Pastor.(2013). Seminal Contributions to Information Systems Engineering, 25 Years of CAISE.
7	John M. Carroll.(2014). Innovative Practices in Teaching Information Sciences and Technology, Experience Reports and Reflections.
8	Gilbert Owusu,John A. W. McCall,Neil F. Doherty,Paul O'Brien.(2013). Transforming Field and Service Operations, Methodologies for Successful Technology-Driven Business Transformation.
9	Gerhard Schmidt,Hüseyin Abut,John H. L. Hansen,Kazuya Takeda.(2014). Smart Mobile In-Vehicle Systems, Next Generation Advancements.
10	Andrei Voronkov,John Alan Robinson.(2001). Handbook of Automated Reasoning (in 2 volumes).
11	John Krogstie,Keng Siau,Terry A. Halpin.(2005). Information Modeling Methods and Methodologies.
12	John A. Stankovic,Krithi Ramamritham.(1993). Advances in Real-Time Systems..
13	John Erickson.(2009). Database Technologies: Concepts, Methodologies, Tools, and Applications (4 Volumes).
14	John Erickson,Keng Siau.(2010). Principle Advancements in Database Management Technologies: New Applications and Frameworks.
15	Dimitris Karagiannis,Heinrich C. Mayr,John Mylopoulos.(2016). Domain-Specific Conceptual Modeling, Concepts, Methods and Tools.

Figura 11. Execució amb resultat

5.7. FUTURES MILLORES

Quan a perspectives de millora d'aquesta aplicació són moltes:

- **Implementació de consultes contra altres repositoris a l'API**

Com ja s'ha dit, es podria implementar la mateixa consulta contra altres repositoris si s'afegeix a la funció `construeix_query(tipus_consulta, offset, urlrepo)` i s'inclou al selector de repositori del formulari principal.

- **Implementació d'altres tipus de consulta contra altres repositoris a l'API**

També s'ha explicat que es podrien implementar noves consultes si s'afegeixen a la funció `construeix_query(tipus_consulta, offset, urlrepo)` i s'inclouen al selector de tipus de consulta del formulari principal.

- **Inclusió d'altres dades a les recerques**

A l'igual que s'ha inclòs la dada de la data de publicació, es podrien afegir més dades sobre la publicació al resultat. Algunes de forma directe, si les dades a extreure són literals, i d'altres, si la dada és un terme IRI, de forma més complexa, ja que poden requerir reestructurar tota la consulta.

- **Implementació de cerques I o O per agrupar recerques de diversos autors en una sola separant els noms o cognoms amb un espai.**

L'opció de fer consultes conjuntes també s'ha explicat anteriorment a l'apartat de Consultes SPARQL de l'aplicació. S'haurien de dividir la cadena del camp de recerca per espais i modificar les consultes per incloure cadascun de les paraules o cognoms clau a la recerca a l'apartat del FILTER de la consulta niada amb l'operador OPTIONAL.

- **Paginació de resultats o scroll infinit als resultats de les consultes.**

L'opció de presentar els resultats en pàgines seria possible, com ja s'ha explicat també, jugant amb els paràmetres X i Y de la consulta que es corresponen al tamany Y de la pàgina (o nombre de publicacions retornades en cada recerca) i l'OFFSET X que és des d'on es comencen a extreure els resultats. Depenent si volem paginació o scroll infinits, les recerques es presentarien en pàgines diferents o s'afegirien a continuació dels anteriors.

- **Exportació o guardar els resultats a fitxers amb altres formats**

Es podria implementar una opció per exportar els resultats de les recerques a un format com a .xls (o text) a fitxers per si es volen utilitzar per altres coses.

6. CONCLUSIONS

Aquest treball ha servit per examinar la feina de projectes que utilitzen les dades oberts enllaçades en l'àmbit de les referències bibliogràfiques i, després de la recerca i l'anàlisi realitzat d'algunes de les principals iniciatives i repositoris d'informació d'aquest tipus, es poden concloure varies idees.

En primer lloc, es pot veure que la Web s'està convertint en un dels majors repositoris globals d'informació, i, gràcies a aquest tipus d'iniciatives baix els principis LOD s'està fent baix un únic estàndard i, tot i que el número dels repositoris LOD continua augmentant, l'ús real de LOD per les institucions és encara limitat però amb un gran potencial en qualsevol àmbit del coneixement.

