

Ontologia del CRTM i d'elements d'interès de la ciutat de Madrid

Autor: David Mor España

Tutor: Felipe Geva Urbano

Professor: Ferran Prados Carrasco

Grau d'Enginyeria Informàtica

Sistemes d'Informació

10/01/2019



Aquesta obra està subjecta a una llicència de Reconeixement-NoComercial-CompartirIgual
[3.0 Espanya de Creative Commons](https://creativecommons.org/licenses/by-nc-sa/3.0/es/)

FITXA DEL TREBALL FINAL

Títol del treball:	<i>Ontologia del CRTM i d'elements d'interès de la ciutat de Madrid</i>
Nom de l'autor:	<i>David Mor España</i>
Nom del col·laborador/a docent:	<i>Felipe Geva Urbano</i>
Nom del PRA:	<i>Ferran Prados Carrasco</i>
Data de lliurament (mm/aaaa):	<i>01/2019</i>
Titulació o programa:	<i>Grau d'Enginyeria Informàtica</i>
Àrea del Treball Final:	<i>Web semàntica</i>
Idioma del treball:	<i>Català</i>
Paraules clau	<i>Web semàntica, Ontologia, Open Data, RDF, OWL, SPARQL, PHP</i>
Resum del Treball:	
<p>El present treball consisteix en l'estudi teòric de la web semàntica i tot el que l'envolta (estàndards, Open Data, etc.), i en la creació d'una ontologia mitjançant la metodologia de la Universitat de Stanford, que conté dades de tots els mitjans de transport públic que ofereix el Consorcio Regional de Transportes de Madrid i d'elements d'interès de la ciutat.</p> <p>Així doncs, l'objectiu del desenvolupament pràctic és permetre que aquesta pugui ser nodrida a partir de diversos repositoris de dades i, per tant, unificar la terminologia de l'estructura d'aquestes fonts amb la finalitat que l'ontologia pugui ser utilitzada per programaris externs que cerquin generar coneixement i proporcionar suport a l'hora de prendre decisions.</p> <p>D'altra banda, la motivació d'aquest projecte també fa referència a un vessant crític amb la gestió dels recursos públics per part dels dirigents polítics. La comparació dels resultats d'algunes consultes a l'ontologia amb l'índex de satisfacció de la població de Madrid respecte diversos àmbits públics i socials, reforcen la sensació de desigualtat per part dels habitants de determinats districtes de la ciutat.</p> <p>Finalment, el treball també presenta el desenvolupament a mida d'un programari web que interactua amb l'ontologia, fet que pretén demostrar una de les seves múltiples aplicacions i els beneficis associats a la web semàntica: dotar les dades de semàntica, generar inferències i gestionar grans fluxos de dades de forma molt eficient, entre d'altres.</p>	

Abstract:

This work is based on the theoretical study of the semantic web and all around it (standards, Open Data, etc.), and on the creation of an ontology using the Stanford University's methodology, which contains data of all the means of public transport offered by the Consorcio Regional de Transportes de Madrid and interesting elements of the city.

Therefore, the aim of the practical development is to allow it to be nourished from several data repositories and, consequently, to unify the terminology of the structure of these sources in order that the ontology could be used by external software which seek to generate knowledge and provide decision-making support.

On the other hand, the motivation of this project also refers to a critical view about the management of public resources by political leaders. The comparative analysis between the results of some queries to the ontology and the Madrid's population satisfaction index about several public and social areas, reinforce the perception of inequality on the inhabitants of certain districts of the city.

Finally, the work also presents the development of a web software which interacts with the ontology, a fact that aims to demonstrate one of its multiple applications and the benefits associated with the semantic web: provide semantic data, generate inferences and manage large flows of data in a very efficient way, among others.

Al meu pare, a la meva mare i al meu germà.

Agraïments

Primerament, m'agradaria agrair la guia contínua del meu tutor a la universitat -en Carlos Álvarez Martínez-, els *feedbacks* personalitzats del professorat de les assignatures que he cursat i tots els recursos que la UOC sempre ha posat a la meua disposició. Tot plegat m'ha ajudat a formar-me durant els darrers tres anys i escaig, i m'ha permès la consecució de molts objectius personals.

D'altra banda, també m'agradaria agrair en particular a en Felipe Geva Urbano -el meu tutor durant el desenvolupament del present treball- les tasques realitzades en els darrers quatre mesos: el seu seguiment puntual, les seves indicacions constructives i tot el seu suport, han ajudat al fet que hagi pogut assolir els objectius plantejats a l'inici d'aquest projecte.

Resum

El present treball consisteix en l'estudi teòric de la web semàntica i tot el que l'envolta (estàndards, Open Data, etc.), i en la creació d'una ontologia mitjançant la metodologia de la Universitat de Stanford, que conté dades de tots els mitjans de transport públic que ofereix el Consorcio Regional de Transportes de Madrid i d'elements d'interès de la ciutat.

Així doncs, l'objectiu del desenvolupament pràctic és permetre que aquesta pugui ser nodrida a partir de diversos repositoris de dades i, per tant, unificar la terminologia de l'estructura d'aquestes fonts amb la finalitat que l'ontologia pugui ser utilitzada per programaris externs que cerquin generar coneixement i proporcionar suport a l'hora de prendre decisions.

D'altra banda, la motivació d'aquest projecte també fa referència a un vessant crític amb la gestió dels recursos públics per part dels dirigents polítics. La comparació dels resultats d'algunes consultes a l'ontologia amb l'índex de satisfacció de la població de Madrid respecte diversos àmbits públics i socials, reforcen la sensació de desigualtat per part dels habitants de determinats districtes de la ciutat.

Finalment, el treball també presenta el desenvolupament a mida d'un programari web que interactua amb l'ontologia, fet que pretén demostrar una de les seves múltiples aplicacions i els beneficis associats a la web semàntica: dotar les dades de semàntica, generar inferències i gestionar grans fluxos de dades de forma molt eficient, entre d'altres.

Abstract

This work is based on the theoretical study of the semantic web and all around it (standards, Open Data, etc.), and on the creation of an ontology using the Stanford University's methodology, which contains data of all the means of public transport offered by the Consorcio Regional de Transportes de Madrid and interesting elements of the city.

Therefore, the aim of the practical development is to allow it to be nourished from several data repositories and, consequently, to unify the terminology of the structure of these sources in order that the ontology could be used by external software which seek to generate knowledge and provide decision-making support.

On the other hand, the motivation of this project also refers to a critical view about the management of public resources by political leaders. The comparative analysis between the results of some queries to the ontology and the Madrid's population satisfaction index about several public and social areas, reinforce the perception of inequality on the inhabitants of certain districts of the city.

Finally, the work also presents the development of a web software which interacts with the ontology, a fact that aims to demonstrate one of its multiple applications and the benefits associated with the semantic web: provide semantic data, generate inferences and manage large flows of data in a very efficient way, among others.

Paraules clau

Web semàntica, Ontologia, Open Data, RDF, OWL, SPARQL, PHP.

Índex

1.	Introducció.....	13
1.1.	Context del projecte.....	13
1.2.	Descripció del projecte	14
1.3.	Objectius del projecte	15
1.3.1.	Objectius principals	15
1.3.2.	Objectius secundaris	16
1.4.	Metodologia i procés de treball.....	16
1.5.	Planificació del projecte.....	17
1.6.	Estructura de la resta del document.....	19
2.	Estudi teòric.....	20
2.1.	Representació del coneixement	20
2.2.	Web semàntica	21
2.3.	Ontologia	25
2.4.	Llenguatges i estàndards	29
2.4.1.	XML	29
2.4.2.	RDF	31
2.4.3.	RDFS	32
2.4.4.	OWL	32
2.4.5.	SPARQL	33
2.4.6.	GeoSPARQL	33
2.5.	Linked Data.....	34
2.6.	Open Data	35
2.6.1.	CSV	35
2.7.	GTFS	36
3.	Anàlisi i disseny del projecte	37
3.1.	Anàlisi inicial	37
3.2.	Disseny de l'ontologia.....	39
3.2.1.	Classes.....	40
3.2.2.	Propietats	42
3.2.3.	Conjunt de dades	43
3.3.	Programari	44
3.3.1.	Protégé	44
3.3.2.	WebProtégé.....	45
3.3.3.	Virtuoso	46
3.3.4.	EasyRdf	46

4.	Desenvolupament del projecte.....	47
4.1.	Maquinari i programari.....	47
4.1.1.	Requisits inicials.....	47
4.1.2.	Instruccions d'instal·lació	47
4.1.3.	Protégé	50
4.1.4.	WebProtégé.....	50
4.1.5.	Virtuoso	52
4.2.	Creació de l'ontologia	52
4.2.1.	Classes.....	53
4.2.2.	Data properties	55
4.2.3.	Object properties	56
4.2.4.	Instàncies	57
4.3.	Consultes a l'ontologia	65
4.4.	Creació d'una aplicació a mida.....	75
5.	Conclusions i línies de futur	82
5.1.	Conclusions	82
5.2.	Línies de futur	83
	Bibliografia.....	84
	Annexos	86

Figures i taules

Índex de figures

Figura 1: Planificació temporal de les tasques del projecte	17
Figura 2: Diagrama de Gantt amb la línia temporal del projecte	18
Figura 3: Web 0.1.....	22
Figura 4: Web 1.0.....	22
Figura 5: Web 2.0.....	22
Figura 6: La web semàntica segons Tim Berners-Lee	23
Figura 7: La web semàntica segons W3C.....	29
Figura 8: Conjunts de dades Linking Open Data interconnectats l'any 2014 (Linked Open Data Cloud).....	34
Figura 9: Pantalla d'ús del programari Protégé	44
Figura 10: Pantalla d'ús del programari WebProtégé.....	45
Figura 11: Pantalla d'ús del programari Virtuoso	46
Figura 12: Memòria RAM assignada a la màquina virtual	48
Figura 13: Tipus d'emmagatzemament del disc dur de la màquina virtual.....	48
Figura 14: Memòria assignada al disc dur de la màquina virtual.....	49
Figura 15: Assignació de la unitat òptica a la màquina virtual per arrencar l'arxiu ISO del sistema operatiu	49
Figura 16: Secció <i>header</i> i <i>prefixes</i> del programari Protégé	53
Figura 17: Llistat de classes creades a l'ontologia	54
Figura 18: Llistat de <i>data properties</i> creats a l'ontologia	55
Figura 19: Llistat d' <i>object properties</i> creats a l'ontologia.....	56
Figura 20: Diagrama de l'ontologia creada que mostra classes, subclasses i relacions	57
Figura 21: Elecció de format per emmagatzemar l'ontologia utilitzant Protégé.....	57
Figura 22: Creació d'un compte d'usuari per utilitzar el programari WebProtégé	58
Figura 23: Importació de l'ontologia a WebProtégé	58
Figura 24: Ontologia importada correctament a WebProtégé	59
Figura 25: Selecció de l'arxiu CSV a importar utilitzant WebProtégé.....	59
Figura 26: Procés d'importació de l'arxiu CSV a WebProtégé	60
Figura 27: Elecció de format per descarregar l'ontologia utilitzant WebProtégé	61
Figura 28: Missatge d'error en executar el raonador semàntic sense la memòria necessària	61
Figura 29: Exemple d'instàncies existents a l'ontologia i les inferències creades per HermiT	62
Figura 30: Missatge d'informació amb el progrés de les tasques que realitza el raonador semàntic.....	63
Figura 31: Selecció d'axiomes a exportar a l'ontologia amb les inferències raonades.....	63
Figura 32: Selecció d'altres axiomes a exportar, incloses les anotacions de les instàncies	64
Figura 33: Selecció del format de l'ontologia inferida a exportar.....	64
Figura 34: Missatge d'informació amb el progrés de les tasques per exportar l'ontologia	65
Figura 35: Resultat de la consulta que mostra els 100 primers triplets de l'ontologia	66
Figura 36: Resultat de la consulta que mostra les 100 primeres estacions de Madrid i el districte on s'ubiquen ..	67
Figura 37: Resultat de la consulta que mostra l'estació anomenada "Vinateros".....	67
Figura 38: Resultat de la consulta que mostra les estacions de Metro ubicades al districte "01" de Madrid.....	68
Figura 39: Resultat de la consulta que mostra les estacions de Metro ubicades al codi postal "28001"	69
Figura 40: Resultat de la consulta que mostra els "Agente de Igualdad" ubicats a Madrid	69

Figura 41: Índex de satisfacció de la població de Madrid l'any 2017 respecte a l'oferta cultural per districte	70
Figura 42: Resultat de la consulta que mostra el nombre de cinemes i filmoteques de cada districte de Madrid ..	70
Figura 43: Resultat de la consulta que mostra el nombre de museus de cada districte de Madrid	71
Figura 44: Resultat de la consulta que mostra la ràtio entre zones d'aparcament i d'estacions de transport	72
Figura 45: Índex de satisfacció de la població de Madrid l'any 2017 respecte a la mobilitat i el transport	72
Figura 46: Resultat de la consulta que mostra les coordenades de latitud i longitud de les estacions de Metro ..	73
Figura 47: Resultat de la consulta que mostra les instàncies de Lotico que es troben a Madrid	74
Figura 48: Menú <i>Quad Store Upload</i> del programari Virtuoso	75
Figura 49: Editor de consultes SPARQL del programari Virtuoso	76
Figura 50: Resultat de la consulta SPARQL a l'ontologia importada al programari Virtuoso	76
Figura 51: Pantalla inicial del programari desenvolupat a mida	77
Figura 52: Selecció de la classe "Universidades" i el format "HTML" per obtenir les seves instàncies	78
Figura 53: Instàncies de la classe "Universidades" existents a l'ontologia	78
Figura 54: Filtre del text "Perú" a la classe "Embajadas_y_Consulados" per obtenir les seves instàncies	79
Figura 55: Instàncies amb el text "Perú" com a nom de la classe "Embajadas_y_Consulados"	79
Figura 56: Execució d'una consulta al programari desenvolupat a mida	80
Figura 57: Resultats de la consulta en format de sortida text pla	81
Figura 58: Exportació de l'ontologia en format de sortida HTML	81

Índex de taules

Taula 1: Dates clau de la planificació del projecte	19
Taula 2: Repositoris de dades escollits de la ciutat de Madrid, amb el seu títol literal	38
Taula 3: Definició dels camps de la classe Distrito	41
Taula 4: Definició dels camps de la classe Elemento	41
Taula 5: Definició dels camps de la classe Estación	42
Taula 6: Definició de les propietats de l'ontologia	42

1.Introducció

1.1. Context del projecte

La realització d'una ontologia que contingui dades de tots els mitjans de transport públic que ofereix el Consorcio Regional de Transportes de Madrid i d'elements d'interès de la ciutat és rellevant per la societat, ja que la consulta a aquesta podria ser factible a través d'una aplicació o servei web que podria ser d'utilitat quant a servei públic. Per exemple:

- Quines zones de Madrid tenen més i menys elements de transport?
- Els centres i serveis públics de la ciutat disposen dels suficients elements de transport?
- Hi ha algun tipus de relació entre les zones on més elements de transport hi ha vers menys àrees d'aparcament de vehicles habilitades?
- Com a turista, a quins elements de transport s'han de dirigir per a arribar a un determinat monument, museu o altre lloc d'interès turístic?
- Hi ha alguna oficina pública on només es pugui arribar fàcilment mitjançant un element de transport?
- Hi ha diferències notables entre el nombre d'elements de transport propers a cadascuna de les seus administratives de la ciutat?

El desenvolupament d'aquest treball també radica en el fet de poder obtenir una ontologia que unifiqui els continguts de diverses fonts i unificar la terminologia de les dades (millorant l'obtenció de la informació per part de programaris o serveis webs externs), amb l'objectiu de generar coneixement i proporcionar suport a l'hora de prendre decisions.

D'altra banda, cal tenir en compte que realitzar aquest projecte en forma d'ontologia és profitós per les noves necessitats que requereixen les aplicacions del present, ja que s'aconsegueix aportar valor en dotar les dades de semàntica i en permetre inferència. Per tant, la seva correcta formalització permetria reutilitzar-la per part de tercers en un futur.

A més, un altre propòsit d'aquest projecte fa referència a un vessant crític amb la gestió dels recursos públics per part dels dirigents polítics. Si bé tots els ciutadans de Madrid hi paguen els impostos corresponents, no és just que tothom pagui la mateixa base quan àrees de determinats districtes de la ciutat són menys accessibles mitjançant transport públic.

Finalment, des d'un punt de vista personal, la realització d'aquest treball m'ofereix l'oportunitat de conèixer la web semàntica i tot el que l'envolta (estàndards, Open Data, etc.), ampliant el meu coneixement acadèmic en una àrea de la informàtica relativament nova i amb interessants expectatives de futur, aprenent nous llenguatges de programació i utilitzant noves eines de programari.

1.2. Descripció del projecte

D'una banda, pel que fa a l'àmbit teòric, el desenvolupament del projecte cerca l'aprenentatge dels coneixements referents a web semàntica, Open Data i l'estàndard de *Google* GTFS; d'altra banda, quant a l'àmbit pràctic, la necessitat a cobrir del present treball és la creació d'una ontologia que relacioni diversos elements d'interès de Madrid i les estacions de transport públic del Consorcio Regional de Transportes de Madrid (CRTM), amb la finalitat d'extreure conclusions a partir dels informes públics de satisfacció per part de la població de la ciutat (siguin usuaris del servei públic o no).

A partir de la investigació inicial, el punt de partida del treball es pot resumir en els següents punts:

- No s'ha trobat cap programari a l'abast dels ciutadans que compleixi la necessitat que es desitja cobrir.
- El CRTM proveeix els repositoris de dades necessaris.
- De la mateixa manera, la ciutat de Madrid també es troba implicada en el fenomen Open Data i es poden descarregar amb facilitat les dades requerides.
- La cerca d'ontologies existents (a *Swoogle*) que haguessin pogut ser reutilitzades en el present projecte, no va obtenir cap resultat.

El desenvolupament d'una ontologia, entre d'altres, aporta:

- Cooperar amb la tendència d'ús de la web semàntica.
- Compartir l'estructura entre persones o diferents programaris (que inclús la podrien editar).
- Permetre la reutilització del coneixement del domini creat i poder-lo analitzar en profunditat.

Quant a l'àmbit tecnològic, destaquen les següents àrees en el treball a desenvolupar:

- OWL (construït sobre RDF i RDF Schema) és el llenguatge escollit per l'ontologia. RDF és un dels estàndards més popular i OWL li afegeix més característiques per descriure propietats i classes. A més, OWL és un llenguatge d'ontologies WEB i és quelcom interessant per futures fites.
- Protégé és l'eina utilitzada per crear l'ontologia. És una eina lliure i de codi obert, de reconegut prestigi i permet treballar amb RDF i OWL.
- Quant al nodriment de l'ontologia, s'ha utilitzat una versió de l'eina WebProtégé que permet la funcionalitat necessària per importar les dades de forma massiva en format CSV.
- Dels llenguatges analitzats per realitzar les consultes, SPARQL és l'escollit. És un llenguatge reconegut de consulta sobre ontologies RDF.

Pel que fa al CRTM, asseguren que el seu servei és excel·lent¹ i ho demostren, a més d'indicar diferents premis i reconeixements obtinguts, utilitzant les següents estadístiques (de l'any 2017):

- 354,3 milions de cotxes/km ferroviaris i 262,9 milions de cotxes/km per carretera.

¹ "Un transporte público de excelencia":

<http://www.comunidad.madrid/inversion/madrid/transporte-publico-excelencia>, consultat el gener del 2019.

- Xarxa de més d'11.000 km que atén a tots els municipis de la regió.
- 1.502,7 milions de desplaçaments de viatgers. Per tant, uns 5 milions de desplaçaments diaris en una ciutat on habiten 3,1 (6,4 a tota la comunitat autònoma).

Però, pel que fa a la societat i concretament la població de Madrid que és qui sufraga el servei, tots els habitants hi estan contents? Totes les àrees de Madrid estan ben comunicades fent ús de les estacions de transport? Hi ha serveis públics de la ciutat similars a d'altres però pitjor comunicats? L'afluència a aquests serveis i aquest desavantatge comparatiu tenen relació entre si?

Realitzant una recerca a fonts oficials, no ha estat possible trobar informes sobre queixes i suggeriments per part de la població de Madrid pel que fa al servei de transport públic. Però, en altre tipus de fonts, s'ha pogut constatar que efectivament hi ha àrees pitjor comunicades: Sanchinarro, Carabanchel Alto, Rejas, Valdebebas, El Cañaveral, Vallecas, Vilaverde, Los Rosales, Butarque, Almendrales...

A més, aquest fet queda palès en els informes del darrer *Observatorio de la Ciudad* (2017), on s'aprecia un retrocés respecte a anys anteriors en la satisfacció quant a mobilitat i transport públic per part dels habitants de Madrid. I, concretament pel que fa a les fonts comentades al paràgraf anterior, queden confirmades analitzant els informes de resultats districte a districte.

Finalment, per tant, la rellevància en l'àmbit pràctic del present treball radica en la consecució d'una ontologia funcional, amb la finalitat d'executar-hi consultes que permetin respondre les preguntes formulades anteriorment o, per exemple, demostrar possibles relacions entre l'anàlisi del grau de satisfacció de la població al *Observatorio de la Ciudad*, els elements d'interès escollits de Madrid i la xarxa de transport públic.

1.3. Objectius del projecte

1.3.1. Objectius principals

Objectius teòrics del treball:

- Definir de forma introductòria la web semàntica, l'Open Data i tots els conceptes que l'envolten i que formen part del desenvolupament pràctic del treball.

Objectius pràctics del projecte:

- Creació d'una ontologia amb les classes i relacions necessàries per al seu ús.
- Nodrir-la amb els repositoris de dades que ofereixen el CRTM i la ciutat de Madrid.
- Poder-hi realitzar consultes per obtenir coneixement útil per la societat.

Objectius pels usuaris:

- Poder disposar lliurement d'una ontologia, que pugui ser utilitzada per tractar el coneixement generat o reutilitzada per generar-ne de nou.

Objectius personals:

- Adquirir l'aprenentatge referent a web semàntica i Open Data.
- Conèixer i aplicar correctament la metodologia adequada per l'ontologia a crear.
- Aprendre el funcionament de les normes relacionades: RDF, OWL, SPARQL...
- Instruir-me en l'ús de noves eines informàtiques.

1.3.2. Objectius secundaris

Si el desenvolupament del treball ho permet, seria molt interessant assolir els següents objectius addicionals per enriquir-lo:

- Desenvolupar un programari a mida que interactuï amb l'ontologia creada, amb la finalitat de demostrar una aplicació pràctica del projecte i el potencial de la web semàntica.
- Poder treballar amb les coordenades de latitud i longitud per calcular distàncies aproximades entre les instàncies de l'ontologia. És a dir: no només agrupar respostes per districte i barri (cal considerar que Madrid en té de ben grans quant a superfície), sinó per alguna limitació d'espai (en metres, per exemple).
- Treballar amb formats de dades GTFS, proposat per *Google* i que s'està estenent en diferents companyies de transport públic d'arreu del món.

1.4. Metodologia i procés de treball

En aquest cas, la metodologia de treball es diferencia en dues branques:

- D'una banda, és necessària una recerca teòrica en profunditat per assolir els coneixements necessaris pel que fa a web semàntica i Open Data. Després de processar tota la bibliografia requerida, s'ha documentat al present treball com a introducció teòrica per entendre el corresponent projecte pràctic.
- D'altra banda, s'ha valorat l'estratègia d'adaptar una ontologia existent a les necessitats del projecte, ja que moltes metodologies de prestigi consideren la reutilització d'aquestes com un dels punts inicials del seu desenvolupament; malauradament, cap satisfà els requisits i s'ha escollit crear una ontologia nova.

Pel que fa específicament al projecte pràctic, del procés de treball destaca el següent:

- S'ha considerat necessari analitzar diverses metodologies per crear ontologies amb la finalitat de triar l'adient pel present treball; finalment, s'ha decidit utilitzar la guia per crear ontologies de la Universitat de Stanford, realitzada per Noy i McGuinness (2001).
- S'han investigat les eines disponibles per crear ontologies i s'ha decidit utilitzar Protégé i WebProtégé basant-se en arguments com prestigi a la comunitat, multiplataforma i codi lliure.
- S'ha analitzat tota la llista de repositoris de dades disponibles sobre la ciutat de Madrid per escollir-ne els de major interès pel projecte en qüestió. A més, s'ha estudiat la documentació associada a cadascun d'ells per dissenyar l'ontologia eficientment.
- S'ha realitzat una recerca en fonts oficials sobre el grau de satisfacció dels habitants de Madrid i s'ha fet l'anàlisi corresponent per extreure conclusions a partir de les consultes a l'ontologia creada.

1.5. Planificació del projecte

La planificació per la realització del treball queda descrita en el següent diagrama de Gantt:

Nombre	Fecha de inicio	Fecha de fin
• PAC1	20/09/18	3/10/18
• Lectura del pla docent	20/09/18	20/09/18
• Lectura dels materials	20/09/18	20/09/18
• Recerca	21/09/18	26/09/18
• Desenvolupament de la proposta	27/09/18	2/10/18
• Revisió i lliurament de l'entrega	3/10/18	3/10/18
• PAC2	4/10/18	26/10/18
• Lectura dels materials	4/10/18	4/10/18
• Recerca en profunditat	5/10/18	15/10/18
• Desenvolupament de l'estat de l'art	16/10/18	24/10/18
• Planificació del projecte	25/10/18	25/10/18
• Revisió i lliurament de l'entrega	26/10/18	26/10/18
• PAC3	29/10/18	14/12/18
• Redacció de continguts sobre web semàntica	29/10/18	2/11/18
• Anàlisi i disseny de l'ontologia	5/11/18	9/11/18
• Desenvolupament de l'ontologia	12/11/18	16/11/18
• Càrrega de dades a l'ontologia	19/11/18	23/11/18
• Consultes a l'ontologia	26/11/18	4/12/18
• Desenvolupament de l'aplicació a mida	5/12/18	13/12/18
• Revisió i lliurament de l'entrega	14/12/18	14/12/18
• PAC4	17/12/18	10/01/19
• Desenvolupament de la memòria	17/12/18	28/12/18
• Desenvolupament de la presentació	31/12/18	9/01/19
• Revisió i lliurament de l'entrega	10/01/19	10/01/19
• DEBAT VIRTUAL	21/01/19	25/01/19
• Intervir al debat	21/01/19	24/01/19
• Revisió i lliurament de l'entrega	25/01/19	25/01/19

Figura 1: Planificació temporal de les tasques del projecte

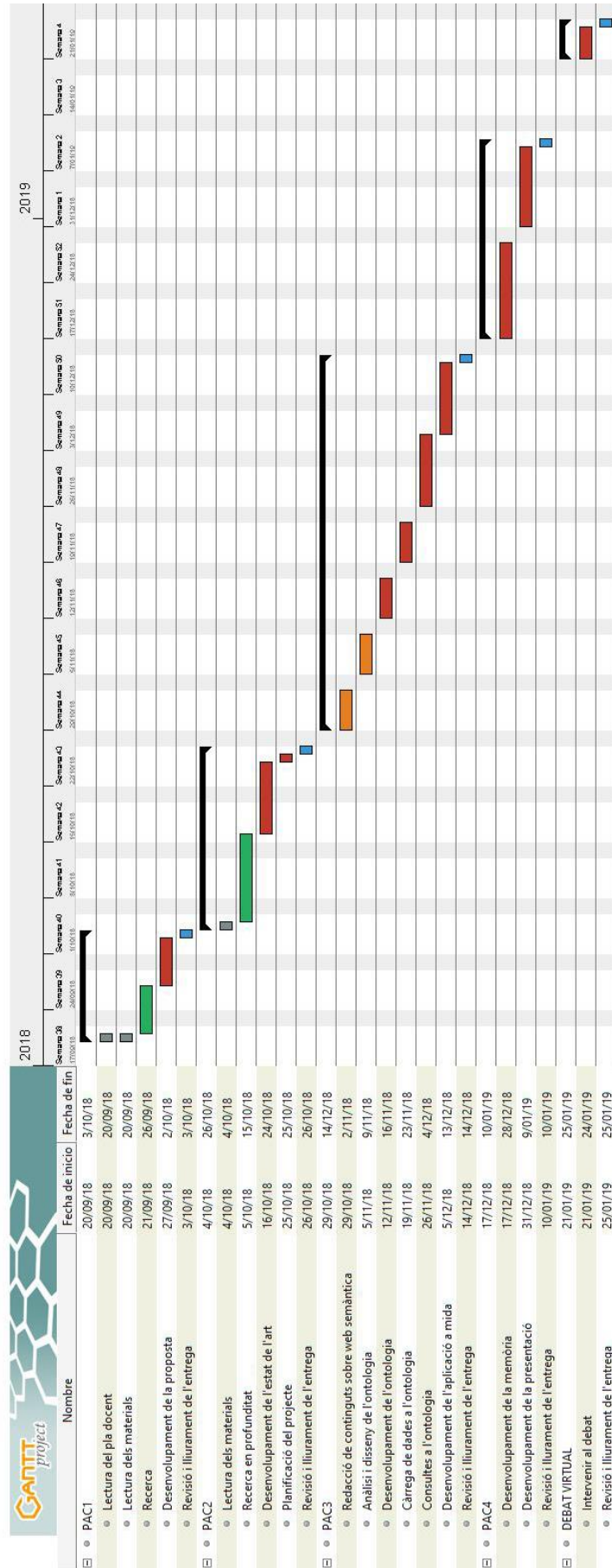


Figura 2: Diagrama de Gantt amb la línia temporal del projecte

Cadascuna de les línies del diagrama descrit anteriorment és una fita important pel desenvolupament del projecte, però les dates clau corresponen als lliuraments de les PAC de l'itinerari de l'assignatura:

	Data d'inici	Data de finalització
PAC1	20/09/2018	03/10/2018
PAC2	04/10/2018	26/10/2018
PAC3	29/10/2018	02/11/2018
PAC4	17/12/2018	10/01/2019
Debat virtual	21/01/2019	25/01/2019

Taula 1: Dates clau de la planificació del projecte

1.6. Estructura de la resta del document

A continuació es troba la resta dels capítols de la memòria. Una breu descripció dels seus continguts és la següent:

- **Estudi teòric:** es tracta d'una introducció teòrica pel que fa als conceptes bàsics del present treball. En essència, s'introdueixen qüestions referents a la representació del coneixement i es tracten les definicions de web semàntica, ontologies, Open Data i llenguatges i estàndards relacionats.
- **Anàlisi i disseny del projecte:** es detalla pas per pas la metodologia utilitzada per crear l'ontologia, des de l'anàlisi de les classes necessàries i les seves respectives relacions, passant per les dades escollides per nodrir-la, fins a les eines requerides pel seu maneig.
- **Desenvolupament del projecte:** es tracta d'una guia per la instal·lació del programari necessari i una descripció detallada de la creació de tots els elements de l'ontologia i del seu nodriment, basat en importacions massives de dades. A més, també s'inclouen les consultes realitzades i la demostració de l'aplicació creada a mida que hi interactua.
- **Conclusions i línies de futur:** s'analitzen els resultats obtinguts en el projecte basant-se en els objectius plantejats inicialment i es descriuen futures línies de treball per fer créixer l'ontologia i la seva funcionalitat.

