

# **Ontologia de punts d'accés wifi emprant repositoris de dades obertes de les ciutats de New York i Buenos Aires**

**Autor: Teresa Tejero Albiñana**

Tutor: Felipe Geva Urbano

Grau Enginyeria Informàtica

12 de gener de 2018

# Crèdits/Copyright



Aquesta obra està subjecta a una llicència de Reconeixement-CompartirIgual

[3.0 Espanya de CreativeCommons](https://creativecommons.org/licenses/by-sa/3.0/)

## FITXA DEL TREBALL FINAL

|  |   |
|--|---|
| <b>Títol del treball:</b>  | <i>Ontologia de punts d'accés wifi emprant repositoris de dades obertes de les ciutats de Buenos Aires i New York</i> |
| <b>Nom de l'autor:</b>   | <i>Teresa Tejero Albiñana</i>   |
| <b>Nom del col·laborador/a docent:</b>   | <i>Felipe Geva Urbano</i>   |
| <b>Data de lliurament (mm/aaaa):</b>   | <i>01/2018</i>  |
| <b>Titulació o programa:</b>   | <i>Pla d'estudis</i>  |
| <b>Àrea del Treball Final:</b>   | <i>XML i Web Semàntica</i>  |
| <b>Idioma del treball:</b>   | <i>Català</i>   |
| <b>Resum del Treball (màxim 250 paraules)</b>  |   |
| <p>El terme <i>ontologia</i> té el seu origen en la filosofia. Ontologia significa «<i>l'estudi del ser</i>». Aquest terme prové del grec <i>ontos</i> (οντος), que significa <i>ser</i> i de <i>logos</i> (λόγος) que significa <i>estudi, discurs, ciència, teoria</i>.</p> <p>Una ontologia és una jerarquia de conceptes amb atributs i relacions, que defineix una terminologia consensuada per definir xarxes semàntiques d'unitats d'informació interrelacionades i amb certes regles.</p> <p>Aquest treball es basa en el camp de la Web Semàntica. La web semàntica és un conjunt d'activitats desenvolupades per World Wide Web Consortium (<i>W3C</i>) per la creació de tecnologies per publicar dades llegibles per aplicacions informàtiques.</p> <p>Aquest treball el que pretén és generar una ontologia sobre punts d'accés wifi que permeti representar com estan distribuïts i la seva informació associada d'una forma estructurada. Aquesta ontologia sobre punts d'accés wifi s'omplirà de dades extretes de repositoris de dades lliures, és a dir, Open Data.</p> <p>Per a realitzar l'ontologia d'aquest treball s'utilitzarà el programa Protégé i WebProtégé. Amb Protégé es realitzarà l'ontologia i amb WebProtégé es realitzarà la importació de les dades que estan en format CSV.</p> <p>Una vegada omplerta l'ontologia de les dades del repositori Kaggle, es realitzaran consultes sobre aquesta. El llenguatge SPARQL, acrònim de l'anglès <i>SPARQL Protocol and RDF Query Language</i>, és un llenguatge estandarditzat per a la consulta de grafs RDF, normalitzat pel RDF Data Access Working Group del World Wide Web Consortium.</p> |   |

**Abstract (in English, 250 words or less):**

The Semantic Web is an extension of the WWW in which the meaning of information and services is well defined. Therefore, it allows understanding and satisfies the requests of the people and the machines that use web contents.

The web is on the verge of change with the emergence of the Semantic Web, since the information that appears on the Internet can be interpreted by computers without the need of human intervention. For this to happen, the information contained in web pages must be codified by ontologies.

An ontology is a formal definition of types, properties and relationships among entities that basically and fundamentally exist for a domain of discussion in particular. An ontology classifies the variables required by some computing set and establishes the relationship among them.

The purpose of this project is to create an ontology of WI-Fi access points in the cities of New York and Buenos Aires.

This project consists of three phases: in the first phase, I have analyzed the data in CSV format. In the second phase, I have done the ontology design and finally I have carried out the implementation and population of the ontology.

For the analysis phase, the data that have been studied were obtained from the open data repository. To carry out the implementation of the ontology, I have used tools for the creation and manipulation of ontologies. These tools are Protegé and WebProtege. To complete the project and demonstrate how the ontology was performed, ontology queries have been performed using The SPARQL language.

**“Aprender es descubrir que algo es posible”**

*Fritz Perls*

## Abstract

Una ontologia és una definició formal de tipus, propietats, i relacions entre entitats que realment o fonamentalment existeixen per a un domini de discussió en particular. Una ontologia cataloga les variables requerides per algun conjunt de computació i estableix les relacions entre ells.