En segon lloc, si bé per problemes de temps i d'extensió d'aquest treball no s'ha pogut abastar més que una part de les iniciatives existents, queda clar que la utilització de les dades que ofereixen possibiliten grans oportunitats per a les persones i organitzacions que vulguin servir les seves dades i que, gràcies a les tecnologies semàntiques que fomenten la connexió entre repositoris, poden augmentar la seva presència i rellevància al Web.

En tercer lloc, el SPARQL és un llenguatge potent de consulta molt similar al SQL però que requereix d'un anàlisi anterior tant de l'estructura del repositori com dels vocabularis utilitzats en ell a on es voldrà fer la recerca per tal de poder construir una consulta que proporcioni les dades exactes que s'estan cercant. La possibilitat d'enllaçar dades pot complicar molt aquest anàlisi i la realització de consultes molt complexes.

Per últim, i amb la realització del senzill cas pràctic de recerques per autor, s'ha pogut constatar que, fàcilment i sense gaires coneixements tècnics, es pot crear una petita aplicació que, cobreixi la necessitat d'obtenir una informació concreta de un repositori LOD, traslladant-li la consulta en SPARQL, retornant la resposta de forma quasi immediata. La facilitat d'aquesta implementació, posa de manifest la potència i les possibilitats de les iniciatives LOD per a persones i organitzacions que vulgui aprofitar-se de la informació continguda en aquest repositoris.

Afegir també que, si bé la planificació del projecte no s'ha pogut seguir de forma estricta per problemes i oportunitats laborals no contemplades a la planificació inicial, s'han pogut augmentar

les hores de dedicació ocupades el projecte per tal de realitzar totes les entregues en els terminis previstos i s'han assolit tots els objectius plantejats a l'inici d'aquest projecte.

En conclusió, un cop analitzat l'estat de les dades enllaçades bibliogràfiques i havent vist els avantatges i oportunitats que ofereixen, juntament amb l'augment d'aquest tipus de repositoris al Web i l'increment brutal de les dades indexades en aquests projectes LOD existents, es pot preveure un brillant futur a tot aquest tipus d'iniciatives que s'encaminen, com ja s'ha dit, a revolucionar l'accés a les dades i transformant la Web en un repositori global i accessible d'informació interrelacionada.

Això tindrà un impacte positiu en l'accés públic a la informació i en la democratització del coneixement tendent a reduir la iniquitat i la bretxa digital. En aquest context, queda clar que les dades obertes enllaçades tenen, i tindran, una gran importància en la integració i reutilització de les dades en els projectes que se'n derivin de la seva utilització, impactant de forma significativa en la societat de la informació tant des del punt de vista social com des de l'econòmic.

Per a les biblioteques o arxius de documents, és una gran oportunitat de fer molt més visible el seu treball i la seva informació. Compartir dades de forma massiva i estable a més ajuda en el seu propi treball, com en la detecció de duplicats, la desambiguació terminològica, l'enriquiment de les dades i a subministrar la seva informació en formats més transparents als usuaris aliens a la seva comunitat professional.

7. GLOSSARI

- **API** (Application Programming Interface): Conjunt de subrutines, procediments, i funcions que proveeix una o varies llibreria per a que un altre programari les utilitzi a mode de capa d'abstracció. Interfície de programació per al desenvolupament d'aplicacions.
- **Diagrama de Gantt**: Tipus de diagrama de barres desenvolupat per Henry L. Gantt el 1910 que permet mostrar la planificació de les tasques d'un projecte en una escala temporal.
- **DOM** (Document Object Model): Interfície de programació orientada a objectes que representa els documents XML mitjançant un conjunt d'objectes en forma d'arbre.
- **DTD** (Document Type Definitions): Document que defineix la gramàtica dels documents XML. Element: Unitat bàsica amb capacitat per representar la lògica i la semàntica d'un document XML.
- **Element arrel**: Element del document XML que penja directament del node arrel.
- **Espai de noms XML**(Namespace). Estàndard del W3C que proporciona elements i atributs identificats de manera unívoca dins d'un document XML. Un document XML pot utilitzar elements o atributs que pertanyin a més d'un vocabulari. Si s'assigna un espai de noms a cadascun d'aquests vocabularis, s'elimina qualsevol possible ambigüitat entre elements o atributs amb el mateix nom que pertanyin a vocabularis diferents.
- **HTML** (Hypertext Markup Language): Llenguatge de marques per a la creació de pàgines web. Interfície: Plataforma de programació que proporciona un conjunt de classes i mètodes per facilitar les tasques de desenvolupament.
- **Linked Data**: Dades enllaçades, fa referència al mètode amb el que pot mostrar, intercanviar i connectar dades mitjançant URIs desreferenciables
- **Linked Open Data** (Dades Obertes Enllaçades): Dades obertes publicades de forma oberta. Agrupa els conceptes de Linked Data i de Open Data.