2. Estudi teòric

2.1. Representació del coneixement

La representació del coneixement pel que fa a la intel·ligència artificial, cerca facilitar l'extracció de conclusions a partir de la interpretació d'una determinada informació. Per tant, primerament cal diferenciar dos conceptes singulars:

- **Informació:** conjunt de dades bàsiques que no reben cap mena d'interpretació i nodreixen uns determinats sistemes. Per exemple, els valors de les dades d'atur laboral a les regions d'un país.
- **Coneixement:** conjunt de dades estructurades a partir d'experiències o d'interpretar informació. Per exemple, la interpretació dels valors de les dades d'atur laboral a les regions d'un país per assegurar si són acceptables, alarmants, etc.

Així doncs, és necessari dotar a la intel·ligència artificial de mecanismes cognitius per relacionar determinats objectes en un entorn i com es relacionen, els processos que hi intervenen, coneixement que els humans tenim assumit com de sentit comú... D'aquesta manera, la resolució de qualsevol problema sempre és més assequible si s'utilitza adequadament tot el coneixement del qual es disposa.

Per tant, representar quelcom es basa en codificar el què se sap i, per fer-ho, és necessari definir els següents elements:

- La seva estructura.
- L'ús que li donaran els éssers humans.
- L'ús que li donaran els sistemes d'intel·ligència artificial.
- L'adquisició de coneixement.
- El seu emmagatzematge i manipulació.

A l'hora de definir un esquema de representació, que és l'instrument per codificar el coneixement d'un domini a un llenguatge simbòlic en un dispositiu informàtic, cal tenir en compte la combinació dels següents elements pel que fa estrictament a l'àmbit tècnic:

- **Part estàtica:**
 - Estructura de dades que codifica el problema.
 - Operacions per crear, modificar i esborrar elements de l'estructura.
 - Predicats que permeten consultar l'estructura de dades.
 - Semàntica de l'estructura per definir la relació entre la realitat i la representació.
- **Part dinàmica:**
 - Estructures de dades que emmagatzemen el coneixement del món real.
 - Procediments per interpretar les dades de la part estàtica a partir del coneixement de la part dinàmica.
 - Procediments per aplicar l'estratègia de control de dades.

- Procediments per generar nou coneixement.

Els esquemes de representació, però, van més enllà d'estructures i bases de dades; així doncs, aquests han de complir les següents propietats pel que fa a la representació i al seu ús:

- **Representació apropiada:** ha de poder representar tot el coneixement del domini definit, amb la finalitat de solucionar el problema donat mitjançant la informació que emmagatzema.
- **Inferència apropiada:** cal que manipuli les estructures de representació per generar-ne de noves a partir dels nous coneixements inferits.
- **Eficiència inferencial:** ha de poder incorporar informació addicional que agilitzi els mecanismes d'inferència amb la finalitat d'optimitzar els còmputos.
- **Eficiència d'adquisició:** cal que sigui capaç d'incorporar amb facilitat nova informació i la seva posterior representació.

A més de les propietats esmentades anteriorment, un sistema de representació digne de ser-ho ha de complir-ne les següents:

- **Claredat:** ha de permetre identificar el coneixement representat fàcilment.
- **Naturalitat:** cal que el coneixement es pugui representar en la seva forma original.
- **Modularitat:** ha de poder fragmentar el coneixement sense perdre eficiència ni eficàcia.

D'altra banda, però, cal tenir en compte que els sistemes de representació són sempre incomplets, ja que el món real és canviant i el volum de coneixement a representar pot convertir-se en més complex per la gran quantitat de detalls que emanen de la realitat del domini definit.

En conclusió, un dels principals problemes de la representació del coneixement radica en el fet de generar un sistema de raonament que extregui les conclusions desitjades dins del marc que defineixen els límits d'un determinat problema. En aquest sentit, la web semàntica i llenguatges i estàndards com *Resource Description Framework (RDF)*, *RDF Schema*, *Web Ontology Language (OWL)* i *SPARQL Protocol and RDF Query Language*, lideren els desenvolupaments més actuals en aquest àmbit.

2.2. Web semàntica

Per saber on anem, cal saber d'on venim. Aquesta frase que encara perdura amb el pas dels anys, també s'aplica a l'evolució de la Web², remuntant-se al 20 de desembre de 1990 quan Tim Berners-Lee³ va posar en marxa la primera pàgina web de la història en el centre d'investigacions nuclears *CERN* –fet que va suposar el naixement de la *World Wide Web*–, passant per la Web 1.0 que oferia una experiència d'usuari passiva on no era possible interactuar amb els documents que es

² Web 0.1: <http://info.cern.ch/hypertext/WWW/TheProject.html>, consultat el gener del 2019.

³ Portal dedicat a la figura de Tim Berners-Lee, per part de W3C: <https://www.w3.org/People/Berners-Lee/>, consultat el gener del 2019.

podien visualitzar a través d'un ordinador, fins a l'arribada de la quotidiana Web 2.0, molt més interactiva i accessible des de diversos tipus de dispositius tecnològics.

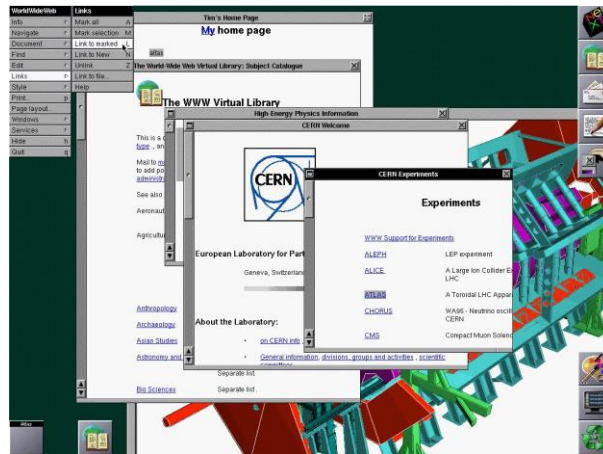


Figura 3: Web 0.1

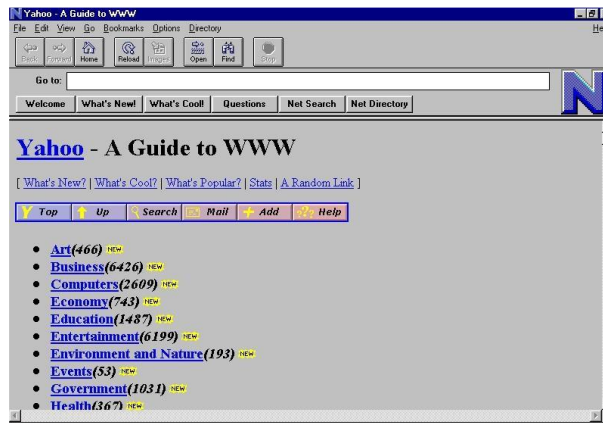


Figura 4: Web 1.0

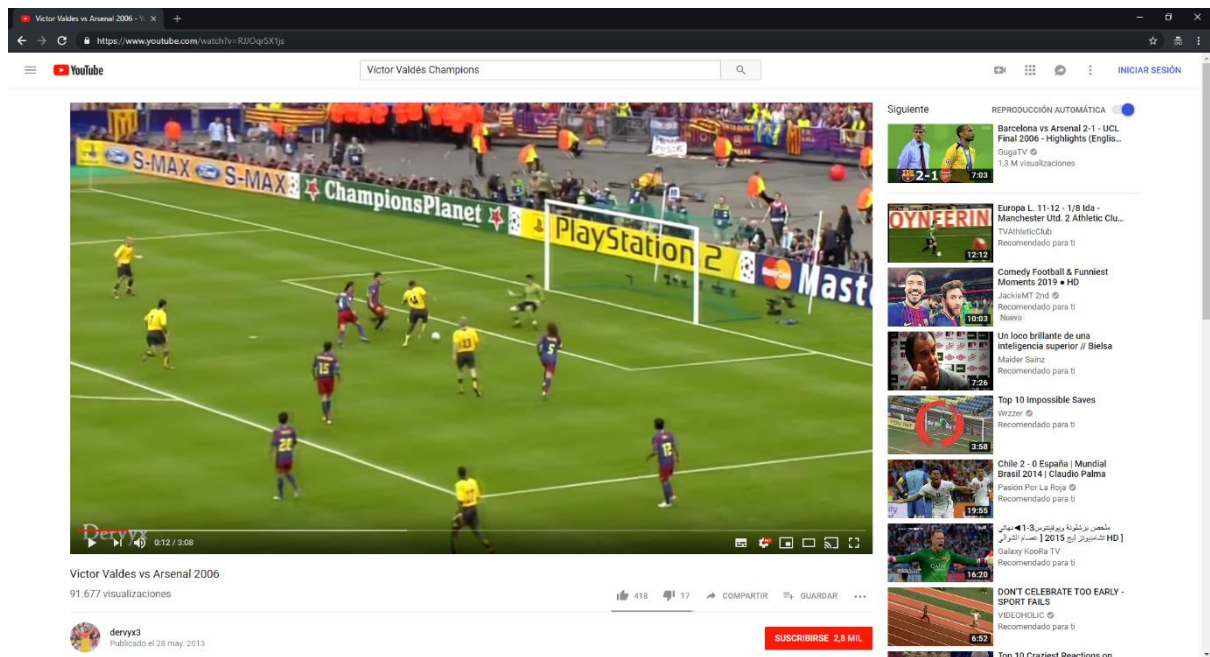


Figura 5: Web 2.0

Si les Web descrites anteriorment es diferencien per la manera en com els usuaris i les usuàries interactuaven i interactuen amb elles, el següent pas es dirigeix cap a la creació de la web semàntica, procés afavorit per l'enorme quantitat de dades que hi ha a les pàgines web d'arreu del món. El mateix Tim Berners-Lee defineix la web semàntica descrivint serveis i agents web que podran entendre el contingut de les pàgines web i relacionar-los entre ells, afavorint la interrelació de recursos.

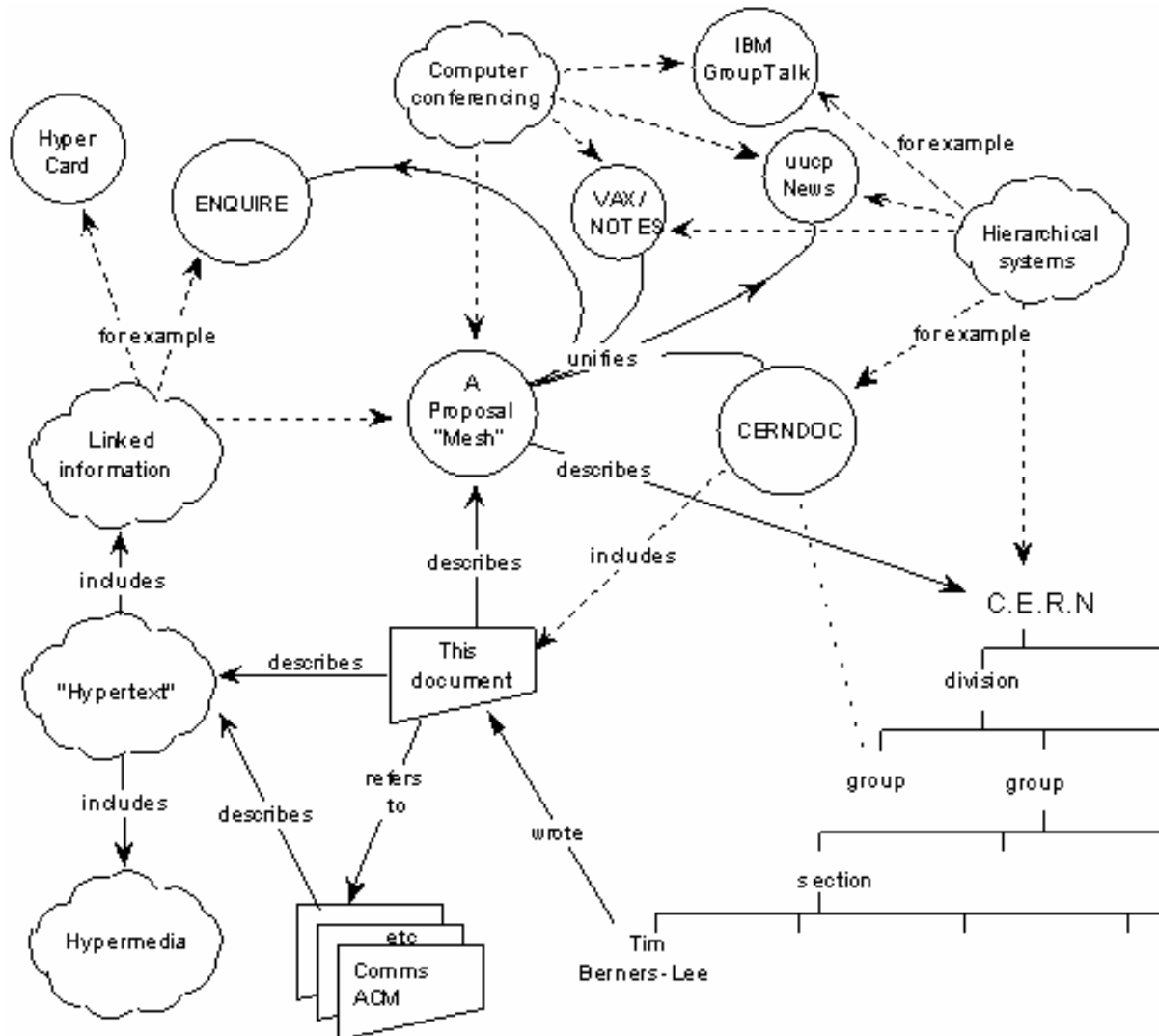


Figura 6: La web semàntica segons Tim Berners-Lee

D'una banda, els serveis web ofereixen de forma pública unes determinades funcionalitats que són accessibles per programari de tercers; per exemple, les pàgines web de comparadors de preus de vols consulten els serveis web de diverses aerolínies per obtenir el seu llistat de tarifes i mostrar-les al públic. D'altra banda, els agents web són entitats capaces de percebre l'entorn on es troba l'usuari o la usuària, processar-lo i actuar en conseqüència per assolir un determinat objectiu utilitzant tècniques d'intel·ligència artificial; per exemple, tornant a l'anterior referent a les pàgines web de comparadors de preus de vol, un agent web hauria de poder oferir, durant el mateix procés de compra, reserves d'allotjament i desplaçament basant-se en informació com la ubicació i la data.

D'aquesta manera, apareixen les tecnologies semàntiques que, en aquest cas, ofereixen una infraestructura per la Web on les dades es troben accessibles gràcies a estàndards ja definits, entrelaçades (tenint en compte, per exemple, que han de ser expressades de manera unívoca) i disponibles des de qualsevol aplicació amb l'objectiu d'aconseguir la màxima integració possible.

Basant-se concretament en termes oficials, W3C⁴ descriu la web semàntica de la següent manera:

“La Web Semàntica és la Web de les dades. Hi ha moltes dades que utilitzem cada dia i que no són part de la Web. Podem veure les anotacions bancàries a la web i fins i tot les nostres fotografies i cites al calendari. Però podem veure les fotos en un calendari per veure què és el que estava fent quan les vaig fer? Puc veure les meves anotacions bancàries al calendari? La resposta a aquestes preguntes és no. I per què no? La resposta és perquè no tenim una Web de dades. I això és a causa que les dades estan controlades per les aplicacions i cadascuna les guarda i tracta de manera particular.

La Web Semàntica tracta sobre dues coses. Sobre formats comuns per a l'intercanvi de dades, mentre que en la Web original solament s'intercanvien documents. I tracta sobre els llenguatges que representen les dades com a objectes del món real. La Web Semàntica proporciona un framework comú que permet a les dades ser compartides i reutilitzades a través dels límits imposats per aplicacions, empreses o comunitats. És un esforç de desenvolupament col·laboratiu liderat per la W3C i un gran nombre d'investigadors i socis industrials.”⁵

A més, segons W3C cal distingir clarament els conceptes de web semàntica i Web 3.0: si bé el primer es refereix al conjunt d'activitats d'àmbits molt diversos que cerca esdevenir en una base de coneixement, el segon, en canvi, es basa sobretot a l'entorn on serveis i agents web intercanvien i processen dades per generar nova informació.

En conclusió, hi esdevenen tres elements clau referents a la web semàntica:

- **Semàntica:** s'encarrega de dotar de significat a les dades per a ser interpretat per màquines.
- **Metadades:** s'utilitzen per catalogar les dades amb etiquetes, que es podran processar posteriorment.
- **Ontologies:** es fan servir per definir un model estructurat de termes i relacions a partir d'unes regles específiques en un domini concret.

⁴ W3C són les sigles de *World Wide Web Consortium*, que es tracta d'un organisme neutre d'àmbit internacional que treballa per generar estàndards que assegurin el creixement de la *World Wide Web* a llarg termini. Va ser fundat l'1 d'octubre de 1994 per Tim Berners-Lee i actualment tenen relacions amb 475 organitzacions membre: <https://www.w3.org/Consortium/Member/List>, consultat el gener del 2019.

⁵ Definició de la web semàntica per part de W3C: <http://www.w3.org/2001/sw>, consultat el gener del 2019.

2.3. Ontologia

Tal com s'ha indicat anteriorment, una ontologia⁶ -en termes informàtics- és un model estructurat de termes i relacions a partir d'unes regles específiques en un domini concret. Acostuma a estar escrita en un llenguatge formal, sota un suport informàtic per poder ser interpretada per diferents tipus de programari i representa la informació que un grup de persones considera necessària per definir el domini i delimitar el problema en qüestió. Per tant, les ontologies són clau per aconseguir l'objectiu de la web semàntica.

El terme ontologia és adoptat per la intel·ligència artificial durant la segona meitat de la dècada dels 90, quan s'incorpora a la base tècnica del món de l'enginyeria web per la inclusió de descripcions semàntiques explícites de recursos. Des d'aleshores, diferents personalitats del sector han tractat de definir què és una ontologia:

- **Tom Gruber (1993):** *“Es tracta d'una especificació explícita d'una conceptualització”*. Aquesta definició és una de les més acceptades en l'àmbit de la intel·ligència artificial.
- **Uschold i Grüninger (1996):** *“Es tracta d'una comprensió compartida d'algun domini d'interès”*.
- **Willem Nico Borst (1997):** *“Es tracta d'una especificació formal d'una conceptualització compartida”*. En aquest cas, Borst va tractar de fer més específica la definició de Gruber.
- **Nicola Guarino (1998):** *“Es tracta d'una teoria lògica que considera el significat intencional d'un vocabulari formal, és a dir, el seu compromís ontològic per a una conceptualització particular del món”*.
- **Joost Breuker (1999):** *“Es tracta d'una representació explícita d'una conceptualització cognitiva, és a dir, la descripció dels components de coneixement rellevants en l'àmbit del modelatge”*.
- **B. Chandrasekaran (1999):** *“Es tracta de les teories sobre tipus d'objectes, les propietats d'aquests objectes, i les relacions entre els objectes que són possibles en un domini específic de coneixement”*.
- **Tim Berners-Lee (2001):** *“Es tracta de la materialització en un document informàtic que defineix formalment les relacions entre termes i constitueix una base de coneixement”*.
- **Noy & McGuinness (2001):** *“Es tracta d'una descripció formal i explícita dels conceptes d'un domini de discurs, propietats de cada concepte descrivint diverses de les seves característiques i atributs, i restriccions de les propietats”*.
- **Vladan Devedžić (2002):** *“Es tracta d'un sistema de conceptes definits i interpretats de manera declarativa. El sistema defineix el vocabulari d'un domini i les restriccions que s'apliquen perquè aquests puguin ser combinats en el model del domini”*.

⁶ Una de les primeres personalitats que va utilitzar el terme *ontologia* va ser Aristòtil, en el seu intent de classificar tot l'existent en l'univers: *“la ciència de l'ésser com a ser”*.

De la mateixa manera, al llarg de la història diverses personalitats han tractat de classificar els tipus d'ontologia segons la perspectiva en què són aplicades: dimensió, formalitat, propòsit, ús, matèria, volum, tipus d'estructura, etc. En el cas particular de Nicola Guarino (1998), les ontologies es poden classificar segons el seu grau de dependència i relació amb una tasca específica, afavorint a la seva modularitat i a la reutilització del coneixement:

- **Ontologies d'alt nivell o genèriques:** descriuen conceptes molt generals (per tant, independents d'un domini particular) com l'espai, el temps, les accions, etc.
- **Ontologies de domini:** descriuen el vocabulari d'un domini genèric amb la finalitat d'especialitzar el coneixement d'ontologies genèriques.
- **Ontologies de tasques o activitats:** descriuen una tasca o activitat tot especialitzant ontologies d'alt nivell.
- **Ontologies d'aplicació:** descriuen els conceptes que depenen de les ontologies de domini o de tasques i acostumen a ser dissenyades a mida per aplicacions específiques. Ja que depenen d'un àmbit concret, es tracten de les més específiques de totes.

Generalment, cada ontologia necessita els següents elements per representar coneixement:

- **Conceptes:** es tracten de les idees bàsiques que es volen formalitzar.
- **Relacions:** representen l'enllaç entre conceptes.
- **Funcions:** es tracta d'un tipus concret de relació que refereix un element que s'obté a partir del càlcul amb d'altres de l'ontologia.
- **Instàncies:** s'utilitzen per representar objectes determinats d'un concepte.
- **Axiomes:** es tracten de teoremes que es declaren sobre relacions que han de complir determinats elements de l'ontologia.

Pel que fa al disseny i creació d'una ontologia, Noy i McGuinness (2001) proposen les següents regles fonamentals:

- No hi ha una única forma correcta de modelar un domini, ja que sempre hi ha alternatives viables depenent de l'ús que es desitgi fer de l'ontologia.
- El desenvolupament d'una ontologia és un procés necessàriament iteratiu.
- Els conceptes de l'ontologia han de ser propers als objectes (substantius) i relacions (verbs) en el domini d'interès.

Noy i McGuinness (2001) també van proposar les activitats fonamentals que conté el procés de disseny i creació d'una ontologia:

- Definir les classes de l'ontologia.
- Organitzar les classes en una jerarquia taxonòmica.
- Definir les propietats i descriure els valors possibles per cadascuna.
- Establir valors a les propietats per cadascuna de les instàncies.

Finalment, hi ha moltes metodologies per dissenyar i crear ontologies. Ja que no hi ha una que sigui millor que les altres, l'elecció de l'adient s'haurà de basar en aspectes com la reutilització, explotació, volum de documents que continguin informació sobre el domini, etc. Algunes de les metodologies més reconegudes són:

- **Guia per crear ontologies de la Universitat de Stanford:** realitzada per Noy i McGuinness (2001), és una metodologia simple que consta dels següents passos:
 1. **Determinar el domini i l'abast de l'ontologia:** és recomanable formular una sèrie de preguntes que la base de coneixement basada en l'ontologia hauria de poder respondre.
 2. **Considerar la reutilització d'ontologies existents:** comprovar si és possible construir l'ontologia a partir d'alguna o algunes ja existents, modificant els seus límits pel domini en qüestió.
 3. **Enumerar els termes rellevants de l'ontologia:** cal enumerar qualsevol terme que tingui relació amb el domini definit. Aquest procés s'ha de realitzar d'una manera precisa i sense ambigüitats.
 4. **Definir les classes i la seva jerarquia:** de l'anterior llista, cal escollir els termes independents per construir les classes i organitzar la seva jerarquia.
 5. **Definir les propietats de les classes:** cadascuna de les classes consta d'una estructura interna formada pels conceptes propis que la descriuen.
 6. **Definir les restriccions de les propietats:** per exemple, el tipus de valor que poden permetre, la seva cardinalitat, etc.
 7. **Crear instàncies:** és el pas on s'escull una classe, es crea una instància individual d'aquesta i s'introdueixen les seves propietats.
- **Methontology:** realitzada al laboratori d'Intel·ligència Artificial de la Universitat Politècnica de Madrid, aquesta metodologia destaca per ser molt útil a l'hora de crear ontologies a escala de coneixement. Els seus passos són:
 1. **Especificació:** s'identifica el propòsit i l'àmbit de la tecnologia.
 2. **Conceptualització:** es descriu en un model conceptual les especificacions de l'ontologia.
 3. **Adquisició de coneixement:** és el procés d'adquirir el coneixement del domini. Aquest pas acostuma a ser realitzat coincidint amb l'execució d'altres.
 4. **Reutilització:** sempre que sigui possible, és recomanable reutilitzar altres ontologies.
 5. **Formalització:** es transforma la descripció conceptual en un model formal.
 6. **Implementació:** s'implementa l'ontologia formalitzada en un llenguatge de representació del coneixement.
 7. **Avaluació:** també de forma periòdica, es jutja el funcionament de l'ontologia en referència al problema en qüestió.
 8. **Documentació:** cal documentar què, com i per què s'ha fet.
 9. **Manteniment:** es revisa periòdicament i, en cas necessari, es modifica l'ontologia.

- **SENSUS-Based:** aquesta metodologia és molt recomanable per a derivar ontologies de domini específic a partir d'una d'alt nivell. Els passos que segueix són els següents:
 1. Inicialment, es recopilen un seguit de termes anomenats termes llavor.
 2. Cadascun d'aquests és vinculat a l'ontologia d'alt nivell.
 3. A l'ontologia de domini específic que s'està creant, s'inclouen tots els conceptes que tenen referència amb els termes llavor i l'arrel de l'ontologia d'alt nivell.
 4. Altrament, també s'inclou la resta de termes que són rellevants en el domini.
 5. Com a últim pas, els termes no seleccionats però que tenen un gran nombre de camins, s'afegeixen a l'ontologia que s'està creant.
- **Tove:** realitzada per Michael Grüninger i Mark Fox, la metodologia va ser creada a partir de l'experiència obtinguda en el projecte TOronto Virtual Enterprise (TOVE). Els seus passos són els següents:
 1. **Capturar escenaris motivadors:** primerament, cal identificar problemes o exemples en el domini que no els resolen ontologies existents.
 2. **Formular preguntes de competència informals:** a partir del pas anterior, cal recopilar un seguit de preguntes que l'ontologia haurà de saber respondre a partir de la representació del coneixement.
 3. **Especificar la terminologia de l'ontologia:** de les preguntes informals, s'extreuen els termes clau de l'ontologia i es codifiquen en un llenguatge formal.
 4. **Formular la llista de preguntes de competència formalment:** a partir del pas anterior, es reescriuran les preguntes utilitzant els termes clau i el seu llenguatge formal.
 5. **Especificar axiomes i definicions pels termes:** amb la finalitat de respondre les preguntes recopilades al punt anterior, cal definir els conceptes i relacions de l'ontologia.
 6. **Avaluar l'ontologia:** a partir d'uns criteris fixats, s'avaluarà la competència de l'ontologia a partir de les respostes que ofereix.
 7. **Definir les condicions sota les quals les solucions a les preguntes són completes.**
- **Enterprise:** realitzada per Uschold i King (1995), és relativament comú trobar ontologies utilitzades actualment basades en aquesta metodologia. Els seus passos són els següents:
 1. **Identificar el propòsit i l'abast de l'ontologia.**
 2. **Crear l'ontologia:** en aquest pas cal identificar els conceptes claus i les seves relacions en el domini, codificar el coneixement utilitzant un llenguatge formal i integrar-lo, si s'escau, examinant ontologies existents.
 3. **Avaluar tècnicament l'ontologia.**
 4. **Documentar l'ontologia:** un bon procés de documentació, garanteix que l'ontologia pugui ser compartida i reutilitzada.

2.4. Llenguatges i estàndards

La web semàntica es pot interpretar com una arquitectura jerarquitzada de capes. A la figura 7 es poden apreciar els nivells que proposa W3C i les tecnologies i estàndards utilitzats per manejar la informació semàntica.

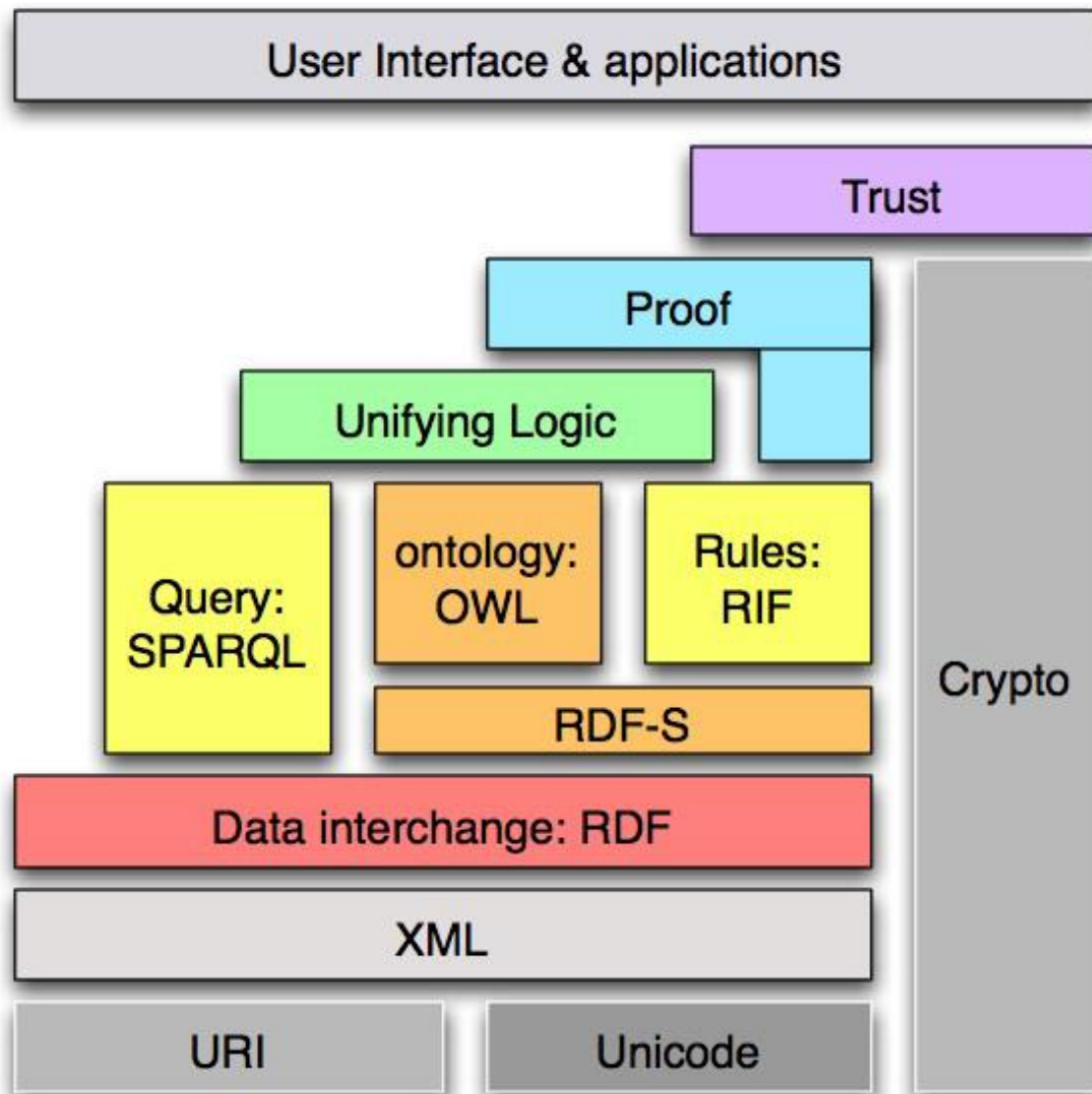


Figura 7: La web semàntica segons W3C

2.4.1. XML

XML són les sigles d'eXtensible Markup Language i va ser publicat pel W3C⁷ l'any 1998 com un llenguatge de marques extensible. Actualment, els estàndards reconeguts són el 1.0 i el 1.1, i és un format obert que utilitza l'extensió ".xml". S'utilitza per emmagatzemar dades en forma llegible i de

⁷ Definició de l'estàndard XML pel W3C: <https://www.w3.org/standards/xml/>, consultat el gener del 2019.

forma estructurada dins d'etiquetes, i divers programari el fa servir per comunicar-se entre si o integrar informació.