La finalitat d'aquest treball és crear una ontologia de punts d'accés wifi de les ciutats de New York i Buenos Aires.

Per realitzar aquest treball s'han utilitzat eines per a la creació d'ontologies: Protégé i WebProtégé. Les dades s'han obtingut de repositoris de dades obertes.

Aquest treball està format per tres fases. En la primera fase he analitzat les dades. En la segona fase he realitzat el disseny de l'ontologia i per última he realitzat la implementació i població de l'ontologia.

Per finalitzar el treball, s'han realitzat consultes a l'ontologia usant el llenguatge SPARQL.

### Paraules clau

Les paraules clau que poden definir aquest Treball Final de Grau són:

**ontology, semantic, opendata, owl, rdf, xml, wifi**

# Índex

|   |           |
|---|-----------|
| <b>1. Introducció</b> .....   | <b>11</b> |
| <b>1.1. Context i justificació del projecte</b> .....               | <b>11</b> |
| <b>1.2. Objectius del treball</b> .....                             | <b>12</b> |
| 1.2.1. Objectius generals.....                                      | 12        |
| 1.2.2. Objectius principals .....                                   | 12        |
| 1.2.3. Metodologia i procés de treball .....                        | 12        |
| <b>1.3. Planificació</b> .....                                      | <b>13</b> |
| 1.3.1. Tasques .....  | 13        |
| 1.3.2. Fites .....  | 14        |
| 1.3.3. Diagrama de Gantt.....                                       | 14        |
| <b>Conceptes bàsics</b> .....                                       | <b>15</b> |
| <b>1.5. Web semàntica</b> .....                                     | <b>15</b> |
| <b>1.6. Ontologia</b> .....   | <b>17</b> |
| <b>1.7. Llenguatges de representació d'ontologies</b> .....         | <b>18</b> |
| 1.7.1. XML.....   | 18        |
| 1.7.2. RDF.....   | 18        |
| 1.7.3. OWL.....   | 18        |
| 1.7.4. SPARQL.....  | 18        |
| 1.7.5. OPEN DATA.....   | 18        |
| <b>2. Disseny de l'ontologia</b> .....                              | <b>21</b> |
| <b>2.1. Metodologia seguida per crear l'ontologia</b> .....         | <b>21</b> |
| <b>2.2. Eines per a la creació d'ontologies</b> .....               | <b>25</b> |
| 2.2.1. Protégé.....   | 26        |
| 2.2.2. WebProtégé.....  | 26        |
| 2.2.3. Apache Jena .....  | 26        |
| 2.2.4. Elecció de l'eina per al desenvolupament de l'ontologia..... | 27        |

|             |  |           |
|-------------|--|-----------|
| <b>3.</b>   | <b>Implementació de l'ontologia</b>          | <b>29</b> |
| <b>3.1.</b> | <b>Instal·lació de les eines necessàries</b> | <b>29</b> |
| 3.1.1.      | Protégé                                      | 29        |
| 3.1.2.      | WebProtégé                                   | 30        |
| 3.1.3.      | Recursos                                     | 30        |
| <b>3.2.</b> | <b>Creació de l'ontologia</b>                | <b>30</b> |
| 3.2.1.      | Classes                                      | 31        |
| 3.2.2.      | Propietats                                   | 33        |
| 3.2.3.      | Instàncies                                   | 35        |
| <b>3.3.</b> | <b>CONSULTES SPARQL</b>                      | <b>41</b> |
| 3.3.1.      | Consultes a l'ontologia                      | 42        |
| <b>4.</b>   | <b>Conclusions i línies de futur</b>         | <b>46</b> |
| <b>4.1.</b> | <b>Conclusions</b>                           | <b>46</b> |
| <b>4.2.</b> | <b>Línies de futur</b>                       | <b>47</b> |
|             | <b>Bibliografia</b>                          | <b>48</b> |
|             | <b>Annexos</b>                               | <b>50</b> |