- **Ontologia:** Especificació formal i explícita d'una conceptualització compartida. Una conceptualització fa referència a un model abstracte d'alguns fenòmens en el món que identifiquen els conceptes d'aquest fenomen. Explícit significa que el tipus de conceptes usats i les seves restriccions s'especifiquen detalladament. Formal referència el fet de que un ontologia pot ser llegida per les màquines. Compartida reflexa la noció de que un ontologia captura el coneixement consensual, que no és quelcom aïllat sinó que és acceptat per un grup.
- **Open Data:** Moviment digital amb l'objectiu de posar les dades a disposició de totes les persones i institucions sense restriccions de copyright o, patents.
- **OWL (Ontology Web Language):** Llenguatge de marques dissenyat per a la publicació i la compartició de dades per Internet mitjançant ontologies. Constitueix una extensió de l'RDF i es deriva del DAML+OIL. Es divideix en tres subllenguatges, cadascun d'ells amb un nivell de complexitat creixent: OWL Lite, OWL DL i OWL Full.
- **RDF (Resource Description Framework):** Llenguatge que es fonamenta amb XML i que serveix per a descriure recursos.
- **RDF Schema (Resource Description Framework Schema):** Llenguatge situat per sobre d'RDF que presenta un conjunt simple de recursos RDF i propietats que permeten crear vocabularis propis.
- **SGML: Standard Generalized Markup Language,** llenguatge de marques que constitueix un estàndard extensible més potent que XML però més difícil d'implementar.
- **SQL:** Llenguatge de consulta de les bases de dades relacionals.
- **SPARQL:** És un llenguatge de consulta per a documents RDF que té un model de consulta propi per RDF i és bastant similar a SQL. Ofereix les anomenades plantilles de tripletes (triple patterns) i les conjuncions que permeten especificar les parts d'un graf RDF que volen ser recuperades.
- **Tripleta:** Element bàsic de la Web Semàntica que s'utilitza per a denotar les classes, propietats de les classes, i valors, podem crear jerarquies de classes per la classificació i descripció d'objectes, i que està format per un subjecte, un objecte i un predicat.
- **URI (Uniform Resource Identifier):** Es tracta d'una cadena de caràcters que serveixen per identificar unívocament qualsevol recurs (servei, lloc, persona, document...)
- **URL (Uniform Resource Locator):** És una cadena de caràcters que identifica de manera unívoca a cadascun dels recursos d'informació disponibles a Internet.
- **Web Semàntica:** La Web Semàntica té com a objectiu la creació d'un medi universal per a l'intercanvi d'informació basat en representacions del significat dels recursos de la web, d'una manera intel·ligible per a les màquines. Amb això es pretén ampliar l'operativitat entre els sistemes informàtics i reduir la mediació d'operadors humans en els processos intel·ligents de flux d'informació. Aquesta iniciativa per a la construcció d'una nova web està encapçalada pel W3C. W3C: World Wide Web Consortium, organisme creat amb l'objectiu de treure el màxim rendiment de la Web mitjançant el desenvolupament de protocols (especificacions, guies d'estil, software, etc.).
- **XML: Extensible Markup Language,** llenguatge extensible de marques que defineix el format estàndard per a l'estructuració de dades i d'informació.
- **XML Schema.** És una recomanació del W3C des del maig del 2001. Aquest llenguatge és una evolució dels DTDs. Permet descriure de manera formal el model de dades d'un document XML en termes de noms d'elements i d'atributs, relacions entre elements, estructura jeràrquica i tipus de dades.
- **XSL (Extensible Stylesheet Language):** Llenguatge que aplica format als documents XML, indicant com s'han de visualitzar i donant certa capacitat de transformació del contingut dels documents.

- **XSLT** (Extensible Stylesheet Language Transformations): Llenguatge que permet transformar un document XML en un altre document XML.