El seu ús té diversos avantatges, però els més destacats són la seva extensibilitat i que un tercer podria entendre i processar amb facilitat el document, fet que millora la compatibilitat entre programaris. A més, XML utilitza una definició de tipus de document (DTD) on consta l'estructura i la sintaxi del document; d'aquesta manera, el contingut de qualsevol XML pot ser validat a partir del seu DTD.

Un document XML està format per:

- **Pròleg:** encara que no és obligatori, es tracta d'unes línies on es descriu la versió d'XML, el tipus de document, l'enllaç al seu DTD i/o comentaris i instruccions referents al seu processament.
- **Cos:** d'ús obligatori, es tracta pròpiament del contingut del document. En forma jerarquizada, s'hi poden trobar elements i els atributs que els caracteritzen, encara que també hi pot haver entitats predefinides (per representar caràcters especials), seccions CDATA (contingut especial que no s'ha d'interpretar com XML) i/o comentaris.

Aquests són uns exemples d'un document XML per emmagatzemar i compartir informació sobre animals, i el seu DTD.

```
<?xml version="1.0"?>
<GOS>
  <NOM>Zoe</NOM>
  <RACA>Siamès</RACA>
  <EDAT>6</EDAT>
  <ALTERAT>1</ALTERAT>
  <OPERAT>0</OPERAT>
  <LLICENCIA>25942143543</LLICENCIA>
  <PROPIETARI>John Doe</PROPIETARI>
</GAT>
```

```
<xsd:element name="GAT">
  <xsd:complexType>
    <xsd:sequence>
      <xsd:element name="NOM" type="xsd:string"/>
      <xsd:element name="RACA" type="xsd:string"/>
      <xsd:element name="EDAT" type="xsd:positiveInteger"/>
      <xsd:element name="ALTERAT" type="xsd:boolean"/>
      <xsd:element name="DECLARAT" type="xsd:boolean"/>
      <xsd:element name="LLICENCIA" type="xsd:string"/>
      <xsd:element name="PROPIETARI" type="xsd:string"/>
    </xsd:sequence>
  </xsd:complexType>
</xsd:element>
```

2.4.2. RDF

RDF són les sigles de *Resource Description Framework* i va ser publicat pel W3C⁸ l'any 1997. Generalment, aquest tipus de documents utilitzen l'extensió ".rdf". Originalment va ser dissenyat com un model de dades per a metadades i, per tant, és molt útil per identificar-les, classificar-les, etc. Així doncs, RDF cerca flexibilitzar aspectes respecte al funcionament d'XML pel que fa a informació semàntica i és considerat un dels millors llenguatges base per representar el coneixement.

El model identifica els recursos i els descriu en termes de propietats i valors. Aquesta descripció està formada per:

- **Subjecte:** és el recurs a descriure.
- **Predicat:** són les propietats del recurs i les seves relacions.
- **Objecte:** és el valor que conté cadascuna de les propietats del recurs.

Encara que aquests components acostumen a ser representats mitjançant grafs (els subjectes i els objectes serien els nodes, i els predicats els arcs), la sintaxi RDF/XML permet emmagatzemar-los i intercanviar-los en un format textual interpretable per màquines. Un exemple podria ser el següent:

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xml:base="http://www.w3.org/XXXX/XX/animals#">

  <rdf:Description rdf:ID="animal">
    <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
  </rdf:Description>

  <rdf:Description rdf:ID="gat">
    <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
    <rdfs:subClassOf rdf:resource="#animal"/>
  </rdf:Description>
</rdf:RDF>
```

A l'exemple anterior apareixen diverses URI (*Uniform Resource Identifier*), que és un format estàndard per representar qualsevol recurs de forma única; d'aquesta manera, les URI ajuden a enllaçar els elements del document RDF amb la Web.

⁸ Relacions entre especificacions referents a RDF, segons les recomanacions dels membres de W3C: https://www.w3.org/standards/techs/rdf#w3c_all, consultat el gener del 2019.

2.4.3. RDFS

RDFS són les sigles de *RDF Schema* i és una extensió de RDF publicada pel W3C⁹ l'any 1998 i editada per Dan Brickley i Ramanathan V. Guha. Tracta d'ampliar les funcionalitats del model de l'any anterior, solucionant diverses limitacions: per exemple, permet construir esquemes i jerarquitzar la informació amb l'ús de classes, subclasses, propietats i restriccions.

Algunes de les definicions de classes més comunes són les següents: `rdfs:Resource`, `rdfs:Property`, `rdfs:Class`, `rdfs:domain`, `rdfs:range`, `rdfs:type`, `rdfs:subClassOf`, `rdfs:subPropertyOf`, `rdfs:seeAlso` i `rdfs:isDefinedBy`. Aquestes permeten jerarquitzar els conceptes descrivint els vocabularis de termes i les seves relacions.

2.4.4. OWL

OWL són les sigles d'*Ontology Web Language* i actualment hi ha dues publicades: la primera l'any 2004 per Mike Dean i Guus Schreiber, i la segona l'any 2009 pel W3C OWL Working Group¹⁰ (anomenada de forma abreviada OWL 2). Es tracta d'un llenguatge estàndard de marcat que permet representar definicions i relacions més complexes respecte als models descrits anteriorment, amb la finalitat de representar i codificar ontologies.

Per tant, considerant que RDF dona suport -prou bo- a un tractament bàsic d'informació semàntica, OWL el millora i permet a dispositius informàtics realitzar-hi tasques de raonament, comparant i combinant recursos amb diferents estructures. Com en els casos anteriors, generalment un OWL està format per un domini, diverses classes i les seves relacions.

Pel que fa específicament a l'OWL publicat l'any 2004, es coneixen tres variants:

- **OWL Lite:** pensada per donar suport a una jerarquia de classificacions i restriccions simples.
- **OWL DL:** dissenyada per oferir una màxima expressivitat mantenint la integritat computacional i perquè qualsevol sentència sigui resolta en temps finit.
- **OWL Full:** a diferència de les anteriors, aquesta variant pretén oferir la màxima expressivitat i llibertat sintàctica, però no pot garantir la computacionalitat.

⁹ RDF Schema 1.1, definit pel W3C: <https://www.w3.org/TR/rdf-schema/>, consultat el gener del 2019.

¹⁰ Relacions entre especificacions referents a OWL, segons les recomanacions dels membres de W3C (el grup *OWL Working Group* es va tancar anys enrere): https://www.w3.org/standards/techs/owl#w3c_all, consultat el gener del 2019.

2.4.5. SPARQL

SPARQL són les sigles de *Simple Protocol And RDF Query Language*, es tracta d'un estàndard per la consulta d'informació basada en RDF (és a dir, un llenguatge de consulta semàntica per ontologies com RDFS o OWL) i va ser publicat oficialment pel W3C¹¹ l'any 2008. A més, en 2013 es publica la versió 1.1¹² i és considerat una de les tecnologies clau pel que fa a la web semàntica.

Tot i que la sintaxi que utilitza és semblant a altres llenguatges de consulta com SQL, SPARQL també és considerat un protocol per accedir a dades RDF. El seu funcionament es basa en, per exemple, interrogar un RDF utilitzant triplets (subjectes, predicats i objectes) i recursos, per obtenir els resultats que hi ha a l'ontologia i que coincideixen amb els arguments de la consulta.

Per exemple, amb aquesta consulta es podria realitzar una cerca dels animals existents a un RDF:

```
SELECT DISTINCT ?Animals ?nombre
WHERE {
  ?Animals skos:subject <http://dbpedia.org/resource/Category:Animals> .
  ?Animals rdfs:label ?nombre .
}
```

2.4.6. GeoSPARQL

GeoSPARQL¹³ es tracta d'un estàndard presentat per OGC¹⁴, que amplia el marc d'extensibilitat de SPARQL per tal d'aconseguir consultar dades geoespaciales, Està dissenyat per ser utilitzat per ontologies que disposen de dades de coordenades de latitud i longitud, representar geometries a partir d'elles i trobar-hi les seves relacions geoespaciales. Conté els seus propis tipus de dades, funcions, etc.

```
SELECT ?Veterinari ?Barri
WHERE {
  ?Veterinari a ex:Veterinari ;
    :hasGeometry ?vgeo .
  ?Barri a ex:Barri ;
    geo:hasGeometry ?bgeo .
  ?vgeo geo:sfWithin ?bgeo .
}
```

¹¹ SPARQL Query Language for RDF: <https://www.w3.org/TR/rdf-sparql-query/>, consultat el gener del 2019.

¹² Definició de la versió 1.1 de SPARQL: <https://www.w3.org/TR/sparql11-overview/>, consultat el gener del 2019.

¹³ Definició de l'estàndard GeoSPARQL per part del OGC: <https://www.opengeospatial.org/standards/geosparql>, consultat el gener del 2019.

¹⁴ Les seves sigles corresponen a *Open Geospatial Consortium* i es tracta d'una organització internacional sense ànim de lucre, que vetlla pel desenvolupament i la implementació d'estàndards lliures per al contingut geoespacial.

2.5. Linked Data

El concepte Linked Data (dades enllaçades o dades vinculades, en català) fa referència a la publicació de dades estructurades que poden ser interconnectades, amb la finalitat de compartir informació computable informàticament i permetre la seva accessibilitat des de qualsevol aplicació. Així doncs, un determinat projecte pot consultar directament els repositoris Linked Data amb l'objectiu d'incorporar-hi totes les seves dades.

L'any 2006, Tim Berners-Lee va definir quatre principis bàsics pel que fa a Linked Data durant la seva presentació pel W3C:

1. Cal utilitzar URIs per identificar tots els recursos que hi ha a la Web.
2. De la mateixa manera, cal aprofitar l'HTTP de cada URI per localitzar-los.
3. S'ha de proporcionar informació útil del recurs mitjançant els estàndards reconeguts.
4. És convenient incloure altres URI relacionades amb els continguts del recurs, amb la finalitat que els usuaris trobin més informació a la Web.

El diagrama següent mostra quins conjunts de dades Linking Open Data estaven interconnectats l'agost de 2014:

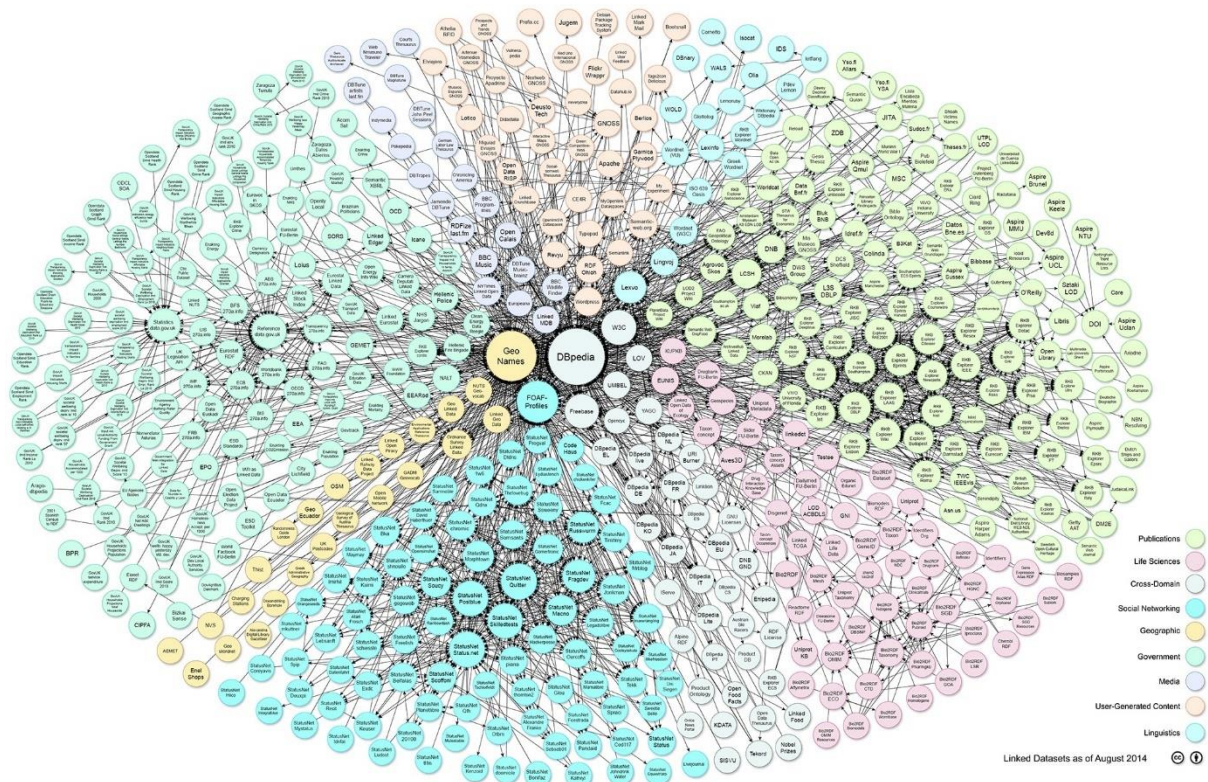


Figura 8: Conjunts de dades Linking Open Data interconnectats l'any 2014 (Linked Open Data Cloud)

Al centre del diagrama es troba *DBpedia*, que es tracta d'un projecte que extreu dades estructurades de la *Wikipedia* amb la finalitat d'enllaçar-les amb altres conjunts de dades existents a la Web. D'altra

banda, Linkeddata.org també és un dels projectes més reconeguts de Linked Data i destaca pel seu enorme volum d'interrelacions, fet que afavoreix navegar d'un concepte a un altre en qualsevol domini.

2.6. Open Data

El concepte Open Data (Dades Obertes, en català) no és pas un llenguatge o un estàndard: es tracta d'un corrent que cerca que determinats tipus de dades siguin accessibles arreu del món de forma lliure, sense sol·licitar cap mena de permís i sense trobar cap mena de restricció (drets d'autor, patents, mecanismes de control...). Per tant, ajuden a fomentar valors com la transparència i la igualtat.

Així doncs, la diferència entre Open Data i Linked Data resideix en el fet que la primera fa referència a llicències obertes i la segona en la interconnexió de les dades; a més, la unió d'ambdues es coneix com a Linked Open Data i un exemple és el descrit a la figura 8. D'específics pel que fa a Open Data, afortunadament s'hi troben cada cop més a la Web: administracions públiques de diferents països d'arreu del món, transports públics de grans ciutats, etc.

Aquests són alguns dels formats més freqüents en els que els repositoris d'Open Data es troben disponibles per la seva descarrega: CSV, GeoRSS, GPX, JSON, KML, KMZ, MDB, RDF, RSS, SHP, TXT, WMS, XLS, entre d'altres.

2.6.1. CSV

CSV són les sigles de *Comma-Separated Values* i es tracta d'un tipus de document senzill que es troba sense estandarditzar de forma completa i oficial (encara que a l'*RFC 4180*¹⁵ s'hi troben definides les seves especificacions actuals des del 2005).

La idea principal és la de contenir de forma organitzada la informació d'una base de dades, separant els elements al document mitjançant caràcters delimitadors: comes (","), punts i comes (";"), tabulacions, salts de línia, etc.

Un exemple de contingut d'un document CSV on els caràcters delimitadors fossin el punt i coma i el salt de línia, seria el següent:

```
NOM;RACA;EDAT;ALTERAT;OPERAT;LLICENCIA;PROPIETARI;
"Zoe";"Siamès";6;1;0;25942143543;"John Doe";
```

¹⁵ Especificacions RFC 4180 per documents CSV: <https://tools.ietf.org/html/rfc4180>, consultat el gener del 2019.

2.7. GTFS

GTFS són les sigles de *General Transit Feed Specification* i va ser publicat per Google l'any 2006. Generalment, aquest tipus de documents són una extensió dels CSV, es tracten d'un format obert i defineixen un format comú per als horaris del transport públic i la informació geogràfica associada.

Així doncs, els *feeds* GTFS permeten que les empreses de transport públic publiquin les seves dades de transport amb la finalitat que programari propi o de tercers les utilitzi de manera interoperable. Un *feed* GTFS acostuma a ser manegat com un arxiu ZIP que conté els següents arxius de text:

- **agency.txt**: es tracta de la informació sobre l'empresa de transport (nom, lloc web, informació de contacte, etc.).
- **stops.txt**: correspon a la informació de les estacions del sistema de transport.
- **routes.txt**: es tracta de l'arxiu on s'emmagatzemen les diferents rutes.
- **trips.txt**: correspon a la informació respectiva als viatges i a quina ruta i servei pertanyen.
- **stop_times.txt**: es tracta del temps d'espera entre aturades dels vehicles del sistema de transport.
- **calendar.txt**: correspon als patrons de servei que funcionen de manera recurrent.
- **calendar_dates.txt**: a diferència de l'anterior, aquest arxiu emmagatzema informació sobre esdeveniments especials (dates úniques, festivitats, etc.).
- **fare_attributes.txt**: correspon a la informació sobre les tarifes corresponents a les rutes descrites prèviament.
- **fare_rules.txt**: es tracten de les regles pertanyents a les tarifes esmenades en el punt anterior.
- **shapes.txt**: correspon a les regles per dibuixar les rutes del sistema de transport en un mapa.
- **frequencies.txt**: es tracta del temps entre viatges per a rutes amb freqüència variable.
- **transfers.txt**: corresponen a les connexions en punts de combinació entre rutes.
- **feed_info.txt**: es tracta d'informació addicional sobre el *feed* (editor, versió, data de venciment, etc.).

Només els sis primers descrits són obligatoris i cal forçar el contingut dels arxius a complir les següents regles de format:

- Tots els arxius han de ser emmagatzemats en format CSV.
- La primera línia de cada arxiu fa referència als noms dels camps.
- El sistema distingeix entre majúscules i minúscules.
- Els valors de cada camp no poden incloure tabulacions, salts de línia, etiquetes HTML, espais addicionals, etc.
- Si algun d'aquests valors conté cometes o comes, cal encapsular-los entre cometes.
- Cada línia ha de finalitzar amb un salt de línia CRLF o LF.
- La codificació dels arxius ha de ser UTF-8 per admetre tots els caràcters Unicode.

3. Anàlisi i disseny del projecte

3.1. Anàlisi inicial

D'una banda, l'ontologia a desenvolupar haurà de contenir la informació relativa a les estacions dels següents mitjans de transports públics de Madrid:

- M4: Metro.
- M5: Rodalies.
- M6: EMT (xarxa d'autobusos).
- M8: Autobusos interurbans.
- M9: Autobusos urbans.
- M10: Metro lleuger / Tramvia.

D'altra banda, estudiant els informes del darrer *Observatorio de la Ciudad* (2017)¹⁶, s'han considerat els següents punts com els més rellevants per assolir els objectius del present treball:

- Mobilitat i transport públic.
- Salut i serveis sanitaris.
- Educació i centres educatius.
- Facilitat per fer esport.
- Oferta cultural.
- Oci i diversió.
- Els aparcaments públics municipals.
- Facilitat per aparcar.
- Els teatres municipals.
- Les biblioteques municipals.
- Les instal·lacions esportives municipals.
- Els centres culturals municipals.
- Els serveis socials.
- Les escoles infantils públiques.
- La informació i promoció turística.
- Els mercats municipals.

¹⁶ Percepció ciutadana al *Observatorio de la Ciudad* realitzat l'any 2017:

<https://www.madrid.es/portales/munimadrid/es/Inicio/El-Ayuntamiento/Observatorio-de-la-Ciudad/Percepcion-Ciudadana/Edicion-2017/?vgnnextfmt=default&vgnextoid=ba643225968b1610VgnVCM2000001f4a900aRCRD&vgnextchannel=f22ff49c4495d310VgnVCM2000000c205a0aRCRD>, consultat el gener del 2019.

Per tant, consultant els repositoris de dades que Madrid ofereix de forma pública i gratuïta, s'ha decidit que els següents seran tractats per nodrir l'ontologia:

	Sector	Freqüència d'actualització
Aparcamientos públicos municipales	Urbanisme i infraestructures	Anual
Aparcamientos residentes P.A.R. y aparcamientos públicos mixtos	Urbanisme i infraestructures	Anual
Bibliotecas y bibliobuses en la ciudad de Madrid	Cultura i oci	Anual
Cementerios de Madrid	Societat i benestar	Anual
Centros Culturales Municipales (incluyen Socioculturales y Juveniles)	Cultura i oci	Anual
Centros de Atención para Menores y Familia	Societat i benestar	Anual
Deportes. Centros Deportivos Municipales (Polideportivos)	Esport	Anual
Divisiones administrativas: distritos, barrios y divisiones históricas	Urbanisme i infraestructures	Anual
Edificios de carácter monumental	Cultura i oci	Anual
Embajadas y Consulados	Urbanisme i infraestructures	Anual
Mercados Municipales	Comerç	Anual
Museos de la ciudad de Madrid	Cultura i oci	Anual
Salas de espectáculos artísticos: teatros, cines, filmotecas, auditorios y salas de conciertos	Cultura i oci	Anual
Salas de estudio y lectura	Educació	Anual
Sedes. Bibliotecas Especializadas Universitarias y Nacionales	Cultura i oci	Anual
Sedes. Centros de Atención Médica	Societat i benestar	Anual
Sedes. Centros de Atención Social	Societat i benestar	Anual
Sedes. Centros de Educación	Educació	Anual
Sedes. Centros de Enseñanza	Educació	Anual
Universidades, colegios mayores, residencias universitarias y otros	Educació	Anual

Taula 2: Repositoris de dades escollits de la ciutat de Madrid, amb el seu títol literal

3.2. Disseny de l'ontologia

Per tal de dissenyar i crear l'ontologia, i considerant els aspectes comentats en capítols anteriors, s'ha decidit utilitzar la metodologia de la Universitat de Stanford realitzada per Noy i McGuinness (2001):

1. **Determinar el domini i l'abast de l'ontologia:** algunes de les preguntes formulades que l'ontologia hauria de respondre són:
 - Quines zones de Madrid tenen més i menys elements de transport?
 - Els centres i serveis públics de la ciutat disposen dels suficients elements de transport?
 - Hi ha algun tipus de relació entre les zones on més elements de transport hi ha vers menys àrees d'aparcament de vehicles habilitades?
 - Com a turista, a quins elements de transport s'han de dirigir per a arribar a un determinat monument, museu o altre lloc d'interès turístic?
 - Hi ha alguna oficina pública on només es pugui arribar fàcilment mitjançant un element de transport?
 - Hi ha diferències notables entre el nombre d'elements de transport propers a cadascuna de les seues administratives de la ciutat?
2. **Considerar la reutilització d'ontologies existents:** realitzant una recerca en profunditat de solucions existents, no s'ha trobat cap que presenti similituds aprofitables. Per exemple, utilitzant el cercador Swoogle¹⁷ apareixen diversos resultats sota la cerca "Madrid", però cap referent a la ubicació d'elements d'interès públic de la ciutat i les estacions del transport públic.
3. **Enumerar els termes rellevants de l'ontologia:** districte, element d'interès i estacions de mitjans de transport públic seran els tres pals de paller de l'ontologia, tot i que hi ha altres elements més específics que dependran d'ells.
4. **Definir les classes i la seva jerarquia.**
5. **Definir les propietats de les classes.**
6. **Definir les restriccions de les propietats.**
7. **Crear instàncies.**

Considerant la rellevància pràctica dels darrers quatre punts de l'anterior guia, seran descrits amb més profunditat a la resta del capítol i al següent.

D'altra banda, cal destacar que s'ha decidit utilitzar l'idioma castellà per la nomenclatura dels elements que formen l'ontologia, ja que un dels objectius d'aquest projecte és crear-ne una que pugui ser publicada a Internet i reutilitzada en un futur, i l'idioma oficial de la Comunidad de Madrid és el castellà.

¹⁷ Swoogle és un motor de cerca per a ontologies, documents, termes i dades publicades a Internet: <http://swoogle.umbc.edu/2006/>, consultat el gener del 2019.

3.2.1. Classes

Com s'ha definit anteriorment, les classes corresponen a les representacions dels conceptes del domini en qüestió. En aquest cas, cal crear-ne les següents juntament amb les seves respectives subclasses:

- Distrito
- Elemento
 - Aparcamientos_públicos_municipales
 - Aparcamientos_residentes
 - Auditorios_y_salas_conciertos
 - Bibliotecas_Especializadas
 - Bibliotecas_y_bibliobuses
 - Cementerios
 - Centros_Atención_Menores_Familia
 - Centros_Atención_Médica
 - Centros_Atención_Social
 - Centros_Culturales_Municipales
 - Centros_Deportivos_Municipales
 - Centros_Educación
 - Centros_Enseñanza
 - Cines_y_filmotecas
 - Embajadas_y_Consulados
 - Mercados_Municipales
 - Monumentos
 - Museos
 - Salas_estudio_y_lectura
 - Teatros
 - Universidades
- Estación
 - M4
 - M5
 - M6
 - M8
 - M9
 - M10

Les subclasses hereten els camps de les classes a les quals pertanyen. Cadascuna té els següents:

Classe Distrito

Camp	Tipus
idDistrito	string
nombreDistrito	string

Taula 3: Definició dels camps de la classe Distrito

Classe Elemento

Camp	Tipus
idElemento	string
nombreElemento	string
esAccesible	integer
tipoViaElemento	string
nombreViaElemento	string
númeroPortalElemento	string
plantaElemento	string
puertaElemento	string
escaleraElemento	string
códigoPostalElemento	integer
coordenadaXElemento	integer
coordenadaYElemento	integer
latElemento	double
longElemento	double

Taula 4: Definició dels camps de la classe Elemento

Classe Estación

Camp	Tipus
idEstación	string
nombreEstación	string
códigoEstación	string
tipoVíaEstación	string
nombreVíaEstación	string
númeroPortalEstación	string
códigoPostalEstación	integer
coordenadaXEstación	integer
coordenadaYEstación	integer
geo:lat	double
geo:long	double

Taula 5: Definició dels camps de la classe Estación

3.2.2. Propietats

Les propietats fan referència a quina és la relació entre les classes que formen l'ontologia. En aquest cas, s'han definit les següents:

	Domini	Rang	Inversa
contieneElemento	Distrito	Elemento	elementoSeUbicaEn
elementoSeUbicaEn	Elemento	Distrito	contieneElemento
contieneEstación	Distrito	Estación	estaciónSeUbicaEn
estaciónSeUbicaEn	Estación	Distrito	contieneEstación

Taula 6: Definició de les propietats de l'ontologia

En resum, un districte conté una sèrie d'elements d'interès i estacions de la ciutat de Madrid i aquests darrers s'ubiquen en un determinat districte, sent precisament aquesta frase la definició de la característica "Inversa" de les propietats definides.

3.2.3. Conjunt de dades

Primerament, del repositori anomenat “Divisiones administrativas: distritos, barrios y divisiones históricas”, s’han seleccionat els següents camps per ser importats a l’ontologia:

- **CODIGO DE DISTRITO:** es tracta d’un identificador únic del districte de la ciutat de Madrid.
- **NOMBRE DE DISTRITO:** correspon a la denominació específica del districte.

Dels repositoris de dades indicats anteriorment referents als elements d’interès de la ciutat de Madrid, s’han seleccionat els següents camps per ser importats a l’ontologia:

- **PK:** es tracta d’un identificador únic de l’element d’interès de la ciutat de Madrid.
- **NOMBRE:** correspon a la denominació específica de l’element.
- **ACCESIBILIDAD:** 1 si és accessible per persones minusvàlides i 0 en cas contrari.
- **NOMBRE-VIA:** es tracta del nom del carrer on es troba ubicat.
- **CLASE-VIAL:** correspon al tipus de via (avinguda, carrer, plaça...).
- **NUM:** es tracta del número pertanyent a l’adreça.
- **PLANTA:** correspon al número de planta.
- **PUERTA:** es tracta del número de porta.
- **ESCALERAS:** correspon a l’escala pertanyent a l’adreça.
- **CODIGO-POSTAL:** es tracta del codi postal on es troba ubicat l’element.
- **DISTRITO:** correspon al districte al qual pertany l’element.
- **COORDENADA-X:** coordenada de 6 caràcters sobre el pla de la ciutat (ED50¹⁸).
- **COORDENADA-Y:** coordenada de 7 caràcters sobre el pla de la ciutat (ED50).
- **LATITUD:** coordenada de latitud en el sistema de referència WGS84¹⁹.
- **LONGITUD:** coordenada de longitud en el sistema de referència WGS84.

Pel que fa als repositoris del CRTM, s’han seleccionat els següents camps per ser importats:

- **IDESTACION:** es tracta d’un identificador únic de l’estació de la ciutat de Madrid.
- **CODIGOESTACION:** correspon al codi de l’estació dins del grup del mateix mitjà de transport.
- **DENOMINACION:** es tracta del nom específica de l’estació.
- **TIPOVIA:** correspon al tipus de via (avinguda, carrer, plaça...).
- **NOMBREVIA:** es tracta del nom del carrer on es troba ubicada.
- **NUMEROPORTAL:** correspon al número pertanyent a l’adreça.
- **CODIGOPOSTAL:** es tracta del codi postal on es troba ubicada l’estació
- **DISTRITO:** correspon al districte al qual pertany l’estació.
- **X:** coordenada de 6 caràcters sobre el pla de la ciutat (ED50).
- **Y:** coordenada de 7 caràcters sobre el pla de la ciutat (ED50).

¹⁸ ED50 (European Datum 1950) és un antic sistema de referència geodèsic emprat a Europa.

¹⁹ WGS 84 (World Geodetic System 1984) és un sistema de coordenades geogràfiques mundial que permet localitzar qualsevol punt de la Terra.

- **LATITUD:** coordenada de latitud en el sistema de referència WGS84.
- **LONGITUD:** coordenada de longitud en el sistema de referència WGS84

3.3. Programari

Durant el desenvolupament pràctic del projecte, s'ha utilitzat el programari descrit a continuació.

3.3.1. Protégé

Protégé és una plataforma lliure i de codi obert que ofereix un conjunt d'eines per construir models de domini i aplicacions basades en l'adquisició de coneixement amb ontologies. El programari està produït en Java i Swing, desenvolupadors externs hi aporten *plugins* i va ser creat per la Universitat de Stanford²⁰, en col·laboració amb la Universitat de Manchester.

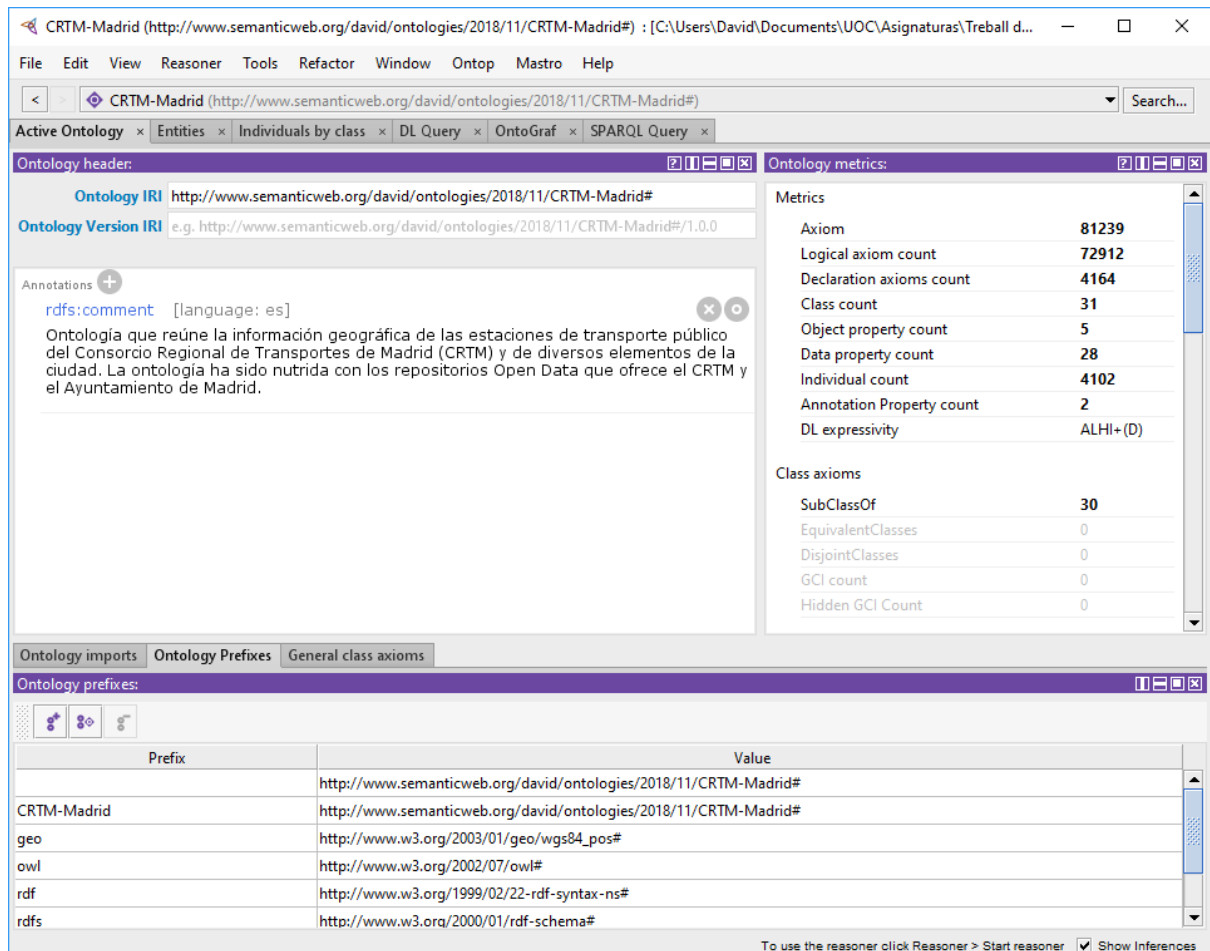


Figura 9: Pantalla d'ús del programari Protégé

²⁰ Portal web oficial de la plataforma Protégé: <https://protege.stanford.edu/>, consultat el gener del 2019.

Adicionalment, cal destacar tres funcionalitats que s'han utilitzat durant el desenvolupament del present treball:

- **HermiT²¹**: es tracta d'un raonador semàntic d'ontologies en format OWL. Aquest és capaç de crear inferències a partir de les mateixes definicions de l'ontologia, amb la finalitat d'identificar relacions i, fins i tot, determinar si el seu contingut és consistent.
- **OntoGraf²²**: correspon a un *plugin* que permet als usuaris de Protégé visualitzar les seves ontologies gràficament per, per exemple, identificar amb facilitat les classes, subclasses, relacions, etc.
- **SPARQL Query²³**: es tracta d'una funcionalitat que permet redactar i executar consultes SPARQL des de la interfície d'usuari de Protégé a una determinada ontologia.

3.3.2. WebProtégé

WebProtégé és un entorn gratuït i de codi obert per desenvolupar ontologies mitjançant un navegador web. Forma part del mateix projecte que Protégé i el programari pot ser utilitzat localment o a la plataforma que es troba allotjada a Internet.

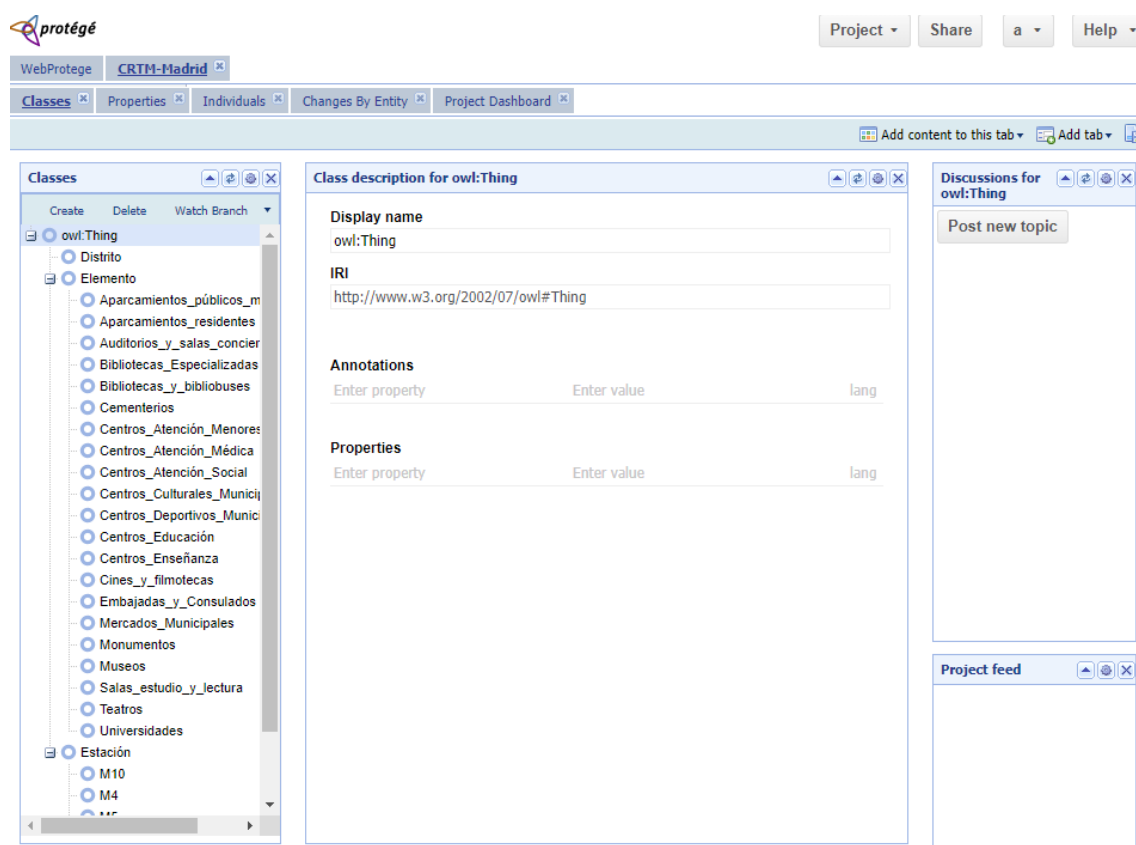


Figura 10: Pantalla d'ús del programari WebProtégé

²¹ Definició oficial del *plugin* HermiT: <https://protegewiki.stanford.edu/wiki/HermiT>, consultat el gener del 2019.

²² Definició oficial del *plugin* OntoGraf: <https://protegewiki.stanford.edu/wiki/OntoGraf>, consultat el gener del 2019.

²³ Definició oficial del *plugin* SPARQL Query: https://protegewiki.stanford.edu/wiki/SPARQL_Query, consultat el gener del 2019.

3.3.3. Virtuoso

Virtuoso²⁴ és un programari d'àmbit empresarial orientat a la gestió de web semàntica provinent de *Linked Data* de diversos repositoris de dades. Actualment combina diverses funcionalitats:

- Aplicació web per gestionar el seu contingut.
- Tractament i integració de bases de dades relacionals, bases de dades orientades a objectes, bases de dades virtuals, ontologies RDF, arxius XML, etc.
- Servidor d'arxius (Web).

El programari és desenvolupat per *OpenLink Software*, multiplataforma i la llicència és GPL i propietària. En el present treball, ha estat d'utilitat per instal·lar un servidor públic que admet consultes SPARQL a l'ontologia creada.

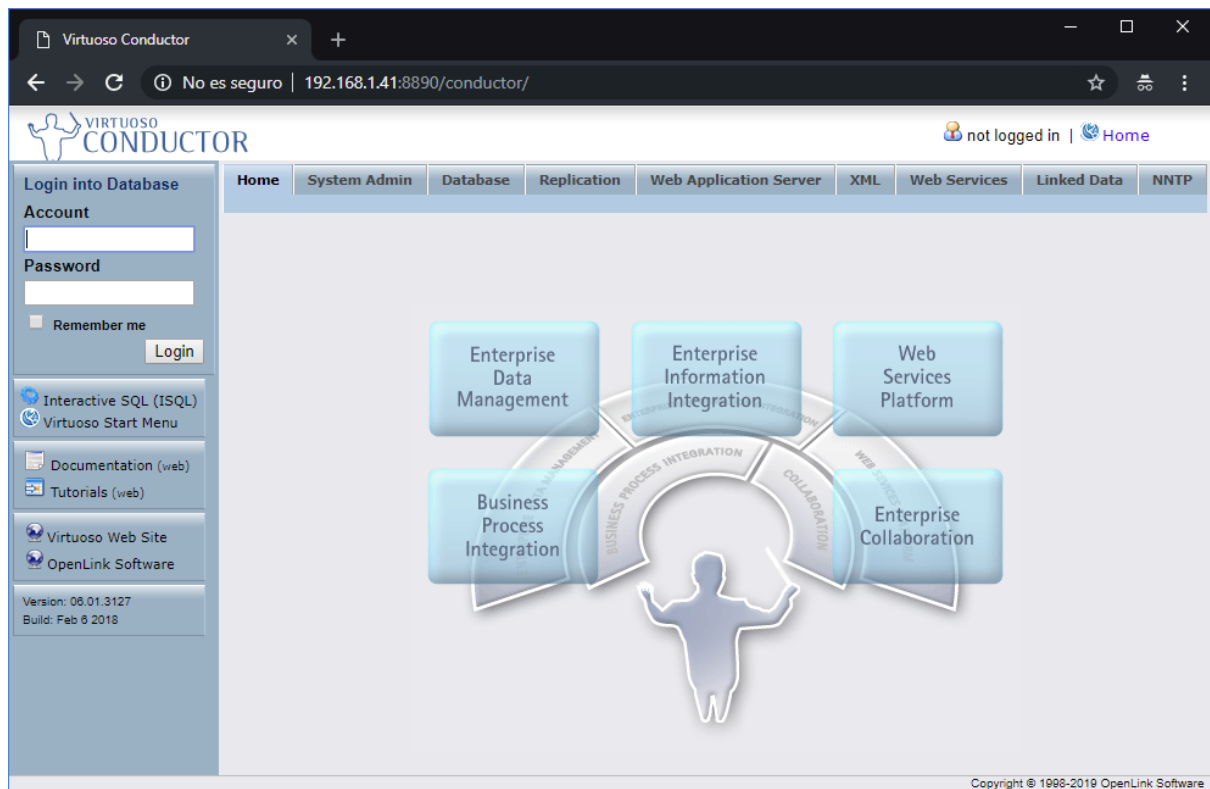


Figura 11: Pantalla d'ús del programari Virtuoso

3.3.4. EasyRdf

EasyRdf²⁵ és una llibreria PHP creada per Nicholas Humfrey sota llicència *3-Clause BSD*, que facilita la gestió i la producció d'arxius RDF. En el present treball, les funcions de la llibreria han estat d'utilitat per aconseguir que el programari desenvolupat a mida manegui l'arxiu de l'ontologia i realitzi consultes al servidor web esmentat en l'anterior subcapítol.

²⁴ Portal web oficial del programari Virtuoso: <https://virtuoso.openlinksw.com/>, consultat el gener del 2019.

²⁵ Portal web oficial de la llibreria PHP EasyRdf: <http://www.easyrdf.org/>, consultat el gener del 2019.

4. Desenvolupament del projecte

4.1. Maquinari i programari

Durant la realització de l'estat de l'art d'aquest projecte, es van poder analitzar -a gran escala- diverses eines molt útils pel seu desenvolupament. Finalment, es van escollir les que a continuació seran descrites basant-se en arguments de compatibilitat, reconeixement de la comunitat i llicenciament.

4.1.1. Requisits inicials

El projecte s'ha desenvolupat a una màquina virtual on s'hi va instal·lar Ubuntu 18.10 de 64 bits. Ubuntu²⁶ és un sistema operatiu basat en Linux, disponible de manera gratuïta i és suportat per la seva comunitat d'usuari i també per professionals del sector. La configuració recomanada per la darrera versió estable en el moment del desenvolupament del projecte, és la següent:

- **Espai de disc dur:** 25 GB.
- **Memòria RAM:** 2 GB.

4.1.2. Instruccions d'instal·lació

Primerament, ha estat necessari escollir el programari de virtualització; en aquest cas, s'ha utilitzat VirtualBox 5.2.22 sota un equip amb les següents característiques:

- **Processador:** Intel® Core(TM) i5-8400 CPU @ 2,80GHz (6 CPUs).
- **Memòria RAM:** 8 GB.
- **Espai de disc dur:** 1 TB.
- **Sistema operatiu:** Windows 10 Education 64 bits (10.0, compilació 17134).
- **Virtualització:** habilitada a la BIOS.

VirtualBox²⁷ és un programari amb llicència GPL²⁸ que s'instal·la en un sistema operatiu i en permet instal·lar-hi d'addicionals per ser executats en un entorn virtual. La seva descàrrega i instal·lació és trivial gràcies al seu assistent per entorns gràfics. A l'hora de crear la màquina virtual descrita anteriorment, la configuració destacada és la referent a l'assignació de memòria RAM i l'espai de disc dur:

²⁶ Portal web oficial del sistema operatiu Ubuntu: <https://www.ubuntu.com/>, consultat el gener del 2019.

²⁷ Portal web oficial del programari VirtualBox: <https://www.virtualbox.org/>, consultat el gener del 2019.

²⁸ GPL correspon a les sigles *General Public License* i es tracta d'una llicència de dret d'autor àmpliament usada en el món del programari lliure i codi obert, que garanteix als usuaris finals la llibertat d'usar, estudiar, compartir, copiar i modificar el programari.

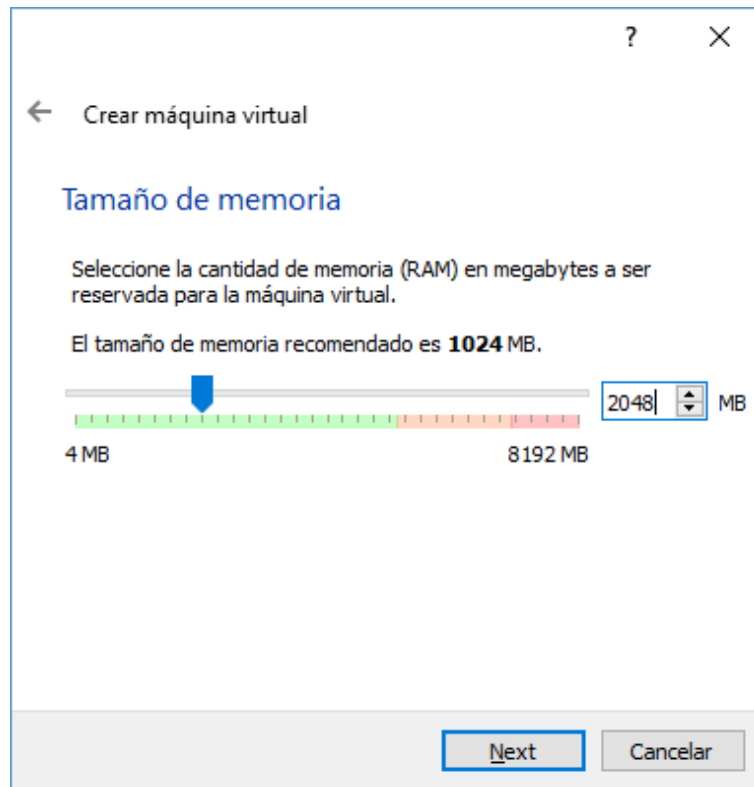


Figura 12: Memòria RAM assignada a la màquina virtual

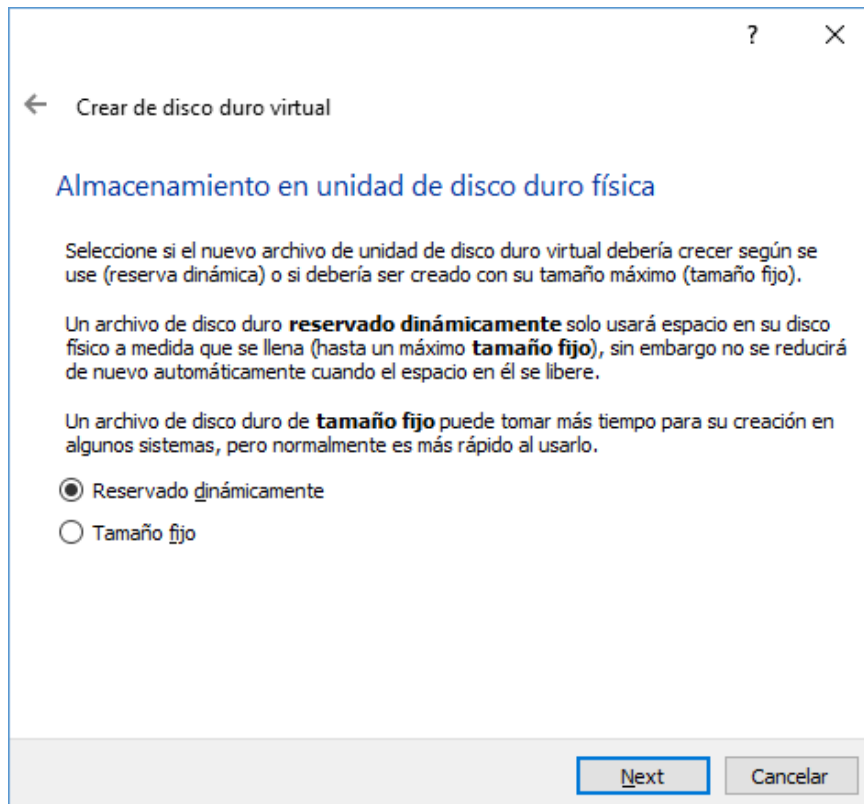


Figura 13: Tipus d'emmagatzemament del disc dur de la màquina virtual

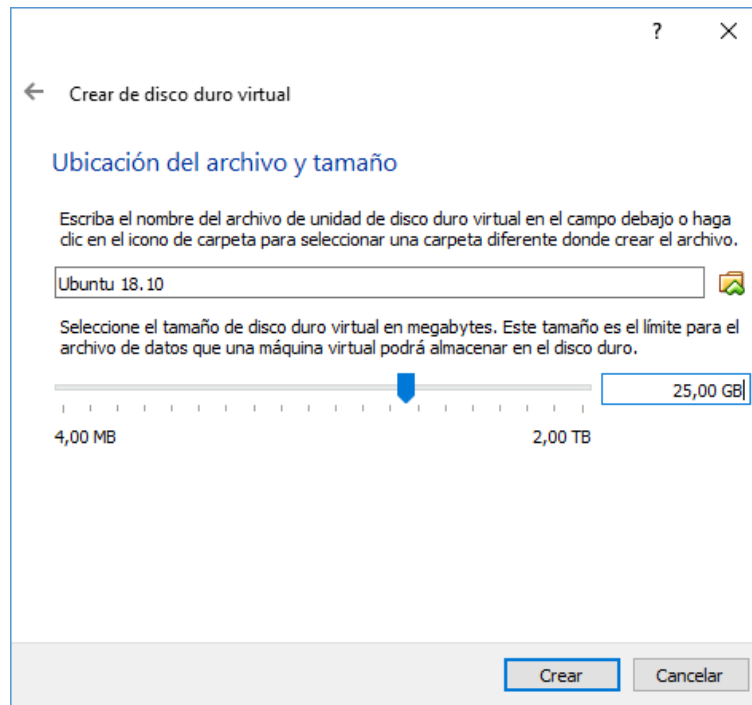


Figura 14: Memòria assignada al disc dur de la màquina virtual

Tot seguit, ha estat necessari descarregar l'arxiu ISO del sistema operatiu a ser-hi instal·lat. Ubuntu ofereix diverses versions i mètodes de descàrrega, sent escollit la versió Desktop de 64 bits mitjançant l'enllaç accessible des del mateix navegador web.

Un cop descarregat l'arxiu, cal indicar a la màquina virtual que la primera arrencada l'ha de realitzar carregant el seu contingut. Per aconseguir-ho, cal accedir fer clic al nom de la màquina, tot seguit a "Configuración", pestanya "Almacenamiento" i "Añadir unidad óptica" al controlador IDE:

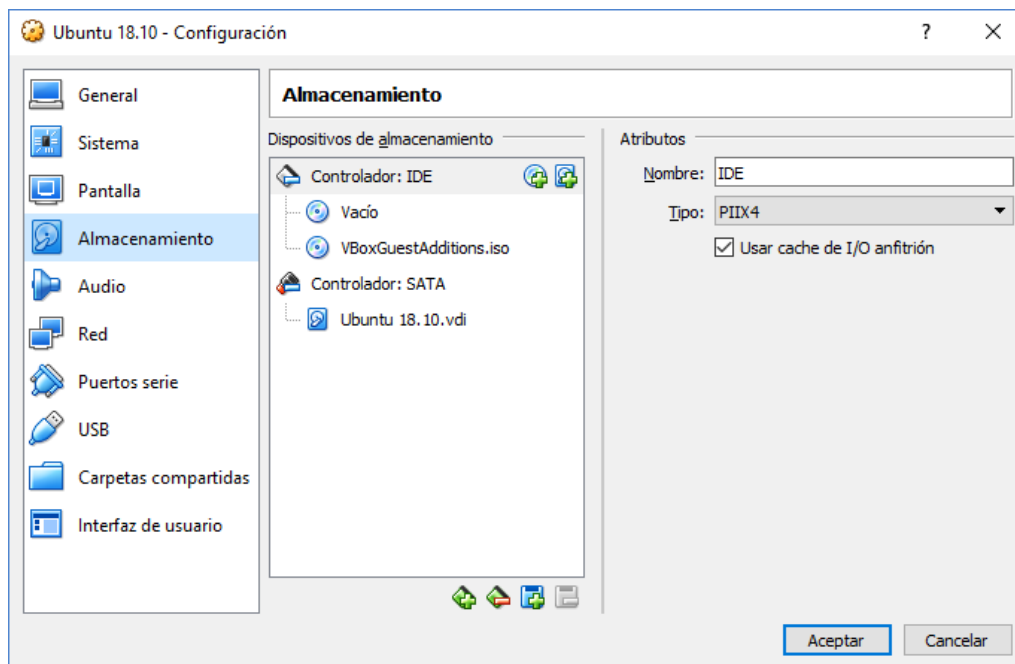


Figura 15: Assignació de la unitat òptica a la màquina virtual per arrencar l'arxiu ISO del sistema operatiu

La instal·lació del sistema operatiu també és trivial gràcies al seu assistent per entorns gràfics. A la configuració, però, s'ha escollit que s'actualitzés tot el programari disponible en finalitzar el procés d'instal·lació.

4.1.3. Protégé

La versió de Protégé utilitzada en aquest projecte és la darrera estable fins al moment: 5.2.0. Si bé és cert que la versió 5.5.0 es trobava disponible en una versió beta, s'ha considerat preferible la utilització de la versió recomanada per Protégé, que pot ser descarregada per sistemes operatius Windows, Linux, macOS i altres plataformes independents (a banda del seu codi font).

Com a requisits previs a la seva instal·lació, és necessari la instal·lació de la màquina virtual Java²⁹ i el seu JDK³⁰. Per realitzar-ho a la màquina virtual d'Ubuntu, cal iniciar una sessió de Terminal i executar la següent ordre:

```
sudo apt-get install openjdk-8-jdk openjdk-8-jre
```

Un cop descarregat el programari Protégé, cal ubicar-lo al directori desitjat i executar la següent ordre pel Terminal:

```
tar zxvf protege-5.2.0-linux.tar.gz
```

Finalment, el binari del programari es troba a l'arrel del directori creat i només és necessària la seva execució per utilitzar-lo:

```
sh run.sh
```

4.1.4. WebProtégé

Encara que en el moment del desenvolupament del projecte la darrera versió estable era la 3.0.0, aquesta presentava incompatibilitats amb funcions que es desitjaven utilitzar i, per tant, s'ha decidit instal·lar l'anterior disponible: 2.6.0.

²⁹ També coneguda per les sigles JVM, s'executa en una plataforma específica i és capaç d'interpretar i executar instruccions expressades en un codi binari especial (el *bytecode* Java), el qual és generat pel compilador del llenguatge Java.

³⁰ Correspon a les sigles *Java Development Kit* i es tracta d'un programari que proveeix eines de desenvolupament per a la creació d'aplicacions en Java.

Els requisits de programari previs a la seva instal·lació són els següents:

- Java 8 (instal·lat anteriorment).
- Apache Tomcat (o qualsevol altre contenidor de *servlets*).
- MongoDB³¹.

Pel que fa a la instal·lació dels dos darrers, cal executar les següents ordres pel Terminal:

```
sudo apt-get install tomcat8 tomcat8-admin tomcat8-common tomcat8-docs tomcat8-examples
tomcat8-user

sudo apt-get install mongodb mongodb-server mongodb-clients
```

Tot seguit, cal assignar la codificació utilitzada pel programari Tomcat a UTF8. Per aconseguir-ho, cal editar el següent arxiu de configuració:

```
$CATALINA_BASE/bin/setenv.sh
```

I afegir-hi la següent línia:

```
export CATALINA_OPTS="$CATALINA_OPTS -Dfile.encoding=UTF-8"
```

Tot seguit cal crear els següents directoris per emmagatzemar les dades de WebProtégé, la seva configuració i els seus logs, respectivament:

```
mkdir /srv/webprotege /etc/webprotege /var/log/webprotege
```

És important que ambdós directoris siguin manipulables perquè Tomcat hi pugui llegir i escriure.

Un cop aquests passos previs de configuració han quedat realitzats, ja pot ser descarregat l'arxiu WAR³² de WebProtégé i ubicar-lo al directori on s'ubiquen les aplicacions que manega Tomcat:

```
wget
https://github.com/protegeproject/webprotege/releases/download/v2.6.0/webprotege-2.
6.0.war
mv webprotege-2.6.0.war webprotege.war
```

³¹ MongoDB és un sistema de base de dades *NoSQL* orientat a documents, desenvolupat sota el concepte de codi obert.

³² Es tracta d'un arxiu JAR utilitzat per a distribuir una col·lecció de *JavaServer Pages*, *servlets*, classes Java, arxius XML, biblioteques de *tags* i pàgines web estàtiques que junts constitueixen una aplicació web.

```
mv webprotege.war /var/lib/tomcat8/webapps/
```

Finalment, és necessari iniciar el servidor Tomcat perquè WebProtégé quedi desplegat i operatiu:

```
sudo service tomcat8 start
```

A partir d'aquest moment, WebProtégé és accessible mitjançant un navegador web a l'adreça <http://localhost:8080/webprotege>.

Cal destacar que la Universitat de Stanford recomana utilitzar la versió de WebProtégé que tenen allotjada³³, eina que no s'ha pogut fer servir en aquest projecte per les incompatibilitats mencionades en paràgrafs anteriors.

4.1.5. Virtuoso

La versió de Virtuoso utilitzada en aquest projecte és la darrera estable fins al moment: 06.01.3127. Hi ha diversos mètodes d'instal·lació, però en aquest cas s'han utilitzat els paquets amb els binaris que inclou el mateix sistema operatiu. Per aconseguir-ho, cal executar les següents ordres pel Terminal.