# Figures i taules

## Índex de figures

|   |    |
|---|----|
| Figura 1: Diagrama de Gantt del treball .....   | 14 |
| Figura 2: Estructura de la Web Semàntica per W3C.....   | 16 |
| Figura 3: Bus de la Web Semàntica.....  | 16 |
| Figura 4: Puntuaciones Open Data. Font Open Data Barometer.....                                   | 20 |
| Figura 5: Interfície cercador ontologies Swoogle .....  | 22 |
| Figura 6: Jerarquia de classes de l'ontologia .....   | 23 |
| Figura 7: Arquitectura Framework Apache Jena.....   | 27 |
| Figura 8: Interfície Protégé.....   | 31 |
| Figura 9: Barra de navegació de Protégé .....   | 31 |
| Figura 10: Classes creades amb Protégé.....   | 32 |
| Figura 11: Data Properties amb Protégé.....   | 33 |
| Figura 12: Descripció de la propietat id.....   | 33 |
| Figura 13: Object Properties amb Protégé.....   | 34 |
| Figura 14: Descripció de la relació estaFormada .....   | 35 |
| Figura 15: Diagrama ontologia.....  | 35 |
| Figura 16: Pantalla Principal Ontologia WebProtégé .....  | 36 |
| Figura 17: Creació instàncies de les classes.....   | 36 |
| Figura 18: Càrrega d'arxius CSV amb WebProtégé.....   | 37 |
| Figura 19: Importació de dades amb WebProtégé.....  | 37 |
| Figura 20: Exemple creació instàncies.....  | 38 |
| Figura 21: Exemple instàncies de la classe Zona .....   | 39 |
| Figura 22: Exemple instància adreça "World Trade Center" .....                                    | 40 |
| Figura 23: Activar Raonador Hermit.....   | 40 |
| Figura 24: Show Inference .....   | 40 |
| Figura 25: Pestanya SPARQL Query en Protégé 4.3 .....   | 41 |
| Figura 26: Resultat consulta "Mostra les ciutats i les seves zones" .....                         | 42 |
| Figura 27: Resultat consulta amb modificadors "Mostra les ciutats i les seves zones" .....        | 43 |
| Figura 28: Resultat consulta "Mostra les zones de la ciutat de New York" .....                    | 44 |
| Figura 29: Resultat consulta "Mostra la localització del punt wifi amb l'identificador 705" ..... | 44 |

Figura 30: Resultat consulta "Mostra el número de punts wifi d'una ciutat, en concret de "Staten Island"" ..... 45

Figura 31: Resultat consulta "Mostra les coordenades del punt wifi amb id 822" ..... 45

## Índex de taules

Taula 1: Tasques del treball ..... 13

Taula 2: Fites del treball ..... 14

Taula 3: Camps classe PuntWifi ..... 31

Taula 4: Camps classe Adreça ..... 31

Taula 5: Camps classe Coordenades ..... 32

Taula 6: Camps classe Ciutat ..... 32

Taula 7: Camps classe Zona ..... 32

Taula 8: Camps classe Zona ..... 32

Taula 9: Propietats tipus Object Property ..... 34

# 1.Introducció

## 1.1. Context i justificació del projecte

Aquest document descriu el pla de projecte final de grau "*Ontologia de punts d'accés wifi emprant repositoris de dades obertes de la ciutat de New York*". A aquest document es descriuen els objectius que s'han proposat, la planificació que s'ha estudiat per realitzar-lo i el procés de desenvolupament de l'ontologia seguint la metodologia escollida.

Es tracta d'un projecte en el qual es pretén realitzar una investigació sobre la web semàntica. Concretament s'ha centrat en les eines que es poden utilitzar per a realitzar ontologies i poder importar dades a aquestes ontologies creades prèviament. Les dades s'han d'importar de fitxers que tenen format csv. El context que s'ha elegit per a la realització de l'ontologia és la ubicació de punts wifi de la ciutat de New York.

## 1.2. Objectius del treball

### 1.2.1. Objectius generals

L'objectiu principal d'aquest treball és crear una ontologia dels punts wifi de la ciutat de New York, que permeti emmagatzemar les ciutats i zones de New York i la ubicació dels seus punts wifi. Per poder realitzar aquesta ontologia es realitzarà mitjançant el fitxer de dades en format CSV que es pot trobar a Kaggle<sup>1</sup>.

### 1.2.2. Objectius principals

Els objectius principals d'aquest treball són:

1. Obtenir el fitxer de dades en format CSV dels punts wifi de les ciutats.
2. Crear o reutilitzar una ontologia per saber la ubicació dels punts wifi de les ciutats.
3. Poblar l'ontologia creada amb les dades dels fitxers amb format CSV.
4. Realitzar consultes amb SPARQL a l'ontologia creada.

### 1.2.3. Metodologia i procés de treball

Com qualsevol creació de programari es poden aplicar metodologies a la creació d'ontologies. Cal aclarir que no existeix una metodologia estàndard. Diferents projectes reals de desenvolupament d'ontologies han donat lloc a diverses aproximacions.