8. BIBLIOGRAFIA

8.1. CONCEPTES PREVIS

- XML: <https://www.w3.org/XML/> (Consultada 10/2016)
- SGML: https://en.wikipedia.org/wiki/Standard_Generalized_Markup_Language (Consultada 10/2016)
- Namespace: <https://www.w3.org/TR/REC-xml-names/> (Consultada 10/2016)
- URI: <https://www.w3.org/wiki/URI> (Consultada 10/2016)
- XML 1.0 W3C Recommendation : <https://www.w3.org/TR/xml/> (Consultada 10/2016)
- XML DTD: <https://www.w3.org/TR/xml/#dt-markupdecl> (Consultada 10/2016)
- XML-Schema (Esquema XML): <https://www.w3.org/XML/Schema> (Consultada 10/2016)
- RDF: <https://www.w3.org/RDF/> (Consultada 10/2016)
- RDF-Schema: <https://www.w3.org/TR/rdf-schema/> (Consultada 10/2016)
- OWL: <https://www.w3.org/OWL/> (Consultada 10/2016)
- Linked Data: <https://www.w3.org/DesignIssues/LinkedData.html> (Consultada 09/2016)

8.2. ESTAT DE LES ARTS OPEN LINKED DATA BIBLIOGRÀFICS

- Library of Congress(LC): <https://www.loc.gov> (Consultada 10/2016)
- FOAF: <http://www.foaf-project.org/> (Consultada 10/2016)
- DBPedia: <http://es.dbpedia.org/> (Consultada 10/2016)
- WorldCat: <https://www.worldcat.org/> (Consultada 10/2016)
- Library Linked Data Incubator Group: <https://www.w3.org/2005/Incubator/llid/> (Consultada 10/2016)
- Schema.org: <http://schema.org/> (Consultada 10/2016)
- Google Structured Data Testing Tool: <https://search.google.com/structured-data/testing-tool> (Consultada 10/2016)
- Schema.org: <http://schema.org> (Consultada 10/2016)
- FRBR: https://en.wikipedia.org/wiki/Functional_Requirements_for_Bibliographic_Records (Consultada 10/2016)
- Variations/FBRB Project: <http://www.dlib.indiana.edu/projects/vfrbr/> (Consultada 10/2016)
- Joint Steering Committee for Development of RDA: www.rda-jsc.org/rda.html (Consultada 10/2016)
- Editeur ONIX: <http://www.editeur.org/83/Overview/> (Consultada 10/2016)
- RNG: <http://relaxng.org/> (Consultada 10/2016)
- Dublin Core Metadata Initiative: <http://dublincore.org/> (Consultada 10/2016)
- British Library (BL): <http://www.bl.uk/> (Consultada 10/2016)
- British National Bibliography(BNB): <http://bnb.bl.uk/> (Consultada 10/2016)
- Repositori complet BNB: <http://thedatahub.org/dataset/bluk-bnb> (Consultada 10/2016)
- Biblioteca Nacional Alemanya (DNB): <http://www.dnb.de/> (Consultada 10/2016)
- Biblioteca Nacional de España: <http://www.bne.es/> (Consultada 10/2016)
- IFLA: <http://www.ifla.org/ES> (Consultada 10/2016)
- Datahub de BNE: <https://datahub.io/es/dataset/datos-bne-es> (Consultada 10/2016)
- WorldCat: <https://www.worldcat.org/> (Consultada 10/2016)

- OCLC: <https://www.oclc.org> (Consultada 10/2016)
- DBLP: <http://dblp.uni-trier.de> (Consultada 10/2016)
- Recursos DBLP: <https://datahub.io/es/dataset/dblp> (Consultada 10/2016)
- LIBRIS (Swedish Union Catalog): <http://libris.kb.se/> (Consultada 10/2016)
- Europeana: <http://labs.europeana.eu/api/linked-open-data-introduction> (Consultada 10/2016)
- CORE: <https://core.ac.uk/> (Consultada 11/2016)
- Bibsonomy.org: <https://www.bibsonomy.org/> (Consultada 10/2016)
- Open University: <https://data.open.ac.uk/> (Consultada 10/2016)
- SKOS: <http://skos.um.es/> (Consultada 10/2016)
- ISBD: <http://www.ifla.org/node/1795> (Consultada 10/2016)
- BIBO: <http://bibliontology.com/> (Consultada 10/2016)
- BIBFRAME: <https://www.loc.gov/bibframe/> (Consultada 10/2016)

8.3. SPARQL

- SPARQL: <http://en.wikipedia.org/wiki/SPARQL> (Consultada 11/2016)
- SPARQL i exemples: <http://www.w3.org/TR/rdf-sparql-query/> (Consultada 11/2016)
- SPAQRL: <http://www.w3.org/TR/2007/CR-rdf-sparql-query-20070614/#modReduced> (Consultada 10/2016)
- SPARQL: <http://librdf.org/query> (Consultada 11/2016)
- http://www.slideshare.net/Hicham_Qaissi/sparql-1637952 (Consultada 11/2016)
- BLOG: <http://ontoguate.wordpress.com/2009/01/25/utilizacion-de-sparql-para-consultas-a-ontologies/> (Consultada 11/2016)
- Exemples SPARQL: <http://serqlsparql.50webs.com/ejsparql.html> (Consultada 11/2016)