```
apt-cache search '^virtuoso'
apt-get install virtuoso-opensource
```

Durant el procés d'instal·lació, és important definir una contrasenya pels usuaris dba i dav (encarregats de l'administració de la base de dades principal i de l'accés intern al portal web); en cas contrari, el sistema no s'iniciarà en finalitzar el procés i quedarà deshabilitat.

A partir d'aquest moment, Virtuoso és accessible mitjançant un navegador web a l'adreça <http://localhost:8890/conductor>.

4.2. Creació de l'ontologia

Els processos de creació de classes i propietats de l'ontologia s'han realitzat amb el programari Protégé; en canvi, WebProtégé ha estat utilitzat per importar els repositoris de dades. Pel que fa a aquest darrer procés, hi ha hagut una sèrie d'entrebancs respecte als formats dels arxius CSV que seran destacats a la subsecció corresponent.

³³ Portal web oficial de la plataforma WebProtégé: <https://webprotege.stanford.edu/>, consultat el gener del 2019.

4.2.1. Classes

En executar Protégé, apareix per defecte una ontologia buida per començar a treballar en la seva creació. És recomanable que un dels primers passos sigui indicar el seu nom a “Ontology IRI”, que es troba a “Ontology header” (figura 16).

En el cas concret de l'ontologia que s'ha desenvolupat, s'ha afegit un prefix³⁴ per treballar amb les coordenades de latitud i longitud en el sistema de referència WGS84. Per aconseguir-ho, cal dirigir-se a “Ontology Prefixes”, indicar el prefix i el seu valor:

The screenshot shows the Protégé ontology editor interface. The 'Ontology header' section is active, displaying the 'Ontology IRI' as <http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#> and the 'Ontology Version IRI' as [e.g. http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#/1.0.0](http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#/1.0.0). Below this, an annotation is shown: `rdfs:comment` [language: es] with the text: 'Ontología que reúne la información geográfica de las estaciones de transporte público del Consorcio Regional de Transportes de Madrid (CRTM) y de diversos elementos de la ciudad. La ontología ha sido nutrida con los repositorios Open Data que ofrece el CRTM y el Ayuntamiento de Madrid.'

The 'Ontology metrics' panel on the right shows various counts: Axiom (81239), Logical axiom count (72912), Declaration axioms count (4164), Class count (31), Object property count (5), Data property count (28), Individual count (4102), Annotation Property count (2), and DL expressivity (ALHI+(D)). It also lists Class axioms (SubClassOf: 30, EquivalentClasses: 0, DisjointClasses: 0, GCI count: 0, Hidden GCI Count: 0) and Object property axioms (SubObjectPropertyOf: 4, EquivalentObjectProperties: 0, InverseObjectProperties: 3, DisjointObjectProperties: 0, FunctionalObjectProperty: 0, InverseFunctionalObjectProperty: 0, TransitiveObjectProperty: 1).

The 'Ontology prefixes' section at the bottom shows a table of prefixes and their values:

Prefix	Value
	http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#
CRTM-Madrid	http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#
geo	http://www.w3.org/2003/01/geo/wgs84_pos#
owl	http://www.w3.org/2002/07/owl#
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs	http://www.w3.org/2000/01/rdf-schema#
xml	http://www.w3.org/XML/1998/namespace
xsd	http://www.w3.org/2001/XMLSchema#

At the bottom right, there is a checkbox for 'Show Inferences' which is checked.

Figura 16: Secció header i prefixes del programari Protégé

³⁴ https://www.w3.org/2003/01/geo/wgs84_pos#, consultat el gener del 2019.

A la pestanya “Entities” i tot seguit a la pestanya “Classes”, es troben totes les classes de l'ontologia. “owl:Thing” ja està creada i totes dependran d'aquesta; així doncs, fent-hi clic amb el botó dret del ratolí i tot seguit a “Add Subclasses”, es poden escriure totes plegades per acabar-les creant fent clic a “Continue”.

A la figura 17 es pot apreciar el llistat de classes creades i, en aquest cas, en estar seleccionada “Aparcamientos_públicos_municipales”, a la capsa “Description” apareix que aquesta és una subclasse de “Elemento”:

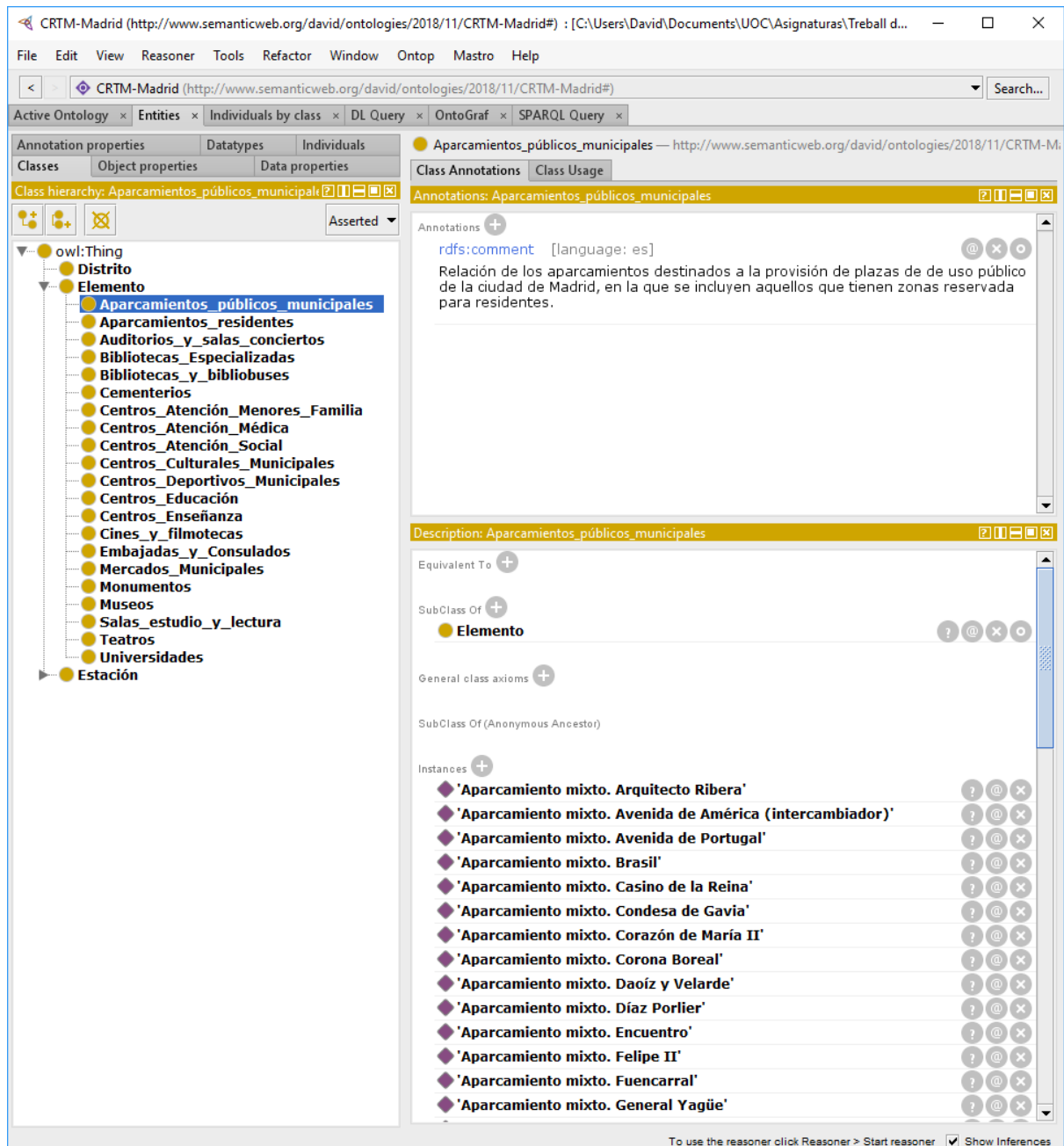


Figura 17: Llistat de classes creades a l'ontologia

4.2.2. Data properties

Tot seguit, s'han creat els camps corresponents a les classes de l'ontologia. Per aconseguir-ho, cal fer clic a "Data properties" on apareix per defecte "owl:topDataProperty" i seguir els mateixos passos descrits anteriorment: fer-hi clic amb el botó dret del ratolí, tot seguit a "Add Sub-properties" i escriure tots plegats a la capsa de text de la nova finestra oberta.

En aquest cas, però, és necessari indicar a cadascun dels camps el seu domini i el seu rang. Seleccionant-ne el desitjat, a la secció "Description" apareixen buides les característiques "Domains (intersection)" i "Ranges": fent-hi clic al pictograma "+" que les acompanya, el programari permet indicar a quines classes pertany i quin tipus de valors admet.

A la figura 18 es pot comprovar la llista de camps creada i, en el cas de "idDistrito" que és el que es troba seleccionat, també que el seu domini és "Distrito" i el seu rang és "string":

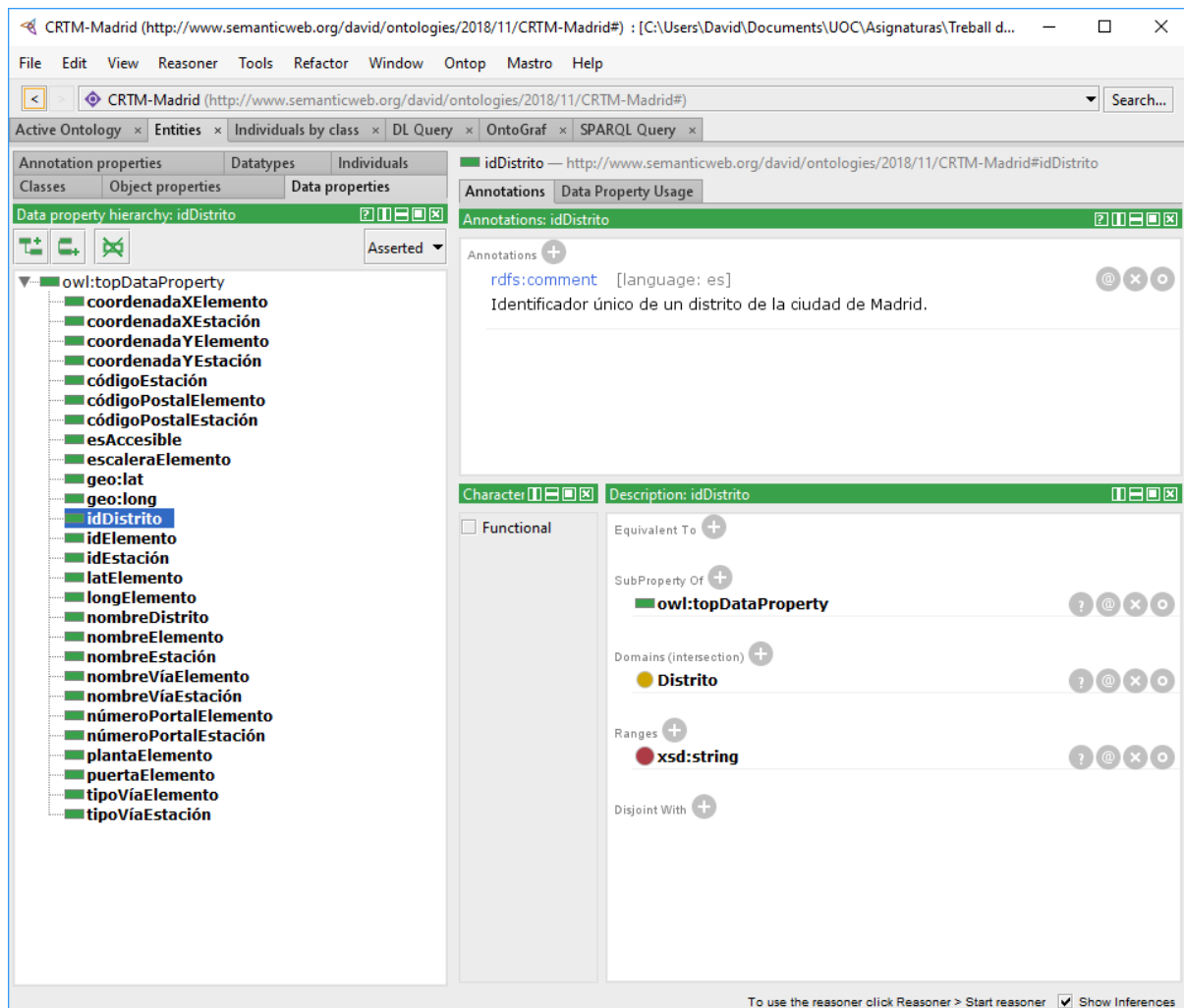


Figura 18: Llistat de *data properties* creats a l'ontologia

4.2.3. Object properties

Finalment, s'han creat les propietats de les classes de l'ontologia, que es tracten de les relacions que les vinculen. Per aconseguir-ho, cal fer clic a "Object properties" on apareix per defecte "owl:topObjectProperty" i seguir els mateixos passos descrits anteriorment: fer-hi clic amb el botó dret del ratolí, tot seguit a "Add Sub-properties" i escriure totes plegades per acabar-les creant fent clic a "Continue".

Tot seguit cal indicar de forma individual quin és el domini, el rang i la inversa de cadascuna. Seleccionant-ne la desitjada, a la secció "Description" apareixen buides les característiques "Inverse Of", "Domains (intersection)" i "Ranges (intersection)": fent-hi clic al pictograma "+" que les acompanya, el programari permet definir cadascuna de les relacions.

Com es pot comprovar a la següent figura, les quatre relacions definides apareixen a llista ubicada al cantó esquerre i, al bloc "Description", es troba tota la definició de la propietat "contieneEstación":

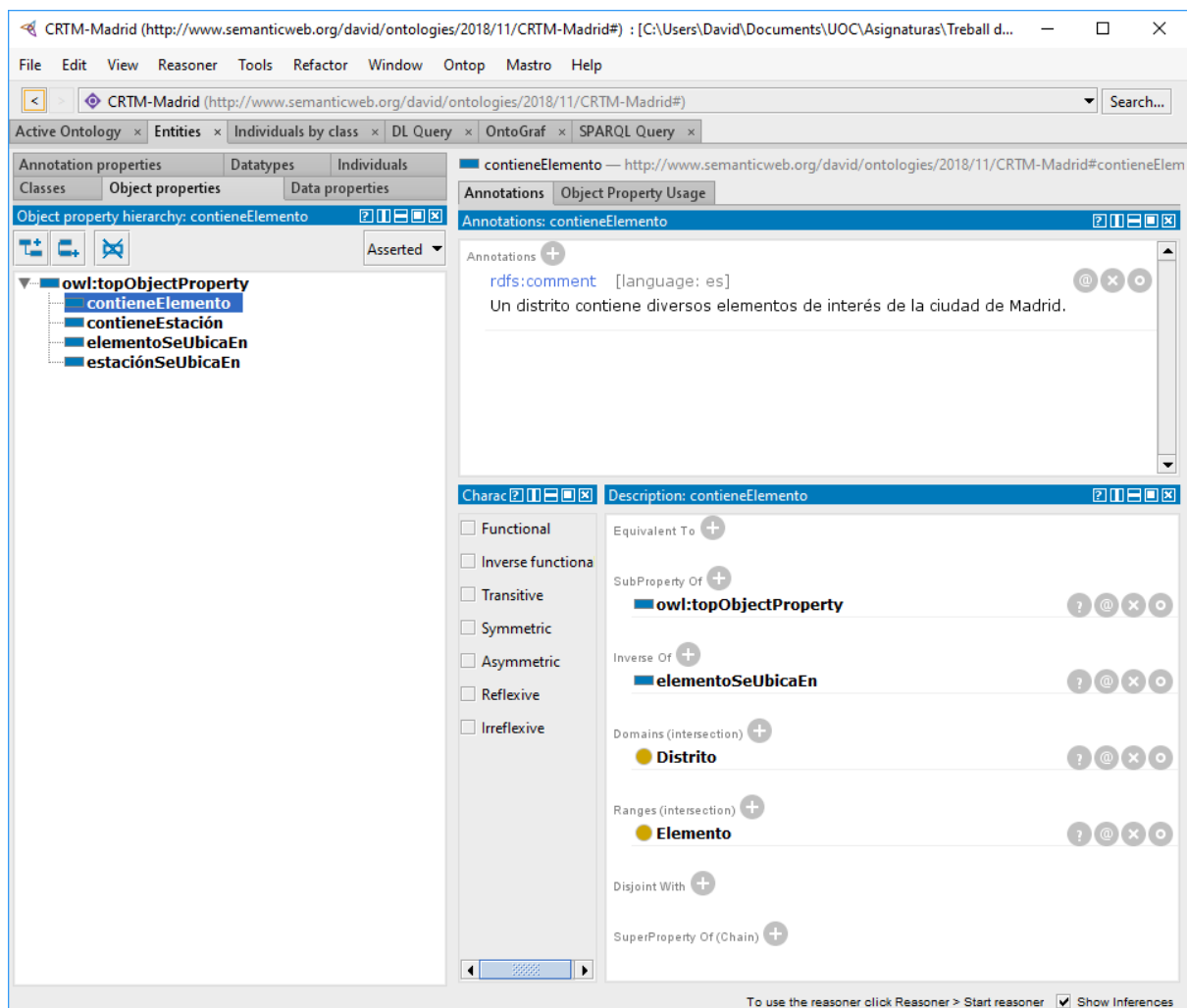
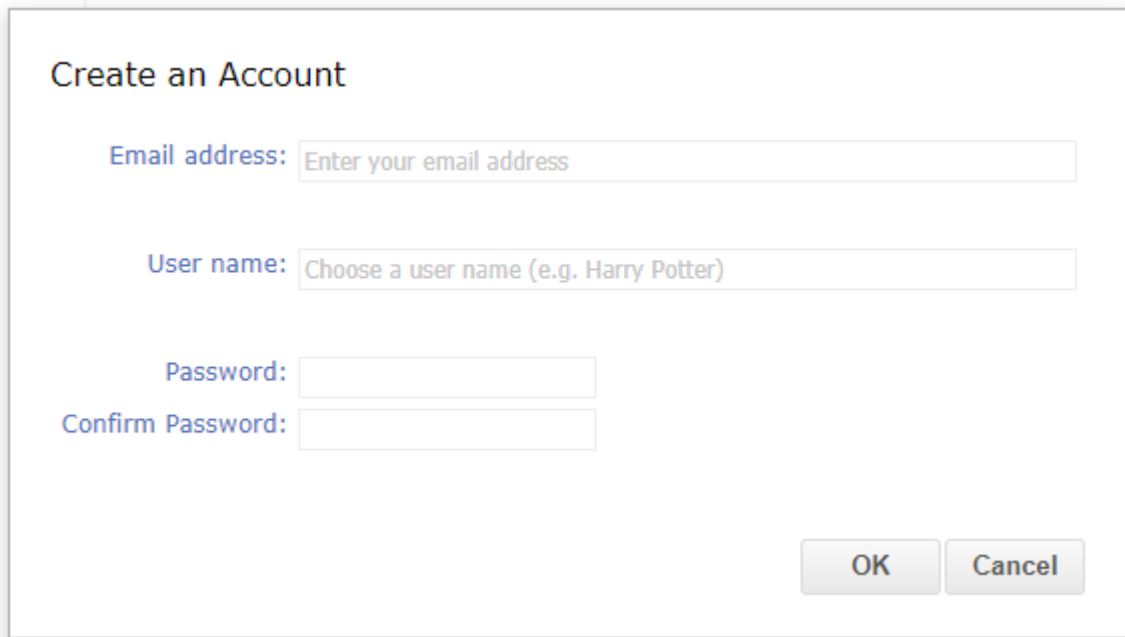


Figura 19: Llistat d'object properties creats a l'ontologia

Accedint-hi mitjançant qualsevol navegador web, el primer pas consisteix en la creació d'un usuari que emmagatzemarà les seves ontologies:



The image shows a 'Create an Account' dialog box. It has a title bar at the top. Below the title, there are four input fields: 'Email address' with a placeholder 'Enter your email address', 'User name' with a placeholder 'Choose a user name (e.g. Harry Potter)', 'Password', and 'Confirm Password'. At the bottom right, there are two buttons: 'OK' and 'Cancel'.

Figura 22: Creació d'un compte d'usuari per utilitzar el programari WebProtégé

Després d'iniciar-hi sessió, a la part superior del programari es troba el botó anomenat "Upload Project..."; fent-hi clic, el sistema permet importar-lo després d'indicar-li un nom:



The image shows an 'Upload ontology' dialog box. It has a title bar at the top. Below the title, there are two input fields: 'Project name' with the value 'CRTM-Madrid' and 'Project description' which is empty. Below these, there is a 'File' section with a button labeled 'Seleccionar archivo' and the text 'crtm-madrid.owl'. At the bottom right, there are two buttons: 'OK' and 'Cancel'.

Figura 23: Importació de l'ontologia a WebProtégé

Tot seguit, WebProtégé mostra l'ontologia creada utilitzant un estil molt similar al programari Protégé:

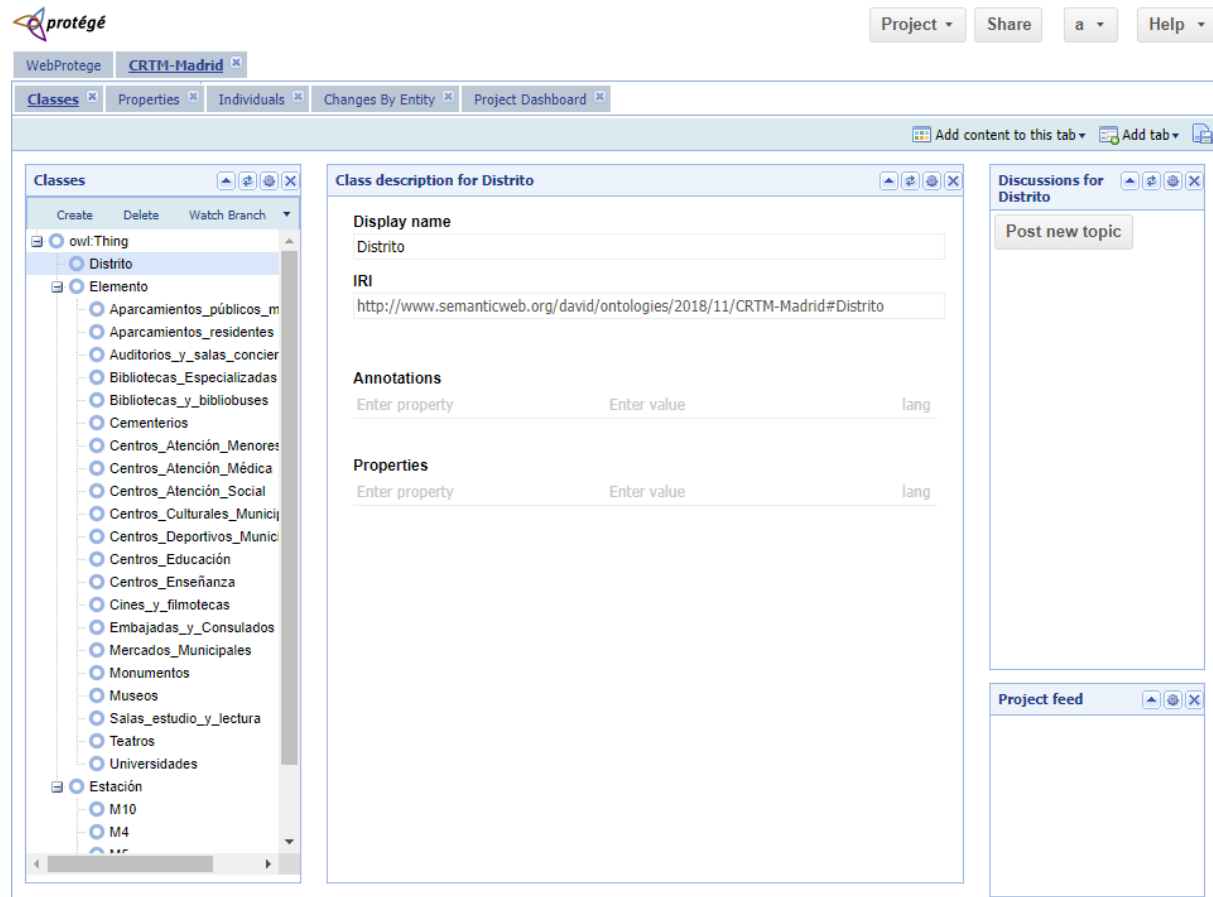


Figura 24: Ontologia importada correctament a WebProtégé

Per aconseguir crear les instàncies, cal seleccionar la classe desitjada, prémer la tecla “Shift” del teclat i fer clic al botó “Create” (que es troba a sobre de la llista de classes), obtenint un quadre de diàleg que demana seleccionar el repositori en qüestió:



Figura 25: Selecció de l'arxiu CSV a importar utilitzant WebProtégé

Tot seguit, s'obre una nova finestra de la qual cal destacar els següents elements:

- “Import rows as” es refereix a com seran afegides les dades a l'ontologia. Permet les opcions de “Classes” i “Individuals”, sent aquesta darrera la utilitzada en aquest projecte.
- “Display name column” fa referència a l'etiqueta que tindrà cada individual de l'ontologia. En aquest cas, s'ha escollit que siguin les mateixes denominacions de les instàncies.
- A “Property value mapping” cal indicar quines columnes del document corresponen als camps de la classe en qüestió -a més del seu tipus-. En el cas de les columnes que relacionen unes classes amb altres (en aquest cas, per exemple, el districte al repositori d'estacions), a “Property name” cal indicar el nom de la relació i a “Column type” “Individual”.

Import CSV File

Import rows as Individual ▼

Display name column Column 2 ▼

Property value mapping

Column	Property name and column type
Column 1	idDistrito String ▼
Column 2	nombreDistrito String ▼

idDistrito	nombreDistrito
01	CENTRO
02	ARGANZUELA
03	RETIRO
04	SALAMANCA
05	CHAMARTIN
06	TETUAN

OK Cancel

Figura 26: Procés d'importació de l'arxiu CSV a WebProtégé

A la part inferior de la figura 26, es pot comprovar com WebProtégé mostra un petit conjunt de dades per mostrar com importarà l'arxiu CSV. Gràcies a aquesta funcionalitat, s'han pogut detectar errades greus de format als repositoris que ofereix la ciutat de Madrid, com per exemple:

- Cadenes de text sense encapsular entre cometes.
- Salts de línia en cadenes de text.

- Comes en diferents camps del document, sent aquest el seu propi caràcter delimitador.

Així doncs, ha estat necessari descarregar els repositoris de dades en format JSON per tractar-los convenientment i convertir-los a CSV mitjançant un servei de tercers.

Finalment, l'ontologia pot ser descarregada des de la llista de projectes de l'usuari a WebProtégé, per tornar a treballar amb ella utilitzant el programari Protégé (cal destacar que aquest darrer és més estable). El sistema permet diversos formats de descàrrega i, en aquest, s'ha utilitzat "OWL/XML".

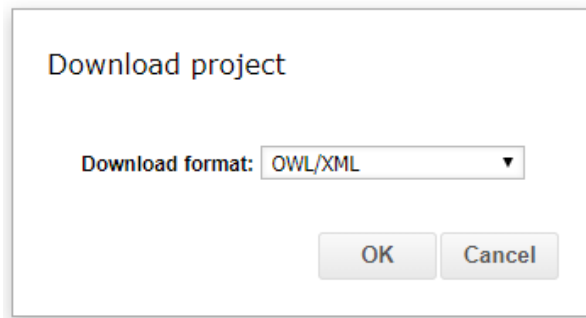


Figura 27: Elecció de format per descarregar l'ontologia utilitzant WebProtégé

Obrint amb Protégé l'arxiu generat, es pot comprovar a la finestra de benvinguda que a l'ontologia hi ha més de 150.000 axiomes, fet que provoca que el programari no funcioni correctament.

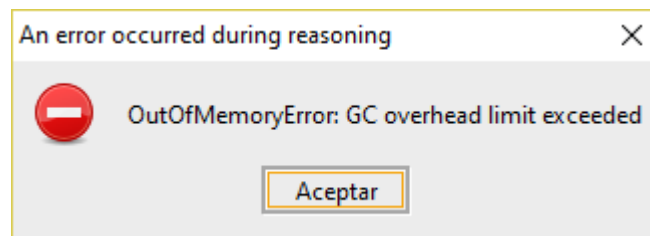


Figura 28: Missatge d'error en executar el raonador semàntic sense la memòria necessària

Repetint-ho amb una versió on només hi ha al voltant de 80.000 i fent clic al menú "Reasoner" i "Start reasoner", el raonador semàntic verifica la consistència de l'ontologia i mostra les inferències creades; per exemple, a la figura 29 es poden comprovar totes les estacions i elements d'interès de Madrid que es troben al districte 01 (Centro):

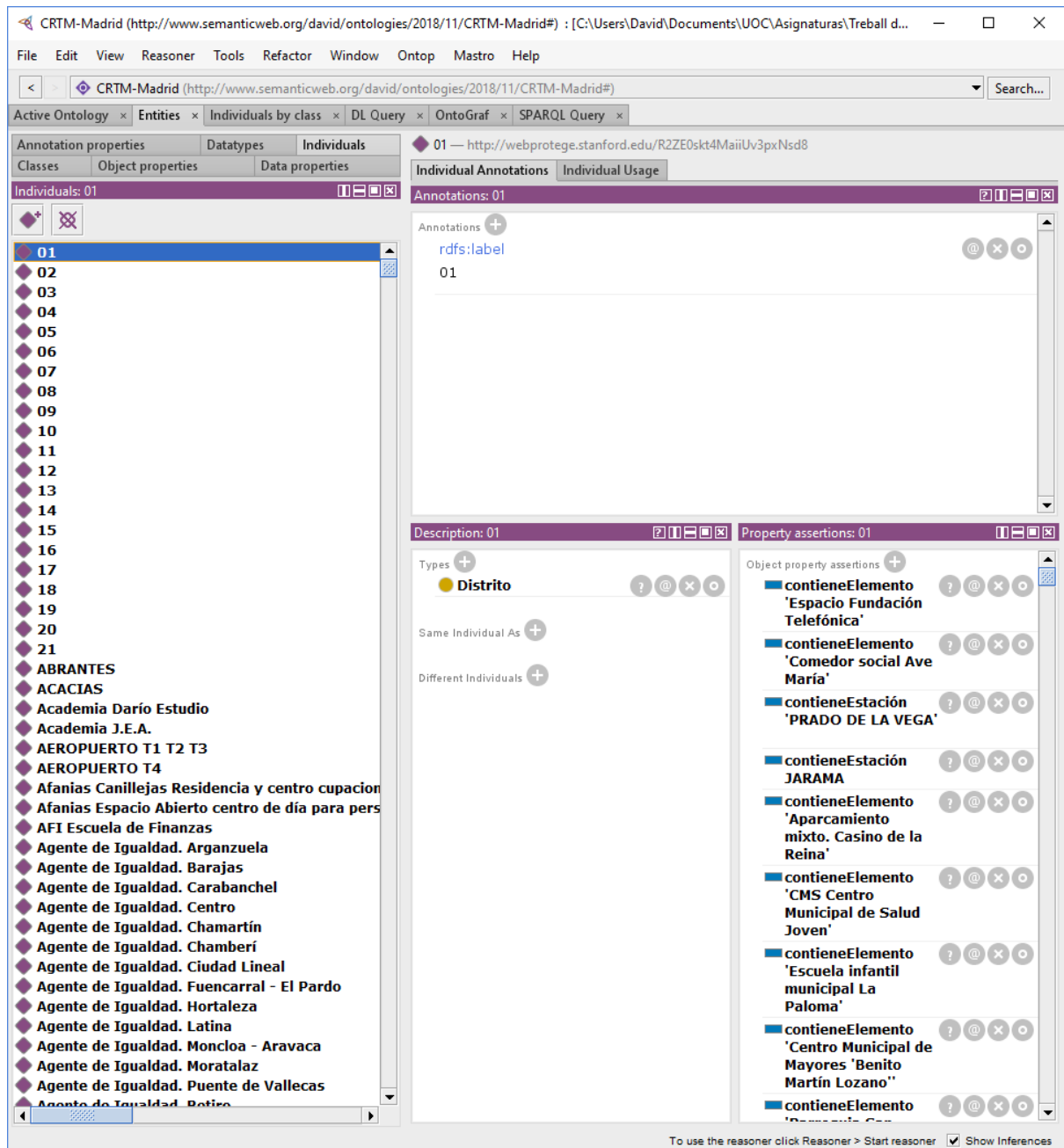


Figura 29: Exemple d'instàncies existents a l'ontologia i les inferències creades per HermiT

Per solucionar aquest problema, cal modificar l'arxiu de configuració inicial de Protégé per tal d'indicar la memòria màxima de la qual pot disposar la màquina virtual Java. Aquest es troba al directori base del programari i, per exemple, el seu contingut hauria de ser el següent per assignar el valor esmentat a 5.000 MB:

```
-Xms5000M
-Xmx5000M
-Xss16M
```

Finalment, obrint el programari, carregant l'ontologia i executant de nou el raonador semàntic, aquest pot realitzar la seva tasca correctament (tot i requerir molt temps: al voltant de 12 hores):

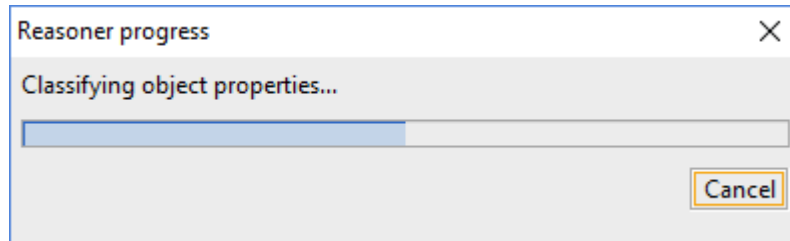


Figura 30: Missatge d'informació amb el progrés de les tasques que realitza el raonador semàntic

Cal tenir en compte que les inferències generades no queden emmagatzemades automàticament a l'ontologia, sinó que han de ser exportades en una de nova. Per aconseguir-ho, cal fer clic al menú "File" i tot seguit a "Export inferred axioms as ontology...", per acabar seleccionant totes les caselles sol·licitades i el format de l'arxiu: "RDF/XML Syntax".

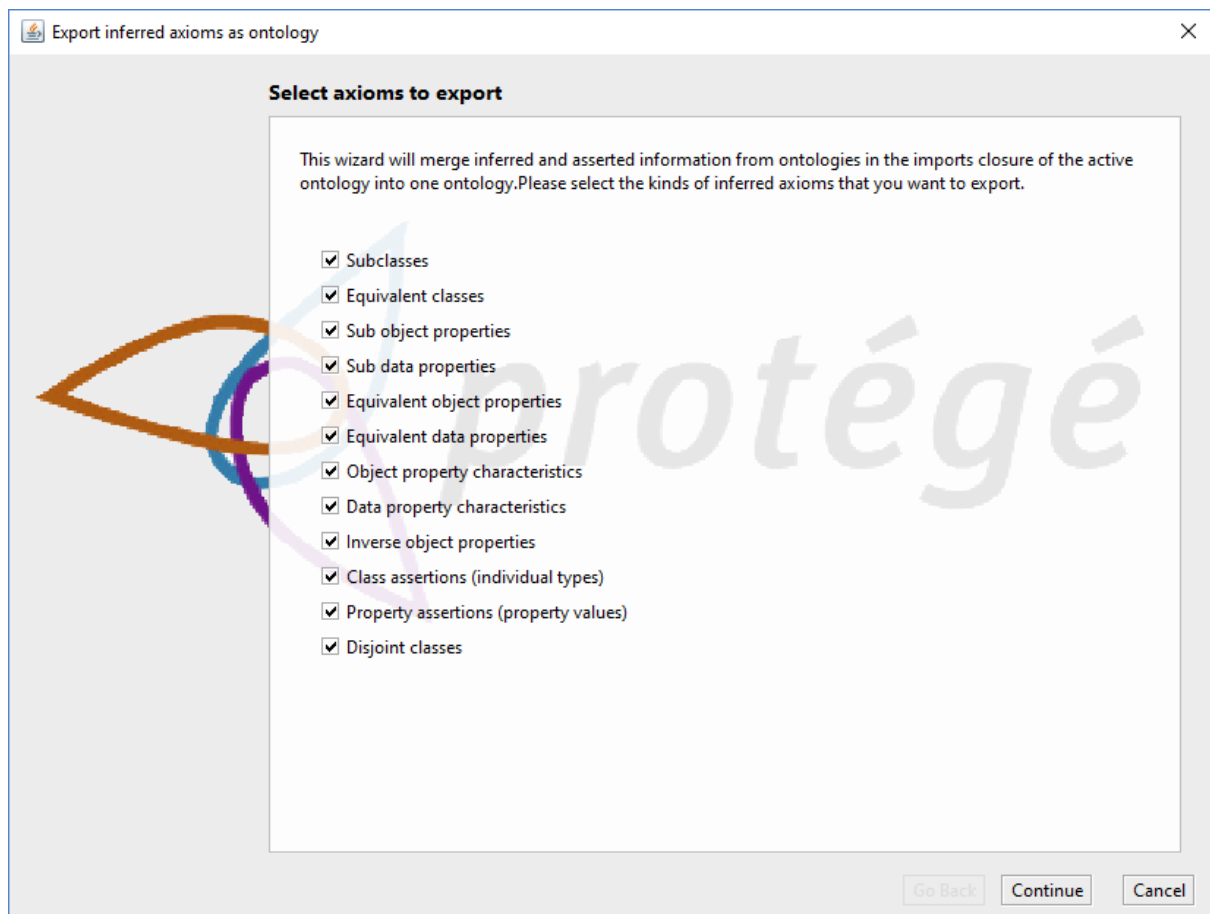


Figura 31: Selecció d'axiomes a exportar a l'ontologia amb les inferències raonades

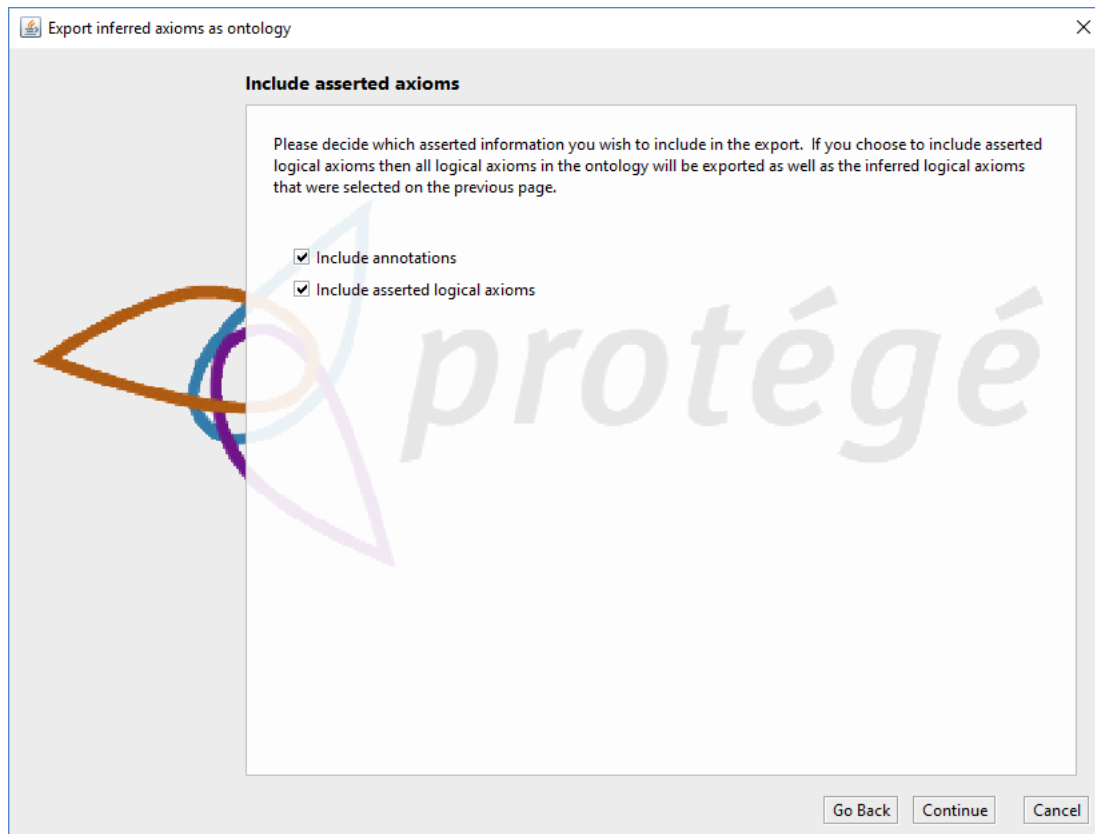


Figura 32: Selecció d'altres axiomes a exportar, incloses les anotacions de les instàncies

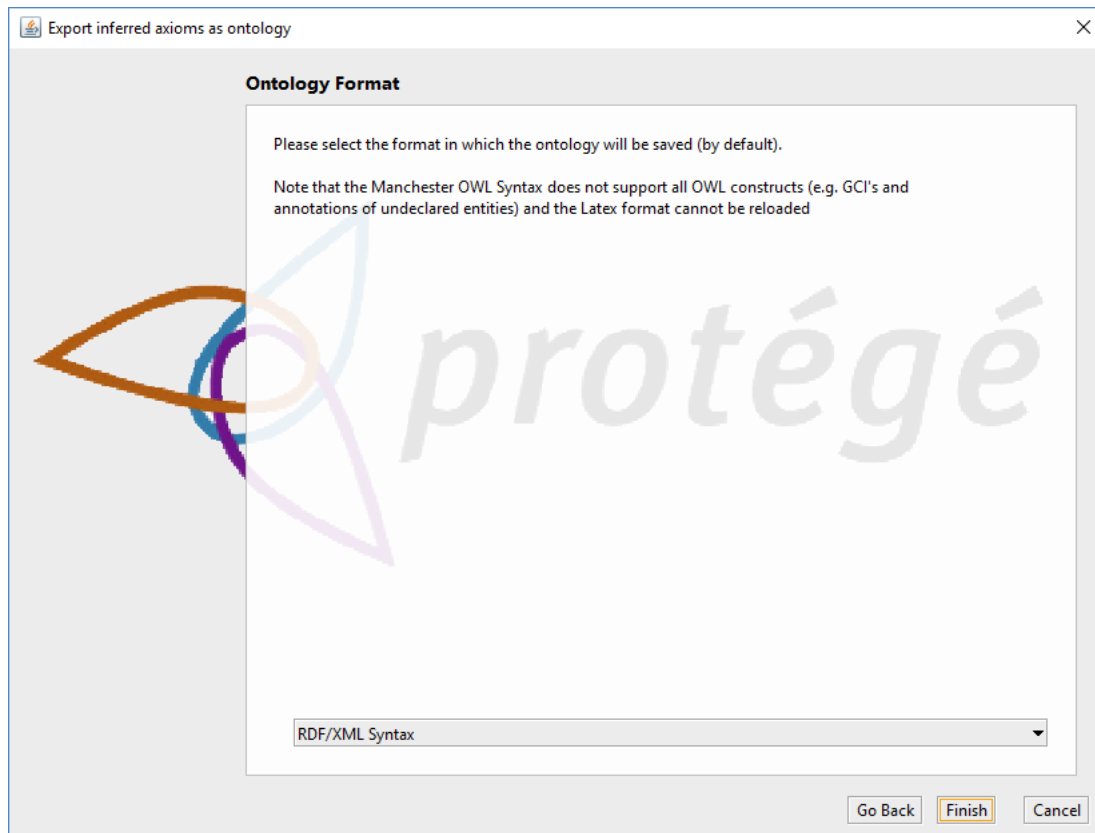


Figura 33: Selecció del format de l'ontologia inferida a exportar

De nou, en el cas de l'ontologia treballada, el procés requereix el voltant d'unes altres 12 hores per obtenir la seva exportació:

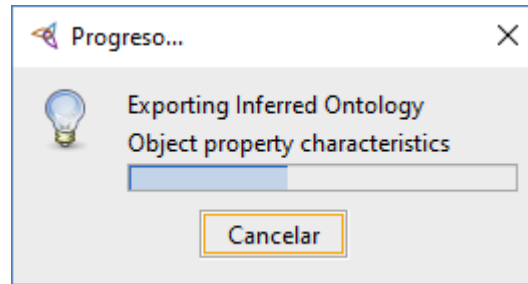


Figura 34: Missatge d'informació amb el progrés de les tasques per exportar l'ontologia

Cal destacar, però, que en aquest cas ha resultat impossible exportar totes les inferències generades: tot i assignar-li 6 GB de RAM -el màxim disponible-, Protégé ha estat incapaç de dur a terme el procés. Així doncs, s'ha decidit eliminar les instàncies d'alguna de les classes i, per tant, l'ontologia inferida no conté la informació de les estacions de la xarxa d'autobús EMT (M6) de la ciutat de Madrid.

4.3. Consultes a l'ontologia

Per tal de realitzar consultes a l'ontologia, ha estat utilitzat el *plugin* SPARQL Query des de la interfície d'usuari del programari Protégé. L'estructura declarativa del llenguatge SPARQL deriva de SQL; per tant, la seva sintaxi utilitza clàusules molt similars com per exemple:

- **PREFIX:** semblant a la clàusula "FROM" de SQL, aquesta s'utilitza per indicar les ontologies o grafs a consultar per la query i així poder conèixer com s'han d'interpretar les dades.
- **SELECT:** es fa servir per indicar els elements a seleccionar (subjecte, predicat o objecte) amb el caràcter "?" al davant d'ells. D'altra banda, es pot indicar "*" per seleccionar tots els elements implicats a la consulta.
- **WHERE:** s'utilitza per filtrar triplets, per exemple, establint una part (subjecte, predicat o objecte) com a variable i assignant-li un nom precedit del símbol "?".
- **GROUP BY:** es fa servir per agrupar els resultats d'una consulta que, per exemple, hagi operat amb "SUM", "COUNT", etc.
- **ORDER BY:** s'utilitza per ordenar els resultats de la consulta (ascendent o descendent).
- **DISTINCT:** es fa servir per filtrar els resultats obtinguts i no mostrar-ne repetits.
- **REDUCED:** s'utilitza per eliminar alguns, tots o cap duplicat.
- **LIMIT:** es fa servir per restringir el nombre de resultats que retorna la consulta.
- **OFFSET:** s'utilitza per establir el punt d'inici de la seqüència global dels resultats obtinguts.

D'altra banda, SPARQL permet l'ús de funcions per convertir i crear tipus de dades (*str*, *lang*, *datatype*, *uri*, *iri*, *bnode*, *strdt*, *strlang*...), per comprovar el seu tipus (*isNumeric*, *isBlank*, *isLiteral*, *isIRI*...) i d'altres condicionals (*bound*, *exists*, *not exists*, *if*, *coalesce*...).

Una primera consulta inicial per entendre el funcionament del llenguatge i els anomenats triplets és la següent, en la que s'obtenen els 100 primers de l'ontologia:

```
SELECT ?subjecte ?predicat ?objecte
WHERE {
    ?subjecte ?predicat ?objecte .
}
LIMIT 100
```

subjecte	predicat	objecte
3253682f_4b69_40f3_bb23_73d9c56e895e	ontologyId	"CRTM-Madrid"
'Residencia para Mayores Reina Victoria'	códigoPostalElemento	"28040"^^<http://www.w3.org/2001/XMLSchema#integer>
'Residencia para Mayores Reina Victoria'	rdf:type	owl:NamedIndividual
códigoPostalElemento	rdf:type	owl:DatatypeProperty
xsd:integer	rdf:type	rdfs:Datatype
57c0a295_1b31_48e0_9fec_d2e6b4e65208	hashCode	"136841256"^^<http://www.w3.org/2001/XMLSchema#integer>
57c0a295_1b31_48e0_9fec_d2e6b4e65208	sourceOntology	3253682f_4b69_40f3_bb23_73d9c56e895e
'Escuela Infantil Guardería Nido la Cigüeña'	esAccesible	"0"^^<http://www.w3.org/2001/XMLSchema#integer>
esAccesible	rdf:type	owl:DatatypeProperty
'Escuela Infantil Guardería Nido la Cigüeña'	rdf:type	owl:NamedIndividual
xsd:integer	rdf:type	rdfs:Datatype
243cb6c9_bf51_45da_89bc_63ef3b7ecd7d	hashCode	"278532251"^^<http://www.w3.org/2001/XMLSchema#integer>
243cb6c9_bf51_45da_89bc_63ef3b7ecd7d	sourceOntology	3253682f_4b69_40f3_bb23_73d9c56e895e
'Escuela Infantil Nuestra Señora de las Victorias'	nombreElemento	"Escuela Infantil Nuestra Señora de las Victorias"
xsd:string	rdf:type	rdfs:Datatype
'Escuela Infantil Nuestra Señora de las Victorias'	rdf:type	owl:NamedIndividual
nombreElemento	rdf:type	owl:DatatypeProperty
a80ee1bf_edbb_4431_87b7_d00395e7e15a	hashCode	"-1589206718"^^<http://www.w3.org/2001/XMLSchema#integer>
a80ee1bf_edbb_4431_87b7_d00395e7e15a	sourceOntology	3253682f_4b69_40f3_bb23_73d9c56e895e
CHAMARTIN	códigoPostalEstación	"28036"^^<http://www.w3.org/2001/XMLSchema#integer>
códigoPostalEstación	rdf:type	owl:DatatypeProperty

Figura 35: Resultat de la consulta que mostra els 100 primers triplets de l'ontologia

Cal destacar que el motor d'SPARQL utilitza per defecte una ontologia sense necessitat d'indicar-la amb la clàusula "PREFIX"; en aquest cas, és la que es troba oberta amb el programari Protégé.

Amb la següent consulta, se cerca obtenir el llistat de les 100 primeres estacions de la ciutat de Madrid que es troben emmagatzemades a l'ontologia i el districte on s'ubiquen:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX tfg: <http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#>
SELECT ?Estació ?Districte
WHERE {
    ?Estació rdf:type tfg:Estación .
    ?Estació tfg:estaciónSeUbicaEn ?Districte .
}
LIMIT 100
```

Estació	Districte
'TORRE ARIAS'	20
'MAR DE CRISTAL'	16
'DOCE DE OCTUBRE'	12
VALDEZARZA	09
O'DONNELL	04
'PUENTE DE VALLECAS'	13
'BANCO DE ESPAÑA'	01
TETUAN	06
'ALONSO CANO'	07
'ESTADIO METROPOLITANO'	20
'PRINCIPE PIO'	09
GOYA	04
'VENTURA RODRIGUEZ'	01
'CASA DE CAMPO'	09
'ALONSO MARTINEZ'	07
ACACIAS	02
URGEL	11
OPANEL	11
'PUERTA DE TOLEDO'	01
SOL	01
BATAN	09

Figura 36: Resultat de la consulta que mostra les 100 primeres estacions de Madrid i el districte on s'ubiquen

Un altre tipus de consulta més específica, podria ser filtrant els resultats; en aquest cas, aquesta és idèntica a l'anterior però només mostrant l'estació anomenada "Vinateros":

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX tfg: <http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#>
SELECT ?Estació ?Districte
WHERE {
    ?Estació rdf:type tfg:Estación .
    ?Estació tfg:estaciónSeUbicaEn ?Districte .
    ?Estació rdfs:label ?nomEstació .
    FILTER(str(?nomEstació)="VINATEROS")
}
```

Estació	Districte
VINATEROS	14

Figura 37: Resultat de la consulta que mostra l'estació anomenada "Vinateros"

I encara més específica podria ser la consulta que retornés totes les estacions de Metro (M4) d'un determinat districte:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX tfg: <http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#>
SELECT ?Estació
WHERE {
    ?Estació rdf:type tfg:M4 .
    ?Estació tfg:estaciónSeUbicaEn ?Districte .
    ?Districte rdfs:label ?codiDistricte .
    FILTER(str(?codiDistricte)="01")
}
```

Estació
'BANCO DE ESPAÑA'
'VENTURA RODRIGUEZ'
'PUERTA DE TOLEDO'
SOL
'ESTACIÓN DEL ARTE'
ARGÜELLES
LAVAPIES
EMBAJADORES
'TIRSO DE MOLINA'
'PLAZA DE ESPAÑA'
'GRAN VIA'
SEVILLA
'SANTO DOMINGO'
'LA LATINA'
TRIBUNAL
OPERA
CHUECA
CALLAO
'ANTON MARTIN'
NOVICIADO

Figura 38: Resultat de la consulta que mostra les estacions de Metro ubicades al districte "01" de Madrid

A la darrera consulta, s'ha convertit el valor de l'etiqueta "codiDistricte" a *string* utilitzant la funció "str", per tal d'aconseguir un correcte filtratge de resultats.

A més, es poden filtrar els resultats a partir de, per exemple, el contingut d'algun camp de les instàncies; en aquest cas, es mostren totes les estacions de Metro (M4) d'un determinat codi postal:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX tfg: <http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#>
SELECT ?Estació
WHERE {
    ?Estació rdf:type tfg:M4 .
    ?Estació tfg:códigoPostalEstación ?codiPostal .
    FILTER(?codiPostal=28001)
}
```

Estació
'PRINCIPE DE VERGARA'
SERRANO
GOYA
VELAZQUEZ

Figura 39: Resultat de la consulta que mostra les estacions de Metro ubicades al codi postal "28001"

Cal destacar que a la darrera consulta no ha estat necessari convertir el valor del camp a filtrar, ja que el camp "codiPostal" està definit a l'ontologia com un *integer*.

No tan sols es poden filtrar a partir d'un determinat contingut literal, sinó també amb una part d'ell. Per exemple, la següent consulta només mostra els centres d'atenció social respectius als "Agente de Igualdad" amb l'ajuda de la funció "regex":

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX tfg: <http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#>
SELECT ?NomCentreAtencióSocial
WHERE {
  ?Centros_Atención_Social rdf:type tfg:Centros_Atención_Social .
  ?Centros_Atención_Social tfg:nombreElemento ?NomCentreAtencióSocial .
  FILTER regex(?NomCentreAtencióSocial, "^Agente de Igualdad", "i")
}
ORDER BY ?NomCentreAtencióSocial
```

NomCentreAtencióSocial
"Agente de Igualdad. Arganzuela"^^<http://www.w3.org/2001/XMLSchema#string>
"Agente de Igualdad. Barajas"^^<http://www.w3.org/2001/XMLSchema#string>
"Agente de Igualdad. Carabanchel"^^<http://www.w3.org/2001/XMLSchema#string>
"Agente de Igualdad. Centro"^^<http://www.w3.org/2001/XMLSchema#string>
"Agente de Igualdad. Chamartín"^^<http://www.w3.org/2001/XMLSchema#string>
"Agente de Igualdad. Chamberí"^^<http://www.w3.org/2001/XMLSchema#string>
"Agente de Igualdad. Ciudad Lineal"^^<http://www.w3.org/2001/XMLSchema#string>
"Agente de Igualdad. Fuencarral - El Pardo"^^<http://www.w3.org/2001/XMLSchema#string>
"Agente de Igualdad. Hortaleza"^^<http://www.w3.org/2001/XMLSchema#string>
"Agente de Igualdad. Latina"^^<http://www.w3.org/2001/XMLSchema#string>
"Agente de Igualdad. Mondoa - Aravaca"^^<http://www.w3.org/2001/XMLSchema#string>
"Agente de Igualdad. Moratalaz"^^<http://www.w3.org/2001/XMLSchema#string>
"Agente de Igualdad. Puente de Vallecas"^^<http://www.w3.org/2001/XMLSchema#string>
"Agente de Igualdad. Retiro"^^<http://www.w3.org/2001/XMLSchema#string>
"Agente de Igualdad. Salamanca"^^<http://www.w3.org/2001/XMLSchema#string>
"Agente de Igualdad. San Blas-Canillejas"^^<http://www.w3.org/2001/XMLSchema#string>
"Agente de Igualdad. Tetuán"^^<http://www.w3.org/2001/XMLSchema#string>
"Agente de Igualdad. Usera"^^<http://www.w3.org/2001/XMLSchema#string>
"Agente de Igualdad. Vicálvaro"^^<http://www.w3.org/2001/XMLSchema#string>
"Agente de Igualdad. Villa de Vallecas"^^<http://www.w3.org/2001/XMLSchema#string>
"Agente de Igualdad. Villaverde"^^<http://www.w3.org/2001/XMLSchema#string>

Figura 40: Resultat de la consulta que mostra els "Agente de Igualdad" ubicats a Madrid

Analitza els resultats del *Observatorio de la Ciudad*, al següent diagrama de Pareto³⁵ destaquen els índexs de satisfacció pel que fa a l'oferta cultural en alguns districtes de Madrid:

³⁵ El diagrama de Pareto -també anomenat corba tancada o Distribució ABC- és una gràfica que mostra una sèrie de dades separades per barres i en ordre descendent (d'esquerra a dreta), i una línia -en un eix secundari- que representa les freqüències acumulades dels valors en qüestió.

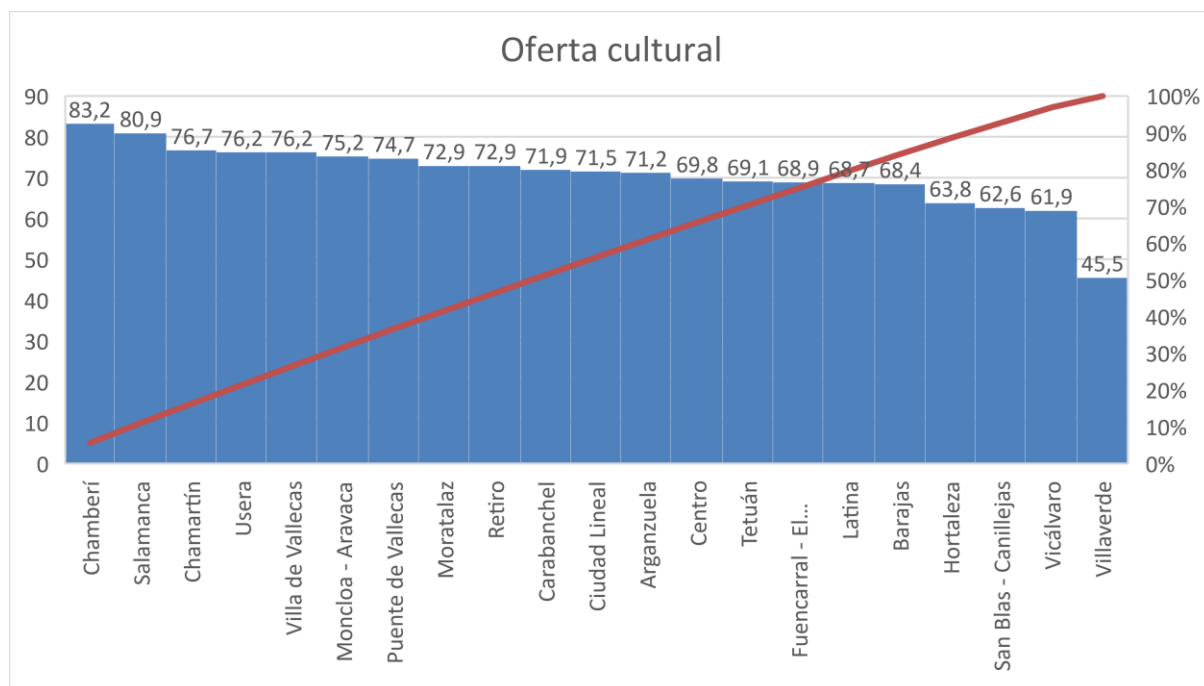


Figura 41: Índex de satisfacció de la població de Madrid l'any 2017 respecte a l'oferta cultural per districte

Una consulta interessant per contrastar l'índex de satisfacció de la figura 41 i que pot respondre l'ontologia, és el nombre de cinemes i filmoteques de cada districte de la ciutat de Madrid:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX tfg: <http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#>
SELECT ?NomDistricte (count(?Cines_y_filmotecas) as ?NombreCinemesIFilmotecas)
WHERE {
    ?Cines_y_filmotecas rdf:type tfg:Cines_y_filmotecas .
    ?Cines_y_filmotecas tfg:elementoSeUbicaEn ?Districte .
    ?Districte tfg:nombreDistrito ?NomDistricte .
}
GROUP BY ?NomDistricte
ORDER BY DESC (?NombreCinemesIFilmotecas)
```

NomDistricte	NombreCinemesIFilmotecas
"CHAMBERI"^^<http://www.w3.org/2001/XMLSchema#string>	"7"^^<http://www.w3.org/2001/XMLSchema#integer>
"CENTRO"^^<http://www.w3.org/2001/XMLSchema#string>	"6"^^<http://www.w3.org/2001/XMLSchema#integer>
"ARGANZUELA"^^<http://www.w3.org/2001/XMLSchema#string>	"3"^^<http://www.w3.org/2001/XMLSchema#integer>
"MONCLOA-ARAVACA"^^<http://www.w3.org/2001/XMLSchema#string>	"3"^^<http://www.w3.org/2001/XMLSchema#integer>
"HORTALEZA"^^<http://www.w3.org/2001/XMLSchema#string>	"3"^^<http://www.w3.org/2001/XMLSchema#integer>
"FUENCARRAL-EL PARDO"^^<http://www.w3.org/2001/XMLSchema#string>	"2"^^<http://www.w3.org/2001/XMLSchema#integer>
"SALAMANCA"^^<http://www.w3.org/2001/XMLSchema#string>	"2"^^<http://www.w3.org/2001/XMLSchema#integer>
"SAN BLAS-CANILLEJAS"^^<http://www.w3.org/2001/XMLSchema#string>	"1"^^<http://www.w3.org/2001/XMLSchema#integer>
"CIUDAD LINEAL"^^<http://www.w3.org/2001/XMLSchema#string>	"1"^^<http://www.w3.org/2001/XMLSchema#integer>
"VILLA DE VALLECAS"^^<http://www.w3.org/2001/XMLSchema#string>	"1"^^<http://www.w3.org/2001/XMLSchema#integer>
"RETIRO"^^<http://www.w3.org/2001/XMLSchema#string>	"1"^^<http://www.w3.org/2001/XMLSchema#integer>
"CARABANCHEL"^^<http://www.w3.org/2001/XMLSchema#string>	"1"^^<http://www.w3.org/2001/XMLSchema#integer>
"CHAMARTIN"^^<http://www.w3.org/2001/XMLSchema#string>	"1"^^<http://www.w3.org/2001/XMLSchema#integer>

Figura 42: Resultat de la consulta que mostra el nombre de cinemes i filmoteques de cada districte de Madrid

Cal destacar que a la consulta anterior s'ha preferit mostrar el nom del districte en comptes de l'anotació de les seves respectives instàncies, ja que en el cas de l'ontologia treballada es tracta del seu identificador (un codi numèric de l'1 al 21).

Adicionalment, una consulta similar però realitzada per conèixer el nombre de museus ubicats a cada districte de la ciutat de Madrid, ajuda a contrastar l'índex de satisfacció de la figura 41:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX tfg: <http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#>
SELECT ?NomDistricte (count(?Museos) as ?NombreMuseos)
WHERE {
    ?Museos rdf:type tfg:Museos .
    ?Museos tfg:elementoSeUbicaEn ?Districte .
    ?Districte tfg:nombreDistrito ?NomDistricte .
}
GROUP BY ?NomDistricte
ORDER BY DESC (?NombreMuseos)
```