La metodologia elegida per a realitzar aquest treball ha estat *Ontology Development 101* proposada per Noy & McGuinness (2001) a la Universitat de Stanford.

Les principals recomanacions d'aquesta metodologia són 7:

1. Determinar el domini i cobertura de l'ontologia
2. Considerar reutilitzar ontologies existents
3. Enumerar els termes importants en l'ontologia
4. Definir les classes i la seva jerarquia
5. Definir les propietats de les classes
6. Definir les característiques de les propietats

---

<sup>1</sup> Kaggle: <https://www.kaggle.com/>, consultat el 25/10/17.

### 1.3. Planificació

Per la planificació del projecte s'han establert les fites en les entregues de les PACs establertes per al treball. Per a la planificació del projecte s'ha utilitzat el diagrama de Gantt i per a la seva representació s'ha utilitzat l'eina Gantt Project.

Per a cadascuna de les PAC's s'estableixen tasques a realitzar. A mesura que es proporcionen els enunciats es continuarà amb la millora de la planificació de les PAC.

#### 1.3.1. Tasques

El primer pas per a la planificació de qualsevol treball és obtenir les tasques que formaran part d'aquest. Les tasques previstes per la realització d'aquest treball són les següents:

| Tasques   |
|---|
| <b>PAC 1</b>                                      |
| Lectura del pla docent i el mòdul 1 del material  |
| Lectura de l'orientació del TFC                   |
| Recerca d'informació sobre ontologies             |
| Elaboració del pla de treball                     |
| Lliurament PAC 1                                  |
| <b>PAC 2</b>                                      |
| Lectura dels mòduls 2 i 3                         |
| Estudi en profunditat de les ontologies           |
| Cercar informació sobre metodologies d'ontologies |
| Elaboració de l'estat de l'art                    |
| Lliurament PAC 2                                  |
| <b>PAC 3</b>                                      |
| Analitzar repositori Kaggle                       |
| Obtenir dades del repositori                      |
| Crear ontologia                                   |
| Implementar script                                |
| Interrogar a l'ontologia                          |
| Lliurament PAC 3                                  |
| <b>PAC 4</b>                                      |
| Redacció de la memòria                            |
| Elaborar la presentació                           |
| Lliurament PAC 4                                  |
| <b>Debat Virtual</b>                              |
| Realitzar intervencions al debat virtual          |
| Respondre a les intervencions                     |
| Lliurament Debat Virtual                          |

Taula 1: Tasques del treball

### 1.3.2. Fites

Com a fites d'aquest treball s'han definit les entregues de cadascuna de les PAC que es realitzaran al llarg del semestre. Les **fites** del treball són les següents:

| Fita                     | Data     |
|--------------------------|----------|
| Inici de curs            | 20/09/17 |
| Lliurament PAC 1         | 04/10/17 |
| Lliurament PAC 2         | 29/10/17 |
| Lliurament PAC 3         | 17/12/17 |
| Lliurament PAC 4         | 11/01/18 |
| Lliurament debat virtual | 26/01/18 |

Taula 2: Fites del treball

### 1.3.3. Diagrama de Gantt

A partir de les tasques descrites a la Taula 1 i les fites especificades a la Taula 2, s'ha obtingut el diagrama de Gantt. A continuació es mostra el diagrama de Gantt corresponent a les PAC's en les que està dividit el treball.