8.4. JAVASCRIPT PER CONSULTES SPARQL

- <http://dailyjs.com/2010/11/26/linked-data-and-javascript/> (Consultat 11/2016)

8.5. MATERIAL ASSIGNATURA

- Orientacions TFC XML-Web semàntica (Consultat 09/2016)

8.6. TESIS I PROJECTES

- “*Dades obertes enllaçades (Linked Open Data) en el món de l'educació. Una visió de l'estat de l'art d'aquesta tecnologia aplicada a l'ensenyament i l'aprenentatge*” (Josep Mascarell i Roca, Jose Javier Samper Zapater, Alexandre Viejo Galicia) (07/2014): <http://openaccess.uoc.edu/webapps/o2/bitstream/10609/34541/6/jmascarellTFM0714mem%C3%B2ria.pdf> (Consultat 10/2016)
- “*Ontologies geogràfiques turístiques de punts d'interès i documents associats*” (Bosch Vega, Daniel) Editat: Universitat Oberta de Catalunya (06/2011): <http://hdl.handle.net/10609/7933> (Consultat 10/2016)
- “*Adoption of the Linked Data Best Practices in Different Topical Domains Max Schmachtenberg*” (Christian Bizer i Heiko Paulheim) (2014): <http://dws.informatik.uni-mannheim.de/fileadmin/lehrstuehle/ki/pub/SchmachtenbergBizerPaulheim-AdoptionOfLinkedDataBestPractices.pdf> (Consultat 10/2016)
- “*Publishing Life Science Data as Linked Open Data: the Case Study of miRBase*” (2013): <http://www.aueb.gr/users/bikakis/papers/miRNALOD.pdf> (Consultat 11/2016)

- “Hacia una infraestructura de datos abiertos enlazados para una gestión agroforestal sostenible” (Alcarria, Ramón; Manso, Miguel Ángel; Robles, Tomás; Navarro, Mariano; Estrada, Jesús) Editat: V Jornadas Ibéricas de Infraestructuras de Datos Espaciais (2014): <http://www.idee.es/resources/presentaciones/JIIDE14/20141106/GestionAgroforestal.pdf> (Consultat 10/2016)
- “DEVELOPMENT OF A METADATA APPLICATION PROFILE AT THE STATE LIBRARY OF NEW SOUTH WALES” (Lynne Billington, Susanne Moir, Katie Wilson i Sue Carpenter (2015):https://www.academia.edu/8289580/Development_of_a_metadata_application_profile_at_the_State_Library_of_New_South_Wales
- “XML i Web Semàntica” (Abarca Corrales, Felix) Editat: Universitat Oberta de Catalunya (02/2010) <http://hdl.handle.net/10609/385> (Consultat 10/2016)
- “RDF i Web Semàntica” (Pascual Aran, Jordi) Editat: Universitat Oberta de Catalunya (02/2010) <http://hdl.handle.net/10609/686> (Consultat 10/2016)
- “Trabajo final de carrera web semántica” (Carvajal Tabasco, Sinesio David) Editat: Universitat Oberta de Catalunya (01/2012) <http://hdl.handle.net/10609/385> (Consultat 10/2016)
- “Llenguatges de consulta en l'àmbit de la Web Semàntica : Cas d'estudi SPARQL” (Jordana Canedo, Aleix) Editat: Universitat Oberta de Catalunya (02/2010) <http://hdl.handle.net/10609/823> (Consultat 10/2016)
- “Magatzems de dades en el context de la web semàntica” (José Sansano Martínez) Editat: Universitat Oberta de Catalunya (02/2012) <http://hdl.handle.net/10609/14980> (Consultat 01/2006)
- “Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services”.(Eric Miller, Uche Ogbuji, Victoria Mueller, and Kathy MacDougall.) Editat: Library of Congress (11/2012) <https://www.loc.gov/bibframe/pdf/marclid-report-11-21-2012.pdf> (Consultat 01/2006)

8.7. LLIBRES I ARTICLES

- Pedro Barahona, "Reasoning Web" Editat: Second International Summer School 2006, (Septembre 2006) Lisboa, Portugal.
- M. Foulonneau, “Generating educational assessment items from linked open data: The case of DBpedia,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2012, vol. 7117 LNCS, pp. 16–27