NomDistricte	NombreMuseos
"CENTRO"^^<http://www.w3.org/2001/XMLSchema#string>	"28"^^<http://www.w3.org/2001/XMLSchema#integer>
"MONCLOA-ARAVACA"^^<http://www.w3.org/2001/XMLSchema#string>	"9"^^<http://www.w3.org/2001/XMLSchema#integer>
"RETIRO"^^<http://www.w3.org/2001/XMLSchema#string>	"7"^^<http://www.w3.org/2001/XMLSchema#integer>
"SALAMANCA"^^<http://www.w3.org/2001/XMLSchema#string>	"6"^^<http://www.w3.org/2001/XMLSchema#integer>
"CHAMBERI"^^<http://www.w3.org/2001/XMLSchema#string>	"6"^^<http://www.w3.org/2001/XMLSchema#integer>
"FUENCARRAL-EL PARDO"^^<http://www.w3.org/2001/XMLSchema#string>	"3"^^<http://www.w3.org/2001/XMLSchema#integer>
"ARGANZUELA"^^<http://www.w3.org/2001/XMLSchema#string>	"2"^^<http://www.w3.org/2001/XMLSchema#integer>
"LATINA"^^<http://www.w3.org/2001/XMLSchema#string>	"2"^^<http://www.w3.org/2001/XMLSchema#integer>
"CHAMARTIN"^^<http://www.w3.org/2001/XMLSchema#string>	"2"^^<http://www.w3.org/2001/XMLSchema#integer>
"TETUAN"^^<http://www.w3.org/2001/XMLSchema#string>	"1"^^<http://www.w3.org/2001/XMLSchema#integer>
"CIUDAD LINEAL"^^<http://www.w3.org/2001/XMLSchema#string>	"1"^^<http://www.w3.org/2001/XMLSchema#integer>
"BARAJAS"^^<http://www.w3.org/2001/XMLSchema#string>	"1"^^<http://www.w3.org/2001/XMLSchema#integer>
"PUENTE DE VALLECAS"^^<http://www.w3.org/2001/XMLSchema#string>	"1"^^<http://www.w3.org/2001/XMLSchema#integer>

Figura 43: Resultat de la consulta que mostra el nombre de museus de cada districte de Madrid

La següent consulta quelcom més complexa, mostra la ràtio entre el nombre de zones d'aparcament per residents i el nombre d'estacions de transport públic per cada districte:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX tfg: <http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#>
SELECT ?NomDistricte (count(distinct ?Aparcamientos_residentes) as
?NombreZonesEstacionamentResidents) (count(distinct ?Estación) as ?NombreEstacions)
(?NombreZonesEstacionamentResidents/?NombreEstacions as ?Ratio)
WHERE {
    ?Aparcamientos_residentes rdf:type tfg:Aparcamientos_residentes .
    ?Aparcamientos_residentes tfg:elementoSeUbicaEn ?Districte .
    ?Estación rdf:type tfg:Estación .
    ?Estación tfg:estaciónSeUbicaEn ?Districte .
    ?Districte tfg:nombreDistrito ?NomDistricte .
}
GROUP BY ?NomDistricte
ORDER BY DESC (?Ratio)
```


NomDistricte	NombreZonesEstacionamentResidents	NombreEstacions	Ratio
"MORATALAZ"	19	4	4.75
"RETIRO"	16	6	2.6666666666666666
"SALAMANCA"	29	12	2.4166666666666666
"TETUAN"	17	8	2.125
"CIUDAD LINEAL"	22	12	1.8333333333333333
"CHAMBERI"	20	14	1.4285714285714285
"LATINA"	16	12	1.3333333333333333
"CHAMARTIN"	18	15	1.2
"ARGANZUELA"	10	9	1.1111111111111111
"FUENCARRAL-EL PARDO"	20	19	1.0526315789473684
"MONCLOA-ARAVACA"	14	14	1
"HORTALEZA"	14	14	1
"CENTRO"	15	20	0.75
"VICALVARO"	10	14	0.7142857142857142
"CARABANCHEL"	10	14	0.7142857142857142
"USERA"	4	6	0.6666666666666666
"VILLAVARDE"	3	8	0.625
"VILLA DE VALLECAS"	3	6	0.5
"SAN BLAS-CANILLEJAS"	4	10	0.4
"PUENTE DE VALLECAS"	3	10	0.3
"BARAJAS"	1	7	0.14285714285714285

Figura 44: Resultat de la consulta que mostra la ràtio entre zones d'aparcament i d'estacions de transport

Cal destacar l'ús de la clàusula "DISTINCT" per separar els resultats a l'hora d'enllaçar-los amb una classe en comú (districte) i l'operació aritmètica bàsica realitzada entre els valors retornats per ambdues funcions "COUNT".

El següent diagrama mostra l'índex de satisfacció dels habitants de la ciutat de Madrid pel que fa a la mobilitat i el transport públic, i denoten certa relació amb la ràtio resultant de l'anterior consulta a l'ontologia.

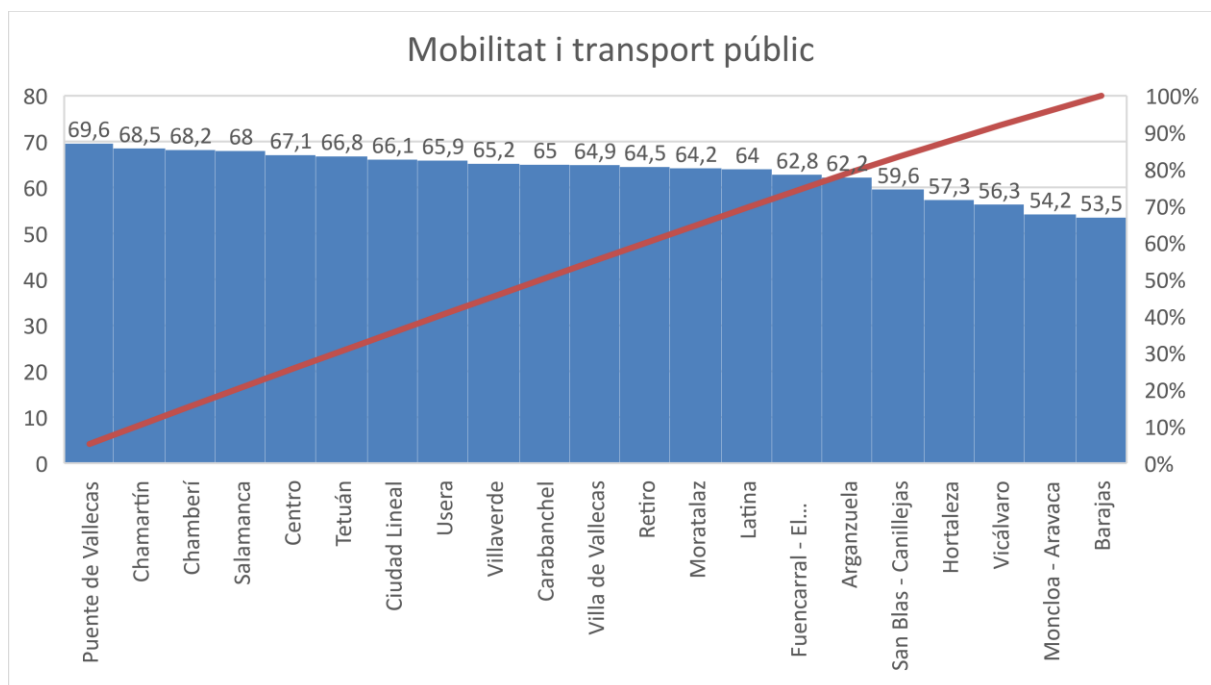


Figura 45: Índex de satisfacció de la població de Madrid l'any 2017 respecte a la mobilitat i el transport

Primerament, cal considerar que els repositoris de dades han permès que les instàncies de la classe “Estació” continguin els camps “lat” i “long” basats en el sistema de referència WGS84, gràcies al fet que prèviament s’ha definit a l’ontologia el prefix necessari per treballar amb estàndards de representació de posicionament. Per exemple, la següent consulta mostra totes les estacions de Metro de la ciutat de Madrid amb les seves respectives coordenades de latitud i longitud:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX tfg: <http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#>
PREFIX geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>
SELECT ?nomEstació ?Latitud ?Longitud
WHERE {
    ?M4 rdf:type tfg:M4 .
    ?M4 geo:lat ?Latitud .
    ?M4 geo:long ?Longitud .
    ?M4 tfg:nombreEstación ?nomEstació .
}
ORDER BY ?nomEstació
```

nomEstació	Latitud	Longitud
"ABRANTES"^^<http://www.w3.org/2001/XMLSchema#>"40.380833614731245"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>"-3.7279002491217077"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>		
"ACACIAS"^^<http://www.w3.org/2001/XMLSchema#>"40.40387164372991"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>"-3.7066410860751864"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>		
"AEROPUERTO T1 T2 T3"^^<http://www.w3.org/2001/XMLSchema#>"40.4686442368363"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>"-3.569538255943205"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>		
"AEROPUERTO T4"^^<http://www.w3.org/2001/XMLSchema#>"40.491767525223615"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>"-3.593245597213427"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>		
"ALAMEDA DE OSUNA"^^<http://www.w3.org/2001/XMLSchema#>"40.45778862852046"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>"-3.5875209290434937"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>		
"ALFONSO XIII"^^<http://www.w3.org/2001/XMLSchema#>"40.44824838816102"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>"-3.6679061529956103"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>		
"ALMENDRALES"^^<http://www.w3.org/2001/XMLSchema#>"40.38409352717396"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>"-3.697881397655217"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>		
"ALONSO CANO"^^<http://www.w3.org/2001/XMLSchema#>"40.43837537611044"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>"-3.6992456432372576"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>		
"ALONSO MARTINEZ"^^<http://www.w3.org/2001/XMLSchema#>"40.42828258137045"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>"-3.6950229019466976"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>		
"ALSACIA"^^<http://www.w3.org/2001/XMLSchema#>"40.41829269464442"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>"-3.6235060015486664"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>		
"ALTO DE EXTREMADURA"^^<http://www.w3.org/2001/XMLSchema#>"40.40992601698508"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>"-3.7389361815722615"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>		
"ALTO DEL ARENAL"^^<http://www.w3.org/2001/XMLSchema#>"40.38977101214874"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>"-3.6452165490428263"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>		
"ALUCHE"^^<http://www.w3.org/2001/XMLSchema#>"40.38562861956385"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>"-3.76080381070444"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>		
"ALVARADO"^^<http://www.w3.org/2001/XMLSchema#>"40.450331975994324"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>"-3.7033142092757845"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>		
"ANTON MARTIN"^^<http://www.w3.org/2001/XMLSchema#>"40.41246356569"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>"-3.6993729633448895"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>		
"ANTONIO MACHADO"^^<http://www.w3.org/2001/XMLSchema#>"40.47021772039073"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>"-3.7176948552412226"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>		
"ARGANZUELA-PLANETARIO"^^<http://www.w3.org/2001/XMLSchema#>"40.39310270492071"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>"-3.688640347447094"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>		
"ARGÜELLES"^^<http://www.w3.org/2001/XMLSchema#>"40.430531839640196"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>"-3.715962651094834"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>		
"ARTILLEROS"^^<http://www.w3.org/2001/XMLSchema#>"40.4052238336675"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>"-3.641812337774728"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>		
"ARTURO SORIA"^^<http://www.w3.org/2001/XMLSchema#>"40.4558324918715"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>"-3.656181218097823"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>		
"ASCAO"^^<http://www.w3.org/2001/XMLSchema#>"40.4302055301581"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>"-3.64105571498423"^^<http://www.w3.org/2003/01/geo/wgs84_pos#>		

Figura 46: Resultat de la consulta que mostra les coordenades de latitud i longitud de les estacions de Metro

Però per realitzar-hi consultes de geoposicionament, és necessari que l'ontologia inclogui la següent classe i les seves respectives subclasses:

- **Geometry:**
 - **Point**³⁶: és un punt format per les coordenades de latitud i longitud.
 - **Polygon**³⁷: és una superfície definida per un límit exterior i cap o més límits interiors. Cada límit es representa mitjançant punts.
 - **MultiPolygon**³⁸: és una superfície múltiple formada polígons.

³⁶ Definició de Point: <http://www.opengis.net/ont/sf#Point>, consultat el gener del 2019.

³⁷ Definició de Polygon: <http://www.opengis.net/ont/sf#Polygon>, consultat el gener del 2019.

³⁸ Definició de MultiPolygon: <http://www.opengis.net/ont/sf#MultiPolygon>, consultat el gener del 2019.

Tractant la càrrega de dades a l'ontologia amb un altre sistema (Apache Jena³⁹, per exemple), hauria permès nodrir les classes descrites prèviament i poder realitzar consultes com quins elements hi ha més a prop d'una determinada estació de transport públic, quins hi ha dins d'un polígon definit, etc.

Utilitzant un servei web de tercers per demostrar el seu funcionament, es pot executar la següent consulta que utilitza la funció “nearby” per retornar els objectes existents a Lotico⁴⁰ que es trobin a un radi de 20 km. (àrea urbana) de les coordenades geogràfiques base de la ciutat de Madrid:

```
PREFIX spatial: <http://jena.apache.org/spatial#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>
PREFIX lotico: <http://www.lotico.com/ontology/>

SELECT *
WHERE {
    ?object spatial:nearby(40.4167 -3.70325 20 'km') .
    ?object rdfs:label ?label .
    ?object geo:lat ?latitud .
    ?object geo:long ?longitud .
}
```

object	label	latitud	longitud
<http://www.lotico.com/resource/member_51658852>	"Big Data Hispano"	"40.42"	"-3.71"
<http://www.lotico.com/resource/member_207547644>	"m"	"40.42"	"-3.71"
<http://www.lotico.com/resource/member_43534792>	"Mar?a Poveda"	"40.42"	"-3.71"
<http://www.lotico.com/resource/member_157170012>	"Santos"	"40.42"	"-3.71"
<http://www.lotico.com/resource/member_117618842>	"Pedro GL"	"40.42"	"-3.71"
<http://www.lotico.com/resource/member_185345929>	"Arnaud2500"	"40.42"	"-3.71"
<http://www.lotico.com/resource/member_179534292>	"albmartin"	"40.42"	"-3.71"
<http://www.lotico.com/resource/member_185009426>	"Erik Escoffier"	"40.42"	"-3.71"
<http://www.lotico.com/resource/member_183534425>	"Maria Jose Peco Ruiz"	"40.42"	"-3.71"
<http://www.lotico.com/resource/member_72422642>	"Eduardo Basterrechea"	"40.42"	"-3.71"
<http://www.lotico.com/resource/member_77169772>	"Elena ?lvarez Mellado"	"40.42"	"-3.71"
<http://www.lotico.com/resource/member_200111042>	"Ignacio Redondo"	"40.42"	"-3.71"
<http://www.lotico.com/resource/member_251092718>	"Louis Metz"	"40.42"	"-3.71"
<http://www.lotico.com/resource/member_182928083>	"Guillermo"	"40.42"	"-3.71"
<http://www.lotico.com/resource/member_195035574>	"Pablo Lasso"	"40.42"	"-3.71"
<http://www.lotico.com/resource/member_39180352>	"Manuel Cornejo"	"40.42"	"-3.71"
<http://www.lotico.com/resource/member_211782705>	"Patricia"	"40.42"	"-3.71"
<http://www.lotico.com/resource/member_8663957>	"elkosmonauta"	"40.42"	"-3.71"
<http://www.lotico.com/resource/member_88254972>	"Luc?a Batista Flores"	"40.42"	"-3.71"
<http://www.lotico.com/resource/member_192003052>	"Andres D Rengifo"	"40.42"	"-3.71"
<http://www.lotico.com/resource/member_115729422>	"Javier"	"40.42"	"-3.71"
<http://www.lotico.com/resource/member_193175401>	"pepcarls"	"40.42"	"-3.71"
<http://www.lotico.com/resource/member_30044212>	"Alison B."	"40.42"	"-3.71"
<http://www.lotico.com/resource/member_52538362>	"Victor AM"	"40.42"	"-3.71"
<http://www.lotico.com/resource/member_68660532>	"Jose"	"40.42"	"-3.71"
<http://www.lotico.com/resource/member_53090592>	"Carlos kuchkovsky"	"40.42"	"-3.71"
<http://www.lotico.com/resource/member_70270482>	"Javier Julian"	"40.42"	"-3.71"

Figura 47: Resultat de la consulta que mostra les instàncies de Lotico que es troben a Madrid

³⁹ Apache Jena és un *framework* de codi obert per a Java, que proporciona una *API* per llegir i escriure dades en arxius RDF.

⁴⁰ Lotico és una xarxa social semàntica i les seves sigles estan formades pels caràcters inicials dels termes que hi aplega: LLocation, Tlme, CCommunity i CConcepts.

4.4. Creació d'una aplicació a mida

Un dels objectius del present treball ha estat la creació d'un programari que fos capaç d'interactuar amb l'ontologia creada. Aquest s'ha desenvolupat utilitzant el llenguatge PHP⁴¹ i, gràcies a la integració de la llibreria EasyRdf, permet les funcionalitats següents:

- Llistar les instàncies de les classes de l'ontologia.
- Cercar instàncies concretes.
- Realitzar-hi consultes personalitzades.
- Exportar el contingut de l'ontologia en format HTML o en text pla.

Cal tenir en compte que l'aplicació és accessible mitjançant un servidor web; per tant, ha estat necessari allotjar-la al directori d'accés públic del programari XAMPP (més informació sobre la seva instal·lació a l'annex C).

Primerament, és necessari importar l'ontologia creada al programari Virtuoso amb la finalitat d'habilitar el servei web (anomenat "SPARQL endpoint"⁴²) que permet realitzar-hi consultes i obtenir els resultats. Per aconseguir-ho, cal accedir a <http://localhost:8890/conductor> mitjançant un navegador web, iniciar sessió amb l'usuari "dba" i fer clic als menús "Linked Data" i "Quad Store Upload":

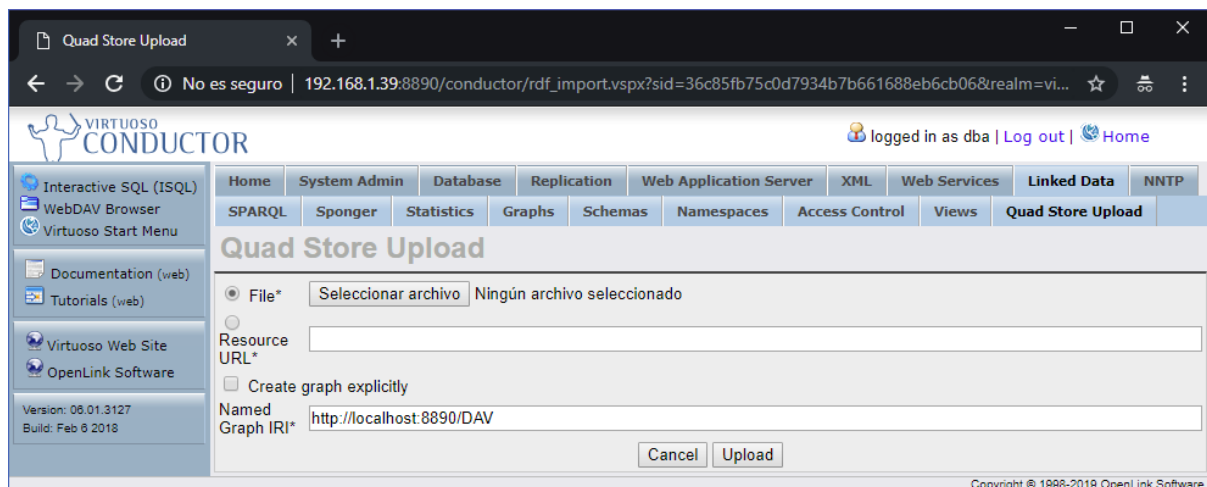


Figura 48: Menú *Quad Store Upload* del programari Virtuoso

És necessari seleccionar l'ontologia creada fent clic a "Seleccionar archivo" de l'opció "File" i finalitzar el procés fent clic al botó "Upload".

Finalment, accedint a <http://localhost:8890/sparql> i executant la següent consulta, es pot comprovar la correcta importació de l'ontologia:

⁴¹ Correspon a les sigles d'*Hypertext Preprocessor* i és un llenguatge de programació de codi obert dissenyat pel desenvolupament web. S'incrusta en el codi HTML, treballa al costat del servidor i permet contingut web dinàmic.

⁴² Llistat de *SPARQL endpoints* disponibles segons W3C: <https://www.w3.org/wiki/SparqlEndpoints>, consultat el gener del 2019.

```
select distinct ?Concept where {[ ] a ?Concept} LIMIT 100
```

Virtuoso SPARQL Query Editor

Default Data Set Name (Graph IRI)

Query Text

```
select distinct ?Concept where {[ ] a ?Concept} LIMIT 100
```

(Security restrictions of this server do not allow you to retrieve remote RDF data, see [details](#).)

Results Format:

Execution timeout: milliseconds (values less than 1000 are ignored)

Options: ☒ Strict checking of void variables

(The result can only be sent back to browser, not saved on the server, see [details](#))

Copyright © 2019 [OpenLink Software](#)
Virtuoso version 06.01.3127 on Linux (x86_64-pc-linux-gnu), Single Server Edition

Figura 49: Editor de consultes SPARQL del programari Virtuoso

http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#Aparcamientos_públicos_municipales
http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#Elemento
http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#Aparcamientos_residentes
http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#Auditorios_y_salas_conciertos
http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#Bibliotecas_Especializadas
http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#Bibliotecas_y_bibliobuses
http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#Cementerios
http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#Centros_Atención_Menores_Familia
http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#Centros_Atención_Médica
http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#Centros_Atención_Social
http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#Centros_Culturales_Municipales
http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#Centros_Deportivos_Municipales
http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#Centros_Educación
http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#Centros_Enseñanza
http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#Cines_y_filmotecas
http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#Distrito
http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#Embajadas_y_Consulados
http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#Estación
http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#M10
http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#M4
http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#M5
http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#Mercados_Municipales
http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#Monumentos
http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#Museos
http://www.semanticweb.org/david/ontologies/2018/11/CRTM-Madrid#Teatros

Figura 50: Resultat de la consulta SPARQL a l'ontologia importada al programari Virtuoso

Per aconseguir les funcionalitats descrites anteriorment, ha estat necessari estudiar l'API⁴³ i conèixer totes les possibilitats que ofereix la llibreria EasyRdf. A grans trets, en aquest cas s'han utilitzat les funcions referents a interactuar amb SPARQL *endpoints* i llegir el contingut d'una determinada ontologia, encara que n'hi ha que també permeten la seva modificació.

Si bé és cert que la documentació del codi PHP del programari desenvolupat a mida no forma part del present treball, cal destacar que l'ús d'EasyRdf es basa a cridar la llibreria, definir la ubicació de l'ontologia i l'URL del SPARQL *endpoint*, i integrar les seves funcions per obtenir els resultats desitjats.

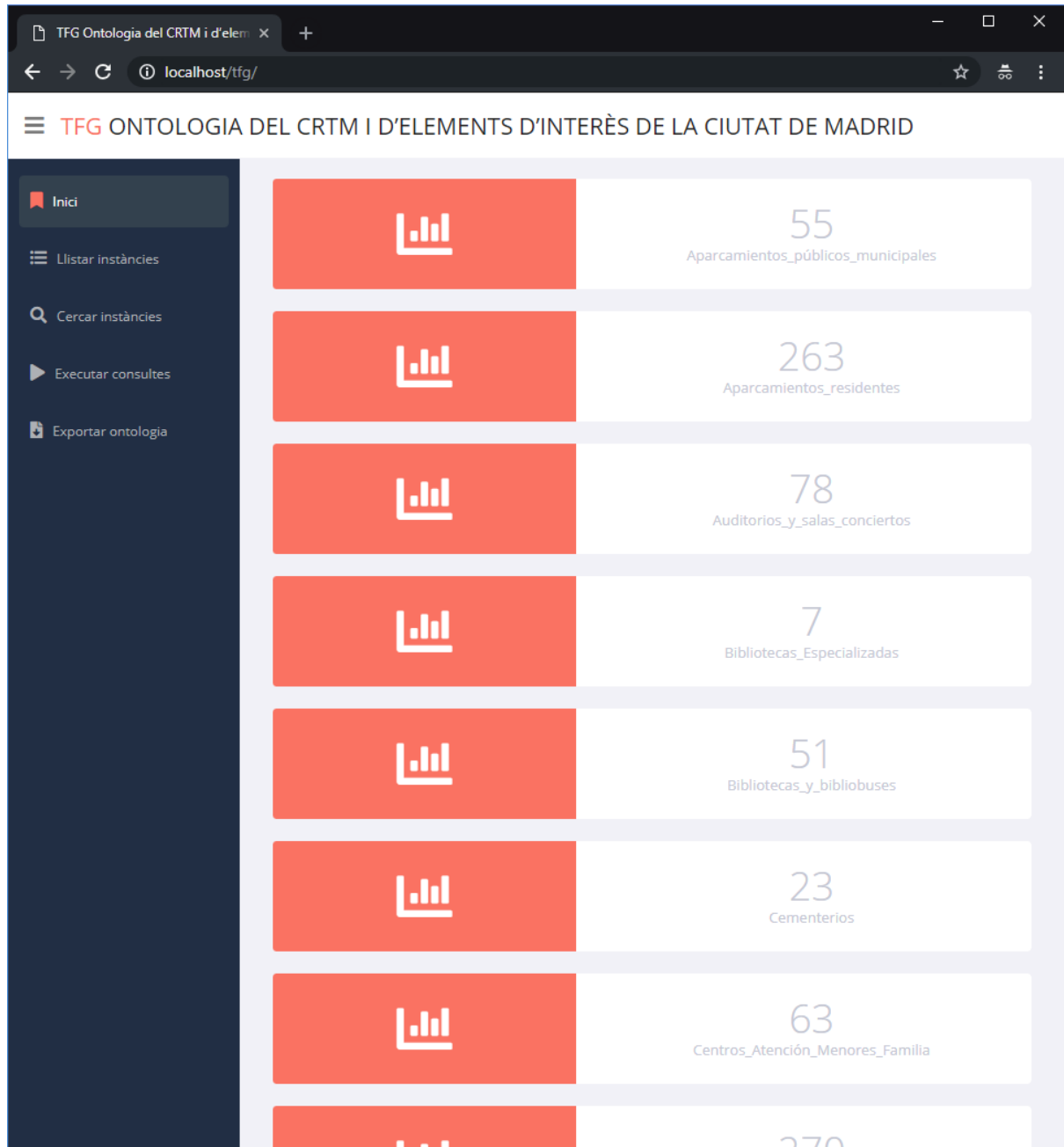


Figura 51: Pantalla inicial del programari desenvolupat a mida

⁴³ Les seves sigles fan referència a *Application Programming Interfaces*, i es tracta de la definició del conjunt de funcions i mètodes pertanyents a una determinada llibreria.

Un cop comprovat l'estat del programari Virtuoso després d'importar-hi l'ontologia, l'aplicació desenvolupada a mida ja hi pot interactuar; per exemple, una primera funcionalitat és obtenir el llistat de les instàncies de la classe "Universidades":

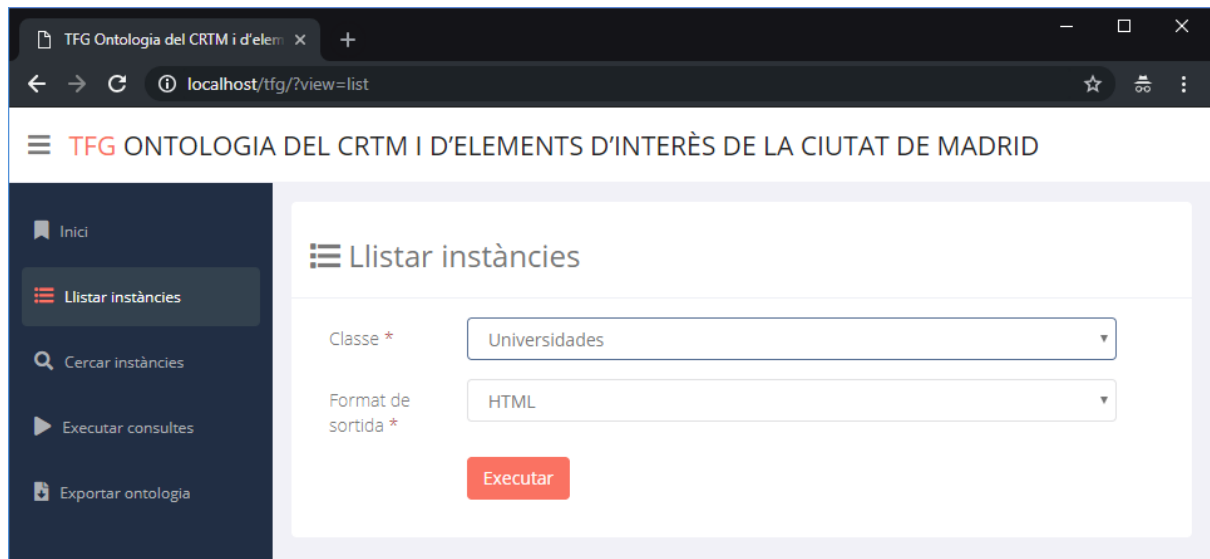


Figura 52: Selecció de la classe "Universidades" i el format "HTML" per obtenir les seves instàncies

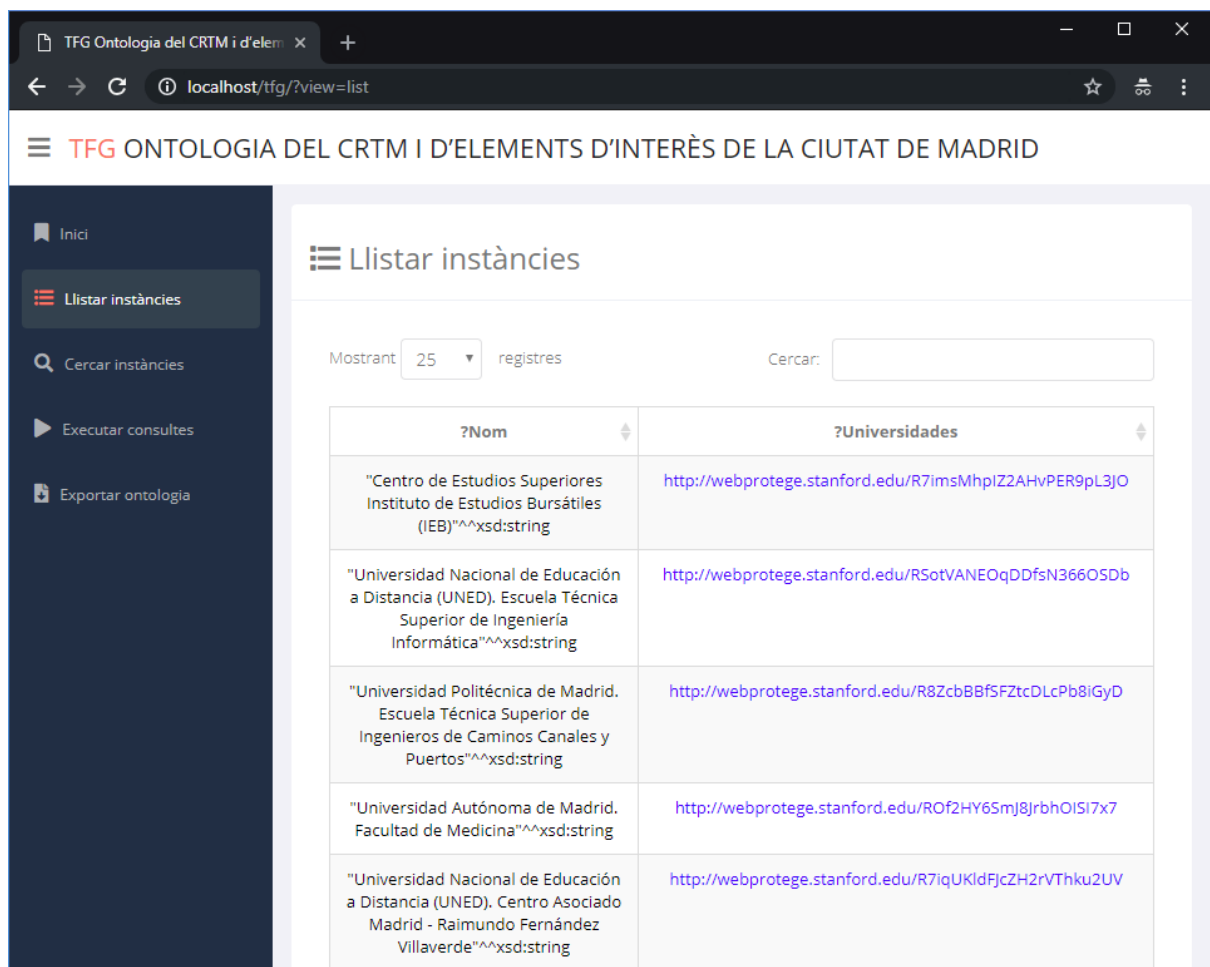


Figura 53: Instàncies de la classe "Universidades" existents a l'ontologia

Amb la finalitat de llistar instàncies de forma personalitzada, el menú “Cercar instàncies” permet filtrar els resultats per algun dels camps de la classe escollida. En aquest cas, es realitza la cerca de totes les instàncies de la classe “Embajadas_y_Consulados” amb el text “Perú” al seu nom:

Figura 54: Filtre del text “Perú” a la classe “Embajadas_y_Consulados” per obtenir les seves instàncies

?Nom	?Embajadas_y_Consulados
"Consulado de Perú"^^xsd:string	http://webprotege.stanford.edu/R8G4ePPbDm78b6ewfBH63IM
"Embajada de Perú"^^xsd:string	http://webprotege.stanford.edu/R8cnWREoZEemDIOpDMIW8hV

Figura 55: Instàncies amb el text “Perú” com a nom de la classe “Embajadas_y_Consulados”

Una altra de les funcionalitats és la d'executar consultes lliures a l'ontologia. Per defecte, el programari ja inclou una sèrie de prefixos que poden ser cridats per la *query*. En aquest cas, s'executa una de les consultes realitzades anteriorment al *plugin* SPARQL Query del programari Protégé, obtenint els mateixos resultats que a la figura 42.

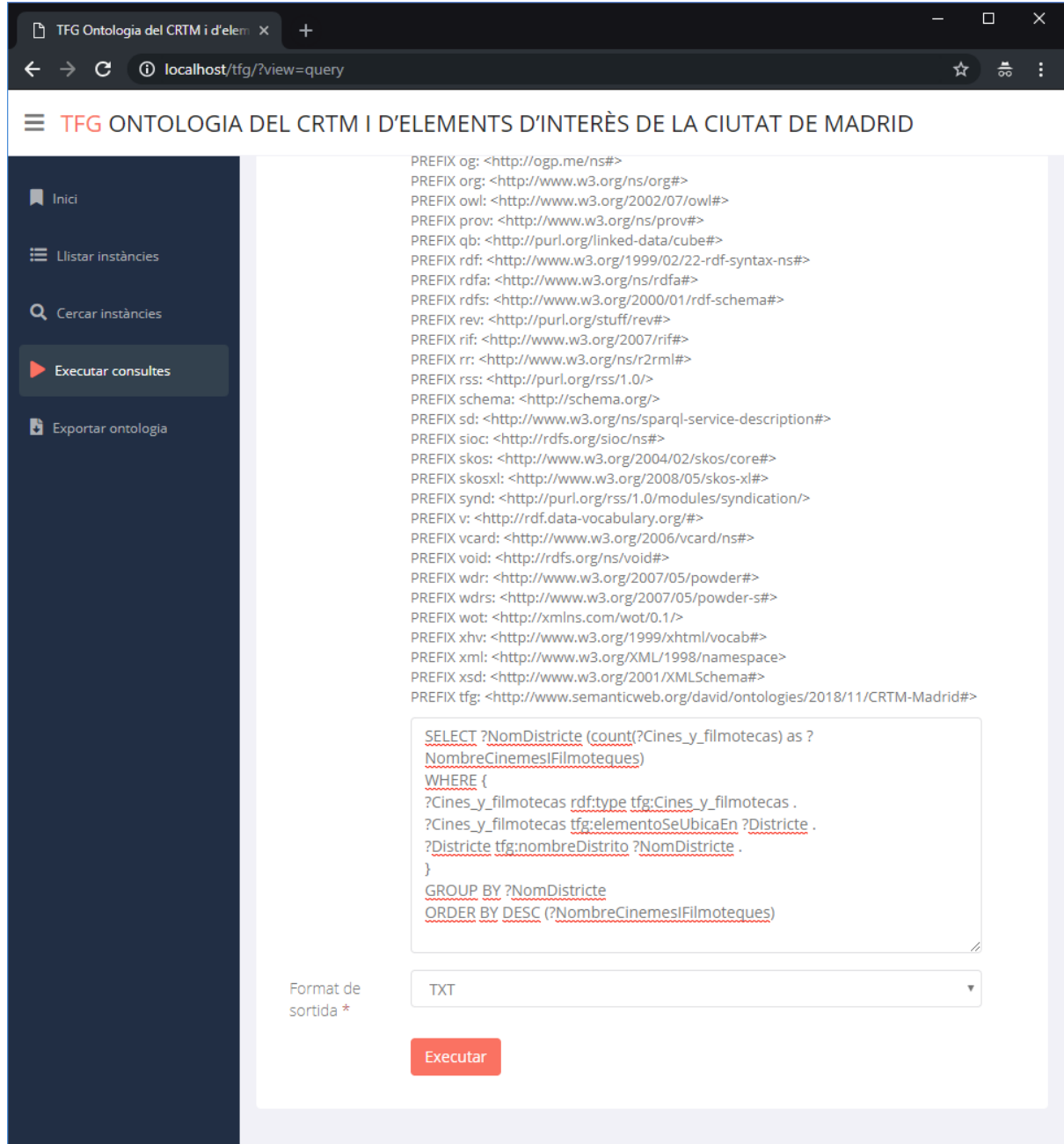



Figura 56: Execució d'una consulta al programari desenvolupat a mida



TFG ONTOLOGIA DEL CRTM I D'ELEMENTS D'INTERÈS DE LA CIUTAT DE MADRID

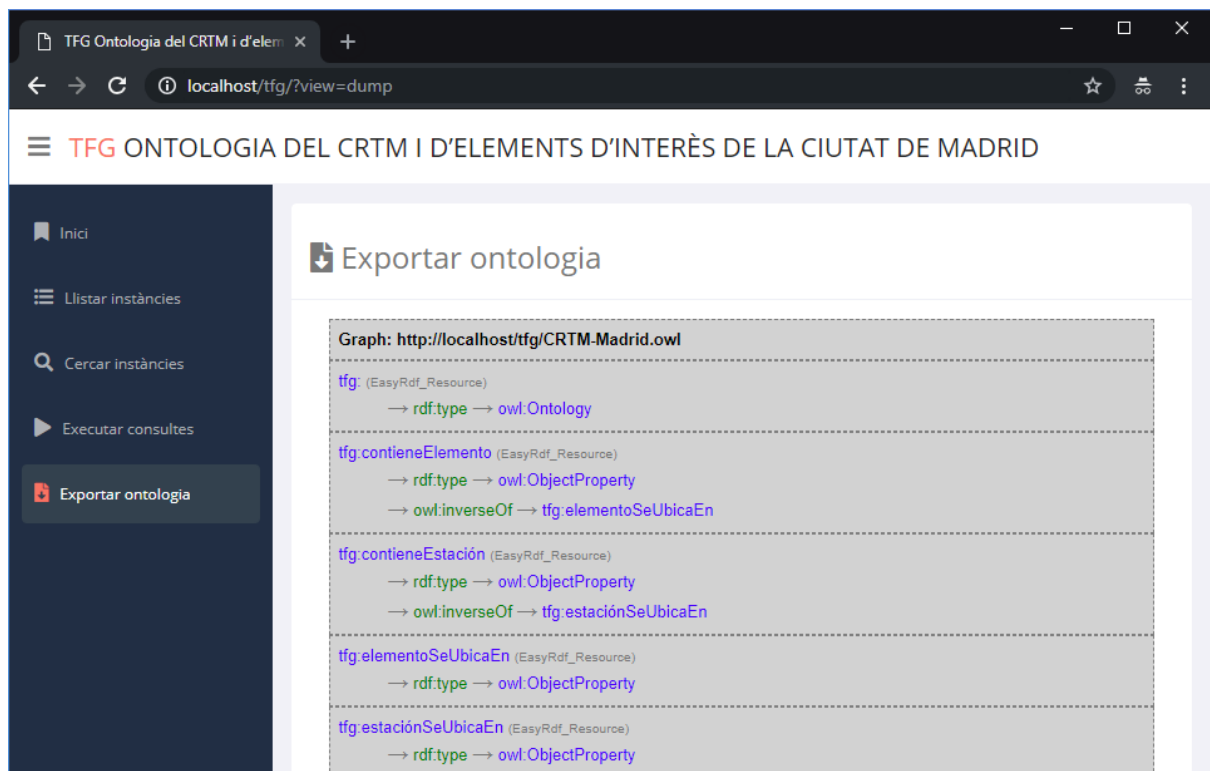
Executar consultes

?NomDistricte	?NombreCinemesIFilmoteques
"CHAMBERI"	"7"^^xsd:integer
"CENTRO"	"6"^^xsd:integer
"HORTALEZA"	"3"^^xsd:integer
"ARGANZUELA"	"3"^^xsd:integer
"MONCLOA-ARAVACA"	"3"^^xsd:integer
"SALAMANCA"	"2"^^xsd:integer
"FUENCARRAL-EL PARDO"	"2"^^xsd:integer
"CIUDAD LINEAL"	"1"^^xsd:integer
"VILLA DE VALLECAS"	"1"^^xsd:integer
"RETIRO"	"1"^^xsd:integer
"CHAMARTIN"	"1"^^xsd:integer
"SAN BLAS-CANILLEJAS"	"1"^^xsd:integer
"CARABANCHEL"	"1"^^xsd:integer

Tornar enrere

Figura 57: Resultats de la consulta en format de sortida text pla

Per últim, el programari també permet exportar el contingut de l'ontologia en format PHP i text pla:



TFG ONTOLOGIA DEL CRTM I D'ELEMENTS D'INTERÈS DE LA CIUTAT DE MADRID

Exportar ontologia

Graph: <http://localhost/tfg/CRTM-Madrid.owl>

tfg: (EasyRdf_Resource)

→ rdf:type → owl:Ontology

tfg:contieneElemento (EasyRdf_Resource)

→ rdf:type → owl:ObjectProperty

→ owl:inverseOf → tfg:elementoSeUbicaEn

tfg:contieneEstación (EasyRdf_Resource)

→ rdf:type → owl:ObjectProperty

→ owl:inverseOf → tfg:estaciónSeUbicaEn

tfg:elementoSeUbicaEn (EasyRdf_Resource)

→ rdf:type → owl:ObjectProperty

tfg:estaciónSeUbicaEn (EasyRdf_Resource)

→ rdf:type → owl:ObjectProperty

Figura 58: Exportació de l'ontologia en format de sortida HTML

5. Conclusions i línies de futur

5.1. Conclusions

És moment per fer la vista enrere i analitzar quins objectius proposats a l'inici del treball s'han assolit:

- Definir de forma introductòria la web semàntica, l'Open Data i tots els conceptes que l'envolten.
- Crear una ontologia amb les classes i relacions necessàries per al seu ús.
- Instruir-me en l'ús de noves eines informàtiques.
- Conèixer i aplicar correctament la metodologia adequada per l'ontologia creada.
- Nodrir-la amb els repositoris de dades que ofereixen el CRTM i la ciutat de Madrid.
- Poder-hi realitzar consultes per obtenir coneixement útil per la societat.
- Poder disposar lliurement d'una ontologia, que pugui ser utilitzada per tractar el coneixement generat o reutilitzada per generar-ne de nou.
- Desenvolupar un programari a mida que interactua amb l'ontologia creada.

Així doncs, s'han pogut assolir tots els objectius principals del projecte i un dels tres secundaris plantejats. D'altra banda, també cal destacar les lliçons apreses durant la realització del present treball: la importància d'una bona planificació i la corresponent previsió de possibles desviaments temporals a causa de qualsevol mena d'imprevist, la necessitat de dedicar el temps requerit per un profitós estat de l'art, l'anàlisi en profunditat de totes les eines i metodologies disponibles per crear l'ontologia, etc.

Si bé és cert que han sorgit diversos problemes durant el desenvolupament del projecte, aquests han estat solucionats satisfactòriament amb esforç i dedicació. Per exemple, el format d'alguns repositoris Open Data de la ciutat de Madrid no era l'adient i hi ha calgut un tractament previ a la importació, el programari Protégé va requerir l'assignació de molta memòria RAM (i hores) per generar les inferències de l'ontologia, algunes funcions de la llibreria EasyRdf van ser modificades per integrar-les correctament sota els objectius del programari desenvolupat a mida, etc.

A banda de tot l'esmentat anteriorment, si hagués de recomanar quelcom a un/a company/a que desitges desenvolupar un TFG d'aquesta àrea de la informàtica, seria l'ús d'Apache Jena per realitzar el nodriment de l'ontologia. Durant la planificació del projecte, vaig assignar tots els recursos de programació al desenvolupament a mida de l'aplicació web en PHP, ja que WebProtégé em facilitava la tasca d'importació de dades; en canvi, si hagués creat un programari utilitzant l'API d'Apache Jena, podria haver nodrit l'ontologia de forma personalitzada i realitzar-hi consultes de geoposicionament.

Pel que fa al vessant crític del projecte referent a la gestió dels recursos públics per part dels dirigents polítics, al treball s'analitzen dos dels exemples més evidents de la desigualtat existent a Madrid: les figures 41, 42 i 43 demostren que l'índex de satisfacció respecte a l'oferta cultural és directament proporcional al nombre de cinemes, filmoteques i museus de cada districte, i la ràtio de la figura 44 mostra un dels motius pels quals algú hagi d'utilitzar un vehicle privat en comptes del transport públic.

Cal tenir en compte que tots els habitants de la ciutat paguen, com a mínim, el mateix impost base independentment de la quantitat d'ús que realitzin (o puguin realitzar) sobre els seus recursos públics.

Finalment, el fet d'haver escollit aquesta àrea -completament desconeguda per mi mesos enrere- i haver declinat d'altres on tenia més experiència, m'ha possibilitat ampliar el meu coneixement acadèmic. Així doncs, m'agradaria destacar l'aprenentatge continu que ha suposat el desenvolupament d'aquest projecte i la satisfacció d'haver assolit els objectius plantejats al seu inici.

5.2. Línies de futur

Pel que fa a les línies de futur, malgrat haver superat les 300 hores previstes durant la planificació del projecte, el resultat obtingut no és quelcom definitiu (potser ho és si es considera com una versió inicial, però no pas com una que ofereixi funcionalitats que podria haver desenvolupat disposant del temps necessari).

Considero que els dos objectius secundaris no assolits en el present treball, l'enriquieren enormement. D'una banda, el fet de crear les classes necessàries a l'ontologia per poder aprofitar el geoposicionament, permetria obtenir resultats més precisos (a la figura 47 s'ha demostrat el potencial de GeoSPARQL realitzant una consulta a un servei extern); per exemple, l'ontologia permetria conèixer l'estació que es troba més a prop d'un determinat element d'interès de la ciutat. Cal destacar, però, que el fet d'agrupar els resultats per districte era necessari per contrastar-los amb l'índex de satisfacció dels habitants de Madrid.

D'altra banda, crear les classes per accessos, andanes, trams, etc. i nodrir-les amb les seves respectives instàncies a partir de les dades que proporciona el CRTM, permetria treballar amb el format GTFS proposat per *Google*. Així doncs, l'ontologia podria facilitar informació de l'hora exacta a la qual una persona s'hauria de trobar en una determinada andana si desitja dirigir-se a un element d'interès de Madrid mitjançant el transport públic de la ciutat.

Finalment, el programari desenvolupat a mida també podria incloure altres funcionalitats interessants. A banda de la conseqüent integració de les dues comentades anteriorment, l'aplicació, per exemple, podria posicionar en un mapa les instàncies obtingudes en comptes de mostrar un llistat de resultats. A més, les noves necessitats que requereixen les aplicacions del present beneficien l'ús d'aquest tipus de programari, ja que és capaç de gestionar grans fluxos de dades de forma molt eficient.

Fa temps que molts gegants del sector doten de semàntica els seus continguts a causa de la necessitat de millorar la cooperació entre màquines i humans. Ja que l'ús de la web semàntica creix any rere any per aprofitar el potencial que suposa la interrelació de dades de procedència molt diversa, tant de bo en un futur -no molt llunyà- pugui afrontar nous reptes en l'àmbit personal i pugui tenir l'oportunitat de millorar aquest projecte amb totes i cadascuna de les funcionalitats esmentades anteriorment.

Bibliografia

Berners-Lee, T. (1999). "Weaving the WEB: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor". New York: Harper Collins Publisher.

Berners-Lee, T., Hendler, J., Lassila, O. (2001). "The Semantic Web". Scientific American.

Fensel, D., van Harmelen, F., Horrocks, I., McGuinness, D. L., Patel-Schneider, P. F. (2001). "OIL: An Ontology Infrastructure for the Semantic Web". IEEE Intelligent Systems.

Grüninger, M. (1996). "Designing and Evaluating Generic Ontologies".

Grüninger, M., Fox, M. S. (1994). "The Design and Evaluation of Ontologies for Enterprise Engineering".

Grüninger, M., Fox, M. S. (1995). "Methodology for the Design and Evaluation of Ontologies".

Guarino, N. (1997). "Understanding, Building and Using Ontologies". International Journal of Human Computer Studies.

Guarino, N. (1998). "Formal Ontology and Information Systems". IOS Press.

Hendler, J., Berners-Lee, T., Miller, E. (2002). "Integrating Applications on the Semantic Web". Journal of the Institute of Electrical Engineers of Japan.

McGuinness, D. L. (2004). "Question Answering on the Semantic Web". IEEE Intelligent Systems.

McGuinness, D. L., Fikes, R., Hendler, J. (2002). "DAML+OIL: An Ontology Language for the Semantic Web". IEEE Intelligent Systems.

Noy, N. F., McGuinness, D. L. (2001). "Ontology Development 101: A Guide to Creating Your First Ontology". Stanford Knowledge Systems Laboratory Technical Report.

Semantic Scripting. [Recurs en línia]. [Data de consulta: gener de 2019].

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

Semantic Web. [Recurs en línia]. [Data de consulta: gener de 2019].

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

SWoogle: Semantic Web Search. [Recurs en línia]. [Data de consulta: gener de 2019].

<<http://swoogle.umbc.edu/>>

W3C. "Introducción al uso de la Web Semántica". Taller SWAD-Europe 13 Junio 2004, Madrid, España. [Recurs en línia]. [Data de consulta: gener de 2019].

<<http://www.w3.org/2001/sw/Europe/events/200406-esp/>>

W3C. “Guía breve de Web Semántica”. [Recurs en línia]. [Data de consulta: gener de 2019].

<<http://www.w3c.es/divulgacion/guiasbreves/WebSemantica/>>

W3C. “SPARQL Protocol for RDF”. [Recurs en línia]. [Data de consulta: gener de 2019].

<<http://www.w3.org/TR/rdf-sparql-protocol/>>

W3C. “SPARQL Query Language for RDF”. [Recurs en línia]. [Data de consulta: gener de 2019].

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

W3C. “Web Ontology Language (OWL) Use Cases and Requirements, W3C Recommendation”. [Recurs en línia]. [Data de consulta: gener de 2019].

<<http://www.w3.org/TR/webont-req/>>

W3C. “OWL Web Ontology Language Reference, W3C Recommendation”. [Recurs en línia]. [Data de consulta: gener de 2019].

<<http://www.w3.org/TR/owl-ref/>>

W3C. “OWL Web Ontology Language Semantics and Abstract Syntax, W3C Recommendation”. [Recurs en línia]. [Data de consulta: gener de 2019].

<<http://www.w3.org/TR/owl-absyn/>>

W3C. “OWL Web Ontology Language Overview, W3C Recommendation”. [Recurs en línia]. [Data de consulta: gener de 2019].

<<http://www.w3.org/TR/owl-features/>>

W3C. “OWL Web Ontology Language Test Cases, W3C Recommendation”. [Recurs en línia]. [Data de consulta: gener de 2019].

<<http://www.w3.org/TR/owl-test/>>

W3C. “OWL Web Ontology Language Guide, W3C Recommendation”. [Recurs en línia]. [Data de consulta: gener de 2019].

<<http://www.w3.org/TR/owl-guide/>>

W3C. “Semantic Web Activity”. [Recurs en línia]. [Data de consulta: gener de 2019].

<<http://www.w3.org/2001/sw/>>

W3C. “Resource Description Framework (RDF)”. [Recurs en línia]. [Data de consulta: gener de 2019].

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

Annexos

Annex A: Glossari

A continuació es descriuen els termes i acrònims utilitzats en el present treball:

- **API:** *Application Programming Interface*, és el conjunt de funcions i mètodes que ofereixen les llibreries de desenvolupament per ser utilitzats en altres programaris.
- **BIOS:** *Basic Input Output System*, és un programari que localitza i reconeix tots els dispositius necessaris per a carregar el sistema operatiu en la memòria RAM.
- **CDATA:** *Character DATA*, és una secció que comprèn una cadena de caràcters que l'analitzador sintàctic no ha d'interpretar com a contingut etiquetat en un document XML.
- **CPU:** *Central Processing Unit*, és la part del maquinari en la qual es troben els elements que manegen les ordres que executen funcions com càlculs binaris, processament de dades...
- **CRTM:** Consorcio Regional de Transportes de Madrid.
- **CSV:** *Comma-Separated Values*, és un tipus de document per representar dades.
- **DTD:** *Document Type Definition*, és una descripció d'estructura i sintaxi d'un document XML.
- **ED50:** *European Datum 1950*, és un antic sistema de referència geodèsic empleat a Europa.
- **GeoRSS:** és un conjunt d'estàndards basat en la família dels RSS, que serveix per a representar informació geogràfica mitjançant l'ús de capes.
- **GeoSPARQL:** és un estàndard basat en SPARQL que s'utilitza per realitzar consultes de dades geoespacial.
- **GPX:** *GPS eXchange*, és un format d'intercanvi de dades GPS basat en XML.
- **GTFS:** *General Transit Feed Specification*, és una especificació que defineix un format comú per als horaris de transport públic i la seva informació geogràfica associada.
- **HTML:** *HyperText Markup Language*, és un llenguatge utilitzat per crear pàgines web.
- **IDE:** *Integrated Drive Electronics*, és un estàndard d'interfícies per a la connexió de dispositius d'emmagatzematge massiu de dades i unitats de discos òptics.
- **IRI:** *Internationalized Resource Identifier*, és una seqüència de caràcters que identifica un recurs i s'amplia al conjunt de caràcters universal codificat.
- **ISO:** és un arxiu informàtic on s'emmagatzema una un sistema d'arxius.
- **JDK:** *Java Development Kit*, és un programari que proveeix eines de desenvolupament per a la creació d'aplicacions en Java.
- **JSON:** *JavaScript Object Notation*, és un format de text d'intercanvi de dades.
- **JVM:** *Java Virtual Machine*, és una màquina virtual capaç d'interpretar i executar instruccions expressades en Java.
- **KML:** *Keyhole Markup Language*, és un llenguatge basat en XML per a representar dades geogràfiques en tres dimensions.
- **KMZ:** *Keyhole Markup Zip*, és un arxiu que n'empaqueta diversos KML.

- **MDB:** *Microsoft Access Database*, és una extensió d'arxiu utilitzada en determinades versions de bases de dades de *Microsoft Access*.
- **OGC:** *Open Geospatial Consortium*, és una organització internacional que vetlla pel desenvolupament i la implementació d'estàndards lliures per al contingut geoespacial
- **OWL:** *Web Ontology Language*, és un llenguatge utilitzat per compartir ontologies a la *World Wide Web*.
- **PAC:** Prova d'Avaluació Contínua.
- **PHP:** *Hypertext Processor*, és un llenguatge de programació dissenyat pel desenvolupament de contingut dinàmic a pàgines i aplicacions web.
- **RAM:** *Random Access Memory*, és la memòria d'accés aleatori utilitzada per dispositius, sistemes operatius i la majoria de programaris.
- **RDF:** *Resource Description Framework*, és una família d'especificacions de la W3C dissenyat com un model de dades per a metadades. Facilita l'intercanvi de dades a la *World Wide Web*.
- **RFC:** *Request For Comments*, és un conjunt de documents que serveixen de referència pel que fa a normes, estàndards, tecnologies, protocols, etc.
- **RSS:** *Really Simple Syndication*, és un format XML per distribuir contingut a la *World Wide Web*.
- **SHP:** *Shapefile*, és un format d'arxiu propietari per intercanviar dades espacials.
- **SPARQL:** *SPARQL Protocol And RDF Query Language*, és un llenguatge estandarditzat per a la consulta de grafs i ontologies RDF.
- **TXT:** *TeXT*, és un format d'arxiu que permet emmagatzemar informació sense format.
- **URI:** *Uniform Resource Identifier*, és una seqüència de caràcters que identifica un recurs.
- **UTF8:** *8-bit Unicode Transformation Format*, és un format de codificació de caràcters Unicode.
- **WAR:** *Web application ARchive*, és un arxiu utilitzat per a distribuir una col·lecció de *JavaServer Pages*, *servlets*, classes Java, arxius XML, biblioteques de *tags* i pàgines web estàtiques que junts constitueixen una aplicació web.
- **WGS84:** *World Geodetic System 1984*) és un sistema de coordenades geogràfiques mundial que permet localitzar qualsevol punt de la Terra.
- **WMS:** *Web Map Service*, és un servei de l'OGC que produeix mapes de dades de forma dinàmica a partir d'informació geogràfica.
- **WWW:** *World Wide Web*, és un sistema de distribució de documents d'hipertext o hipermèdia interconnectats i accessibles via Internet.
- **W3C:** *World Wide Web Consortium*, és un organisme neutre d'àmbit internacional que treballa per generar estàndards que assegurin el creixement de la *World Wide Web* a llarg termini.
- **XLS:** *Microsoft Excel File Format*, és una extensió d'arxiu utilitzada en determinades versions de *Microsoft Excel*.
- **XML:** *eXtensible Markup Language*, és un llenguatge obert de marques extensible que s'utilitza per emmagatzemar dades en forma llegible i de forma estructurada dins d'etiquetes.

Annex B: Lliurables del projecte

El llistat d'arxius lliurats és el següent:

1. **Memòria:** s'inclou la present memòria en format PDF.
2. **Ontologia:** conté l'ontologia inferida durant el desenvolupament d'aquest projecte i, com extres, la versió buida creada amb el programari Protégé i la nodrida amb WebProtégé.
3. **Repositoris de dades:** s'inclouen tots els arxius en format CSV utilitzats a l'hora d'introduir a l'ontologia les dades escollides del CRTM i la ciutat de Madrid.
4. **Aplicació a mida:** conté un arxiu en format ZIP amb el programari web desenvolupat a mida en llenguatge PHP.
5. **Observatorio de la Ciudad (2017):** s'inclouen, districte a districte, tots els arxius en format PDF corresponents a l'anàlisi de resultats pel que fa a la satisfacció dels habitants de la ciutat de Madrid amb els seus serveis públics.

Annex C: Instal·lació del programari XAMPP

XAMPP és un paquet de programari lliure, que ajuda a instal·lar fàcilment el sistema de gestió de bases de dades MariaDB, el servidor web Apache i els intèrprets per a llenguatges de script PHP i Perl. Per aconseguir-ho en el sistema operatiu utilitzat en el present treball, primerament cal realitzar la descàrrega des de la seva pàgina web oficial: <https://www.apachefriends.org/index.html>.

Tot seguit, cal executar les següents ordres pel Terminal per assignar a l'arxiu permisos d'execució i iniciar-lo:

```
cd /home/[nomUsuari]/Downloads
chmod 755 [nomPaquet]
sudo ./[nomPaquet]
```

A continuació s'obre un assistent d'instal·lació que cal seguir amb les opcions seleccionades per defecte. Un cop hagi finalitzat, cal executar la següent ordre pel Terminal:

```
sudo /opt/lampp/lampp start
```

Per verificar l'estat correcte de la instal·lació, el següent enllaç ha de mostrar la informació de benvinguda del programari en el navegador web: <http://localhost>.

Annex D: Currículum Vitae

Sóc en David Mor España i vaig néixer a Barcelona 29 anys enrere. Després de finalitzar el batxillerat tecnològic a l'institut Juan Manuel Zafra, vaig estudiar dos cicles de formació professional de grau superior a Jesuïtes El Clot: Desenvolupament d'Aplicacions Informàtiques i Administració de Sistemes Informàtics.

Pel que fa a l'àmbit de la informàtica, vaig treballar durant dos anys al departament de suport IT de la companyia AdQualis i, a continuació, cinc més com a *Full Stack PHP Developer* a Ngloba Devices. Finalment, l'any 2015 vaig decidir reprendre la meva formació acadèmica i estudiar el grau d'Enginyeria Informàtica a la UOC.