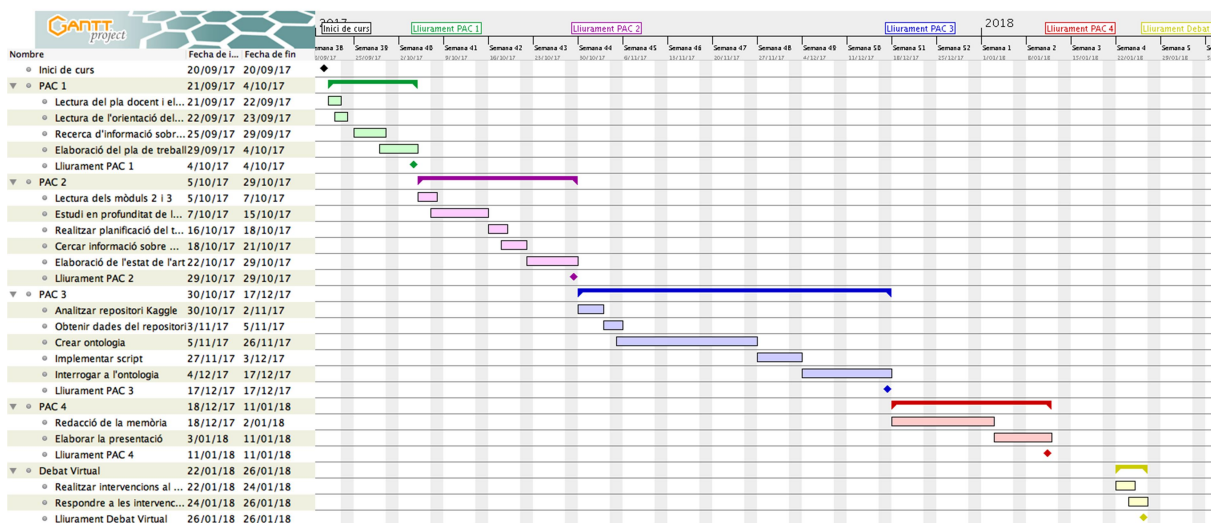


Figura 1: Diagrama de Gantt del treball

## 2. Conceptes bàsics

Abans de començar amb la implementació del projecte, cal fer una introducció dels conceptes bàsics sobre el tema que es tracta, per tal de conèixer la Web Semàntica i com està composta.

### 2.1. Web semàntica

Tenint en compte el que afirma W3C, la Web Semàntica és:

*"La Web Semántica es una Web extendida, dotada de mayor significado en la que cualquier usuario en Internet podrá encontrar respuestas a sus preguntas de forma más rápida y sencilla gracias a una información mejor definida. Al dotar a la Web de más significado y, por lo tanto, de más semántica, se pueden obtener soluciones a problemas habituales en la búsqueda de información gracias a la utilización de una infraestructura común, mediante la cual, es posible compartir, procesar y transferir información de forma sencilla. Esta Web extendida y basada en el significado, se apoya en lenguajes universales que resuelven los problemas ocasionados por una Web carente de semántica en la que, en ocasiones, el acceso a la información se convierte en una tarea difícil y frustrante"<sup>2</sup>.*

Segons el W3C es pot definir a la Web Semàntica com l'evolució de la Web actual en la qual la informació que contenen les pàgines web ha d'estar representada de forma formal per a que no sols pugui ser entesa per les persones, sinó que els ordinadors puguin processar-la per a ser entesa per ells.

La web semàntica es basa en dos punts fonamentals: la descripció del significat i la manipulació automàtica d'aquestes descripcions. La descripció del significat s'articula amb:

- **Semàntica:** segons la RAE, semàntica és *"la disciplina que estudia el significado de las unidades lingüísticas y de sus combinaciones"*<sup>3</sup>. En el context d'aquest treball, semàntica es pot definir com la dotació de significat interpretable per part de les màquines.
- **Metadades:** són dades que descriuen altres dades per facilitar la seva catalogació. Les metadades agreguen semàntica als documents, d'aquesta manera es facilita als motors de cerca la indexació de les dades.
- **Ontologies:** és un conjunt de termes i relacions entre termes que descriuen un domini d'aplicació concret. El seu objectiu és la creació d'un diccionari de termes compartits i comprensibles per a diferents aplicacions i comunitats que permeten la interoperabilitat.

---

<sup>2</sup> Guía Breve de Web Semántica: <https://www.w3c.es/Divulgacion/GuiasBreves/WebSemantica>, consultat 1/10/2017.

<sup>3</sup> Real Academia de la Lengua Española: [http://dle.rae.es/srv/fetch?id=XVRDns5\\_](http://dle.rae.es/srv/fetch?id=XVRDns5_), consultat 1/10/2017.

El W3C proposa una estructura per a la Web Semàntica. Els principis de Web Semàntica s'implementen en les capes de tecnologies i estàndards web. Les capes es presenten en la Figura 2.