9. ANNEXOS

9.1. CODI DE L'API

```

//var endpoint = "http://dblp.l3s.de/d2r/sparql";
var tamany_pagina = 15;

function mostraErrors(xhr, textStatus, errorThrown) {
/* Funció que mostra l'error en la consulta */
    alert(textStatus);
    console.log(errorThrown);
}

function consulta_SPARQL_JSON(queryStr, tipus_consulta,urlrepo) {
    var querypart = "query=" + escape(queryStr);
/* Funció que llança la conuslta per AJAX al repositori i espera la resposata o fallida de la mateixa */
    $.ajax({
        xhrFields: { withCredentials: true },
        url: urlrepo,

```



```

    type: 'POST',
    data: querypart,
    dataType: 'json',
    crossDomain: true,
    headers: {
        'Content-type': 'application/x-www-form-urlencoded',
        'Accept': 'application/sparql-results+json'
    },
    contentType: 'application/json; charset=utf-8',
    success: function(data){ tractaResposta(data, tipus_consulta) },
    error: mostraErrors
});
};

function tractaResposta(jsonObj,tipus_consulta) {
    /* Funció que tracta el resultat de la resposta a la consulta SPARQL dependent del tipu de consulta*/
    switch(tipus_consulta) {
    case "cerca_biblio_autor":
        // Build up a table of results.
        var autors="";
            var anticpub="";
            var autor="";
            var nompub="";
            var datapub="";
            var pub="";
            var num=0;
            var result = " <table border='2' cellpadding='9' class='table-striped table-condensed'> ";
        var numres=0;
            var primer=true;
            result += "<thead><tr><th colspan=2>Publicacions</th></thead><tbody>";
            for(var i = 0; i< jsonObj.results.bindings.length; i++) {

                numres++;
                pub=jsonObj.results.bindings[i]["pub"].value;

                if (anticpub!=pub){
                    if (anticpub!=""){ // no es primer elemento hay que escribir el elemento anterior
                        result+="<tr><td>"+num+"</td><td>"+autors+".( "+datapub+" ).<b>"+nompub+".</b></td></tr>";
                    }
                    num=num+1;
                    autor=jsonObj.results.bindings[i]["nomautor"].value;
                    nompub=jsonObj.results.bindings[i]["nompub"].value;
                    datapub=jsonObj.results.bindings[i]["datapub"].value;
                    autors=autor;
                    anticpub=pub;
                }else{
                    autor=jsonObj.results.bindings[i]["nomautor"].value;
                    autors+=","+autor;
                }
            }
            //escribir ultimo elemento
            result+="<tr><td>"+num+"</td><td>"+autors+".( "+datapub+" ).<b>"+nompub+".</b></td></tr>";
            result += "</tbody></table><br>";
            if (numres>0){
                document.getElementById("results").innerHTML = result;
            }else{
                document.getElementById("results").innerHTML = "No existeix camp resultat per a la consulta
realitzada.";
            }
            break;
        default: // Consulta no reconeguda
            alert("Consulta no reconeguda");
        }
    }
}

function construeix_query(tipus_consulta, offset, urlrepo){
    /* Segons el tipus de consulta, construeix un tipus de consulta SPARQL o un altra */
    switch(tipus_consulta) {
    case "cerca_biblio_autor":
        var autor= document.getElementById("autorsearched").value;
        var query= [
            " PREFIX swrc: <http://swrc.ontoware.org/ontology#> ",
            " PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> ",

```

```

" PREFIX owl: <http://www.w3.org/2002/07/owl#> ",
" PREFIX d2r: <http://sites.wiwiss.fu-berlin.de/suhl/bizer/d2r-server/config.rdf#> ",
" PREFIX xsd: <http://www.w3.org/2001/XMLSchema#> ",
" PREFIX dcterms: <http://purl.org/dc/terms/> ",
" PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> ",
" PREFIX map: <file:///home/diederich/d2r-server-0.3.2/dblp-mapping.n3#> ",
" PREFIX foaf: <http://xmlns.com/foaf/0.1/> ",
" PREFIX dc: <http://purl.org/dc/elements/1.1/> ",
" SELECT (Str(?name2) as ?nomautor) (Str(?pubname) as ?nompub) (Str(?date) as ?datapub) ?pub ",
"   WHERE{ ",
"     ?autor2 foaf:name ?name2 . ",
"     ?pub dc:creator ?autor2 . ",
"     ?pub dcterms:issued ?date . ",
"     ?pub dc:title ?pubname . ",
"     { ",
"       SELECT DISTINCT ?pub ",
"       WHERE { ",
"         ?pub dc:creator ?autor . ",
"         ?autor foaf:name ?name . ",
"         FILTER regex(str(?name), "+autor+", 'i') . ",
"       }ORDER BY ?pub ",
"       OFFSET "+offset+" ",
"       LIMIT "+tamany_pagina+" ",
"     }",
"   }ORDER BY ?pub ?autor2 "
].join(" ");
return query;
default: // Consulta no reconeguda
return false;
}
}

function Consultar(offset){
/* Funció de cerca al pitjar el botó, mira el tipus de consulta seleccionat i crida a montra la query i llançar-la si tot es
correcte */
var tipus_consulta = document.getElementById("listacerques").value;
var urlrepo = document.getElementById("listarepositori").value;
document.getElementById("results").innerHTML = "<br><div class='message' align='center'>En uns instants
apareixeran els resultats de la consulta</div><br><div class='loader'>Loading...</div>";
var query = construeix_query(tipus_consulta,offset,urlrepo);
if(query){
consulta_SPARQL_JSON(query, tipus_consulta,urlrepo);
}else{
alert("Consulta: "+tipus_consulta+" no reconeguda.");
}
}
}