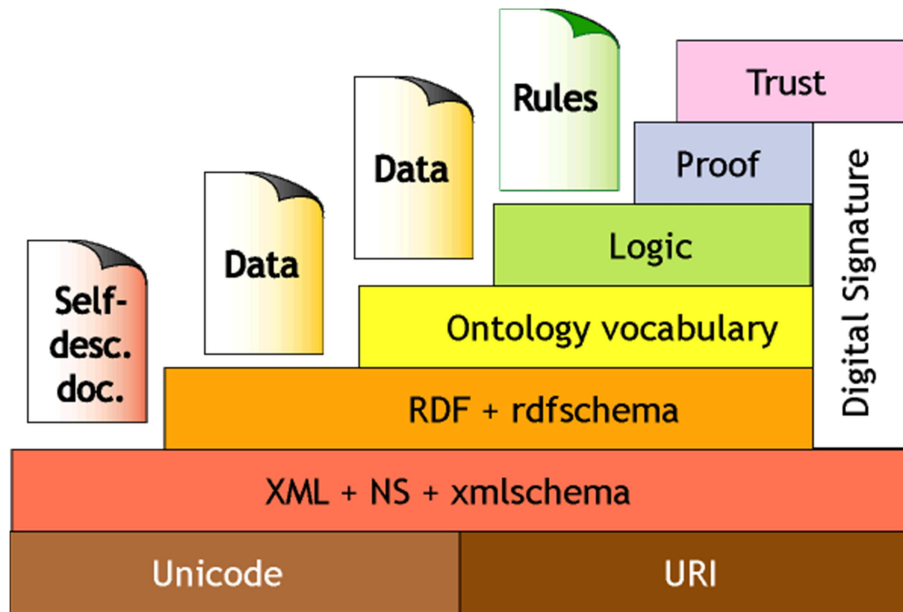


Figura 2: Estructura de la Web Semàntica per W3C

En la figura 1, es poden observar, els nivells més baixos de l'arquitectura de la Web Semàntica. La figura reflecteix que els models RDF es poden codificar com documents XML i que altres llenguatges d'ontologies RDF Schema i OWL són models RDF. També és interessant observar que les tecnologies de Signatura Digital també apareixen en la figura.

Cal, doncs, crear una ontologia o biblioteca de vocabularis descriptius/semàntics, definits en format RDF i ubicats a la web per a determinar el significat contextual d'una paraula per mitjà de la consulta a l'ontologia apropiada.

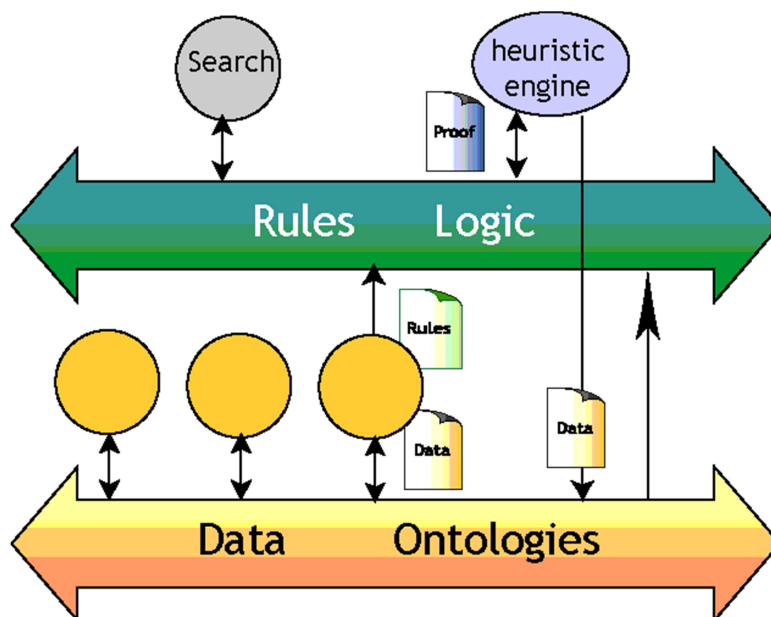


Figura 3: Bus de la Web Semàntica



D'aquesta manera, agents intel·ligents i programes autònoms podrien rastrejar la web de forma automàtica i localitzar, exclusivament, les pàgines que facin referència a la paraula buscada amb el significat i concepte precisos amb el que interpretem aquest terme.

En conclusió, es pot afirmar que, l'objectiu de la Web Semàntica és que es passi d'una col·lecció de documents a convertir-se en una base de coneixement.

## 2.2. Ontologia

El terme ontologia prové del grec i està format per "ὄντος" o "ontos" que vol dir "l'ésser", més "λόγος" o "logos" que significa "ciència", "estudi" o "teoria". Aquest camp ve del camp de la Filosofia però en Intel·ligència Artificial té diferents significats. La definició més consolidada d'ontologia és la realitzada per *Studer i altres autors*<sup>4</sup>, basada en la definició de *Gruber*<sup>5</sup>: "*Una ontologia es una especificación formal y explícita de una conceptualización compartida*".

Els elements d'aquesta definició, també són explicats de la següent forma:

- **Formal:** es refereix a que una ontologia ha de ser llegible per un ordinador excloent el llenguatge natural.
- **Explícita:** significa que els conceptes que s'utilitzen i les seves limitacions es defineixen explícitament.
- **Conceptualització:** es refereix a la identificació dels conceptes més rellevants d'un fenomen en el món.
- **Compartida:** reflecteix la noció que una ontologia captura un consens de coneixement, és a dir, que el coneixement no prové d'un sol individu, sinó que és acceptat per un grup.

Les ontologies estan formades pels següents components que serviran per representar el coneixement d'algun domini:

- **Conceptes:** són les idees bàsiques que s'intenten formalitzar. Els conceptes poden ser classes d'objectes, mètodes, plans, estratègies, processos de raonament, etc.
- **Relacions:** representen la interacció i enllaç entre els conceptes del domini. Solen formar la taxonomia del domini. Per exemple: subclasse-de, part-de, part-exhaustiva-de, connectat-a, etc.
- **Funcions:** són un tipus concret de relació on s'identifica un element mitjançant el càlcul d'una funció que considera diversos elements de l'ontologia. Per exemple, poden aparèixer funcions com categoritzar-classe, assignar-data, etc.

---

<sup>4</sup> Knowledge Engineering: Principles and Methods: <http://www.inf.furb.br/~paulofermando/downloads/EGC/fundamentos%20de%20ec/KnowledgeEngineering-PrinciplesAndMethods.pdf>, consultat 28/10/2017.

<sup>5</sup> A translation Approach to Portable Ontology Specifications: <http://tomgruber.org/writing/ontolingua-kaj-1993.pdf>, consultat 28/10/2017.

- **Instàncies:** s'utilitzen per a representar objectes determinats d'un concepte.
- **Axiomes:** són teoremes que es declaren sobre relacions que han de complir els elements de l'ontologia.

## **2.3. Llenguatges de representació d'ontologies**

Les tecnologies actuals més importants per construir la Web Semàntica son XML, arquitectura RDF i el llenguatge OWL. Totes aquestes tecnologies estan recolzades pel W3C. A continuació, s'analitzaran els llenguatges que s'utilitzen a aquest treball.

### **2.3.1. XML**

El llenguatge XML (eXtensible Markup Language) és un llenguatge de marques derivat de SGML. És un meta-llenguatge que permet definir llenguatges de marques desenvolupades per W3C. S'utilitza per emmagatzemar dades de forma llegible. XML és l'eina utilitzada per estructurar i presentar els continguts web. No obstant això, ofereix una capacitat limitada per expressar semàntica

### **2.3.2. RDF**

RDF (Resource Description Framework) és un llenguatge, desenvolupat pel W3C, capaç de representar metadades, és a dir, dades sobre dades, que ajuden a la identificació, classificació o recuperació de documents en línia. RDF és a la Semàntica el que XML és a la Sintaxi. Gràcies a RDF s'expressen afirmacions i mitjançant la seva llengua de base XML es defineix l'estructura de tals afirmacions. Per tant, XML respon a la necessitat de representar de manera sintàctica el model plantejat per RDF per a arxius legibles per ordinador.

### **2.3.3. OWL**

El Llenguatge d'Ontologies Web (OWL) està dissenyat per a ser usat en aplicacions que necessiten processar el contingut de la informació en lloc d'únicament representar informació per als humans. Està basat en el llenguatge d'ontologies DAM+OIL. OWL pot ser usat per a representar explícitament el significat de termes en vocabularis i les relacions entre aquests termes. OWL té més capacitat per expressar significat i semàntica que XML, i RDF, i, d'aquesta manera, OWL va més enllà d'aquests llenguatges en la seva capacitat per representar contingut interpretable per un ordinador a la web.

### **2.3.4. SPARQL**

SPARQL és un acrònim recursiu de l'anglès SPARQL Protocol and RDF Query Language. Es tracta d'un llenguatge estandarditzat per a la consulta de grafs RDF, normalitzat pel RDF Data Access Working Group (Dawg) del W3C. És una tecnologia clau en el desenvolupament de la web semàntica que es va constituir com a recomanació oficial del W3C el 15 de gener del 2008.

### **2.3.5. OPEN DATA**

Open Data (dades obertes) és una filosofia i pràctica que requereix que certes dades siguin d'accés lliure per a tots, sense limitacions tècniques o legals. Té una ètica similar a altres moviments i

comunitats obertes com el programari lliure (open source, en anglès) i l'accés lliure (open acces, en anglès).

Diversos estudis afirmen que als últims anys ha augmentat l'ús de dades obertes. Cada vegada més organismes públics i empreses privades estan implementant una política de dades obertes perquè els ciutadans els puguin utilitzar i compartir lliurement. Al sector públic, tenir accés a les dades de l'Administració garanteix: transparència, eficiència i la igualtat d'oportunitat; a més de que s'aporta valor. Hi ha repositoris de dades obertes que permeten la consulta i la construcció d'aplicacions que utilitzen la informació alliberada com a font. Alguns repositoris on es poden consultar dades obertes són:

- Administració pública:
  - Govern d'Espanya: <http://datos.gob.es/es>
  - Generalitat de Catalunya: <http://dadesobertes.gencat.cat/es/cataleg-dades-gencat/>
  - Ajuntament de Barcelona: <http://opendata-ajuntament.barcelona.cat/es>
- Economia i Finances:
  - AssetMacro: <https://www.assetmacro.com>
  - Quandl: <https://www.quandl.com>
- Medi Ambient:
  - Sentinel: <https://sentinels.copernicus.eu/web/sentinel/home>
- Música:
  - Million Song Dataset: <https://labrosa.ee.columbia.edu/millionsong/>
- Competicions de dades:
  - Kaggle: <https://www.kaggle.com>

Per aquest treball s'obtindrà la informació dels punts wifi de les ciutats de Buenos Aires i New York del repositori Kaggle. Kaggle és la plataforma líder per a competicions de dades, però també compta amb conjunts de dades sense competicions, les quals es poden obtenir si et dones d'alta a la seva pàgina web.

El Baròmetre de dades obertes<sup>6</sup> (ODB) pretén descobrir la veritable prevalença i l'impacte de les iniciatives de dades obertes a tot el món. Els índexs i baròmetres d'Open Data pretenen reflectir el rànquing internacional dels països en la seva aplicació de polítiques de dades obertes. Encara que és pràcticament impossible mesurar de manera exacta el nivell d'obertura, hi ha algunes iniciatives que mostren una visió general de l'interès en Open Data en els diferents països del món. El passat 1 de juny

---

<sup>6</sup> Open Data Barometer: <http://opendatabarometer.org>, consultat 27/10/2017.

va ser publicat l'últim Baròmetre Open Data, elaborat amb caràcter anual per la World Wide Web Foundation que es realitza des de l'any 2013.

El Baròmetre Open Data avalua les polítiques i pràctiques de dades obertes que estan desenvolupant 115 països d'arreu del món. En la següent figura es pot observar el ranking dels 15 països i el seu nivell d'aplicació d'Open Data:

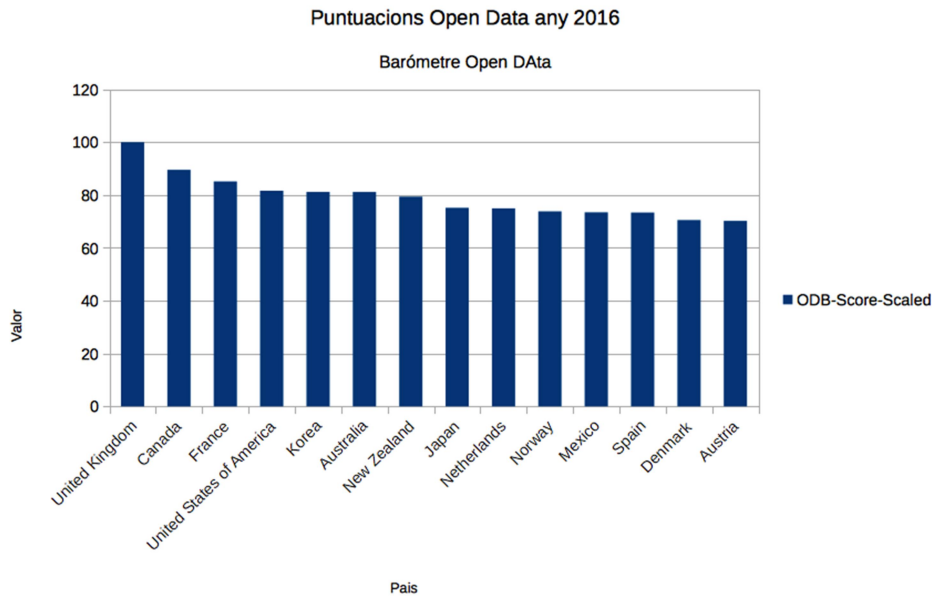


Figura 4: Puntuaciones Open Data. Font Open Data Barometer.

## 3