```

9.2. CODI DE L'APLICACIÓ

```

<html>
<head>
<title> SPARQL JavaScript </title>
<meta charset="UTF-8">
<!-- CSS -->
<link rel="stylesheet" href="https://maxcdn.bootstrapcdn.com/bootstrap/3.3.7/css/bootstrap.min.css"
integrity="sha384-BVYiisSIFeK1dGmJRAkycuHAHRg32OmUcww7on3R9Ydg4Va+PmSTsz/K68vbdEjh4u"
crossorigin="anonymous">
<link rel="stylesheet" href="https://maxcdn.bootstrapcdn.com/bootstrap/3.3.7/css/bootstrap-theme.min.css"
integrity="sha384-rHyoN1iRsVXV4nD0JutlnGaslCJuC7uwjduW9SVrLvRYooPp2bWYgmgJQIXwl/Sp"
crossorigin="anonymous">
<style>
.message{
color:#333333;
font-size:20px;
}
.loader {
color: #333333;
font-size: 10px;
margin: 50px auto;
width: 1em;
height: 1em;
}

```

```

border-radius: 50%;
position: relative;
text-indent: -9999em;
-webkit-animation: load4 1.3s infinite linear;
animation: load4 1.3s infinite linear;
-webkit-transform: translateZ(0);
-ms-transform: translateZ(0);
transform: translateZ(0);
}
@-webkit-keyframes load4 {
0%,
100% {
box-shadow: 0 -3em 0 0.2em, 2em -2em 0 0em, 3em 0 0 -1em, 2em 2em 0 -1em, 0 3em 0 -1em, -
2em 2em 0 -1em, -3em 0 0 -1em, -2em -2em 0 0;
}
12.5% {
box-shadow: 0 -3em 0 0, 2em -2em 0 0.2em, 3em 0 0 0, 2em 2em 0 -1em, 0 3em 0 -1em, -2em 2em
0 -1em, -3em 0 0 -1em, -2em -2em 0 -1em;
}
25% {
box-shadow: 0 -3em 0 -0.5em, 2em -2em 0 0, 3em 0 0 0.2em, 2em 2em 0 0, 0 3em 0 -1em, -2em
2em 0 -1em, -3em 0 0 -1em, -2em -2em 0 -1em;
}
37.5% {
box-shadow: 0 -3em 0 -1em, 2em -2em 0 -1em, 3em 0em 0 0, 2em 2em 0 0.2em, 0 3em 0 0em, -
2em 2em 0 -1em, -3em 0em 0 -1em, -2em -2em 0 -1em;
}
50% {
box-shadow: 0 -3em 0 -1em, 2em -2em 0 -1em, 3em 0 0 -1em, 2em 2em 0 0em, 0 3em 0 0.2em, -
2em 2em 0 0, -3em 0em 0 -1em, -2em -2em 0 -1em;
}
62.5% {
box-shadow: 0 -3em 0 -1em, 2em -2em 0 -1em, 3em 0 0 -1em, 2em 2em 0 -1em, 0 3em 0 0, -2em
2em 0 0.2em, -3em 0 0 0, -2em -2em 0 -1em;
}
75% {
box-shadow: 0em -3em 0 -1em, 2em -2em 0 -1em, 3em 0em 0 -1em, 2em 2em 0 -1em, 0 3em 0 -
1em, -2em 2em 0 0, -3em 0em 0 0.2em, -2em -2em 0 0;
}
87.5% {
box-shadow: 0em -3em 0 0, 2em -2em 0 -1em, 3em 0 0 -1em, 2em 2em 0 -1em, 0 3em 0 -1em, -
2em 2em 0 0, -3em 0em 0 0, -2em -2em 0 0.2em;
}
}
@keyframes load4 {
0%,
100% {
box-shadow: 0 -3em 0 0.2em, 2em -2em 0 0em, 3em 0 0 -1em, 2em 2em 0 -1em, 0 3em 0 -1em, -
2em 2em 0 -1em, -3em 0 0 -1em, -2em -2em 0 0;
}
12.5% {
box-shadow: 0 -3em 0 0, 2em -2em 0 0.2em, 3em 0 0 0, 2em 2em 0 -1em, 0 3em 0 -1em, -2em 2em
0 -1em, -3em 0 0 -1em, -2em -2em 0 -1em;
}
25% {
box-shadow: 0 -3em 0 -0.5em, 2em -2em 0 0, 3em 0 0 0.2em, 2em 2em 0 0, 0 3em 0 -1em, -2em
2em 0 -1em, -3em 0 0 -1em, -2em -2em 0 -1em;
}
37.5% {
box-shadow: 0 -3em 0 -1em, 2em -2em 0 -1em, 3em 0em 0 0, 2em 2em 0 0.2em, 0 3em 0 0em, -
2em 2em 0 -1em, -3em 0em 0 -1em, -2em -2em 0 -1em;
}
50% {
box-shadow: 0 -3em 0 -1em, 2em -2em 0 -1em, 3em 0 0 -1em, 2em 2em 0 0em, 0 3em 0 0.2em, -
2em 2em 0 0, -3em 0em 0 -1em, -2em -2em 0 -1em;
}
62.5% {
box-shadow: 0 -3em 0 -1em, 2em -2em 0 -1em, 3em 0 0 -1em, 2em 2em 0 -1em, 0 3em 0 0, -2em
2em 0 0.2em, -3em 0 0 0, -2em -2em 0 -1em;
}
75% {
box-shadow: 0em -3em 0 -1em, 2em -2em 0 -1em, 3em 0em 0 -1em, 2em 2em 0 -1em, 0 3em 0 -
1em, -2em 2em 0 0, -3em 0em 0 0.2em, -2em -2em 0 0;
}
}

```

```

    }
    87.5% {
        box-shadow: 0em -3em 0 0, 2em -2em 0 -1em, 3em 0 0 -1em, 2em 2em 0 -1em, 0 3em 0 -1em, -
        2em 2em 0 0, -3em 0em 0 0, -2em -2em 0 0.2em;
    }
}

</style>
<!-- Lliberies -->
<script src="https://ajax.googleapis.com/ajax/libs/jquery/3.1.1/jquery.min.js"></script>
<script src="https://maxcdn.bootstrapcdn.com/bootstrap/3.3.7/js/bootstrap.min.js" integrity="sha384-
Tc5lQib027qvyjSMfHjOMaLkfuWVxZxUPnCJA7I2mCWNlP9mGCD8wGNlCPD7Txa"
crossorigin="anonymous"></script>
<script src="/API.js"></script>
</head>
<body>

    <div id="div_formulari_cerca" class="container">
        <h2>Recerca Bibliogr fica</h2>
        <form id="formulari_cerca" class="form-horizontal">
            <div class="form-group">
                <label class="control-label col-sm-1" for="autorseached">Autor:</label>
            <div class="col-sm-4">
                <input type="text" class="form-control" id="autorseached" >
            </div>
                <label class="control-label col-sm-2" for="llistacerques">Tipus de Recerca:</label>
            <div class="col-sm-5">
                <select name="llistacerques" form="formulari_cerca" id="llistacerques" class="form-control">
                    <option value="cerca_biblio_autor">Bibliografia d'un autor</option>
                </select>
            </div>
            </div>
                <div class="form-group">
                    <label class="control-label col-sm-2 col-sm-offset-5"
for="llistacerques">Repositori:</label>
                    <div class="col-sm-5">
                        <select name="llistarepositori" form="formulari_cerca" id="llistarepositori" class="form-control">
                            <option value="http://dblp.l3s.de/d2r/sparql">DBLP</option>
                        </select>
                    </div>
                </div>
                <div class="form-group">
                    <div class="col-sm-12">
                        <button type="button" id="querybutton" class="btn btn-default col-sm-12"
onclick="Consultar(0);">Cercar</button>
                    </div>
                </div>
            </form>

        </div>

        <div id="results" class="container" class="well well-sm">

    </div>
</body>
</html>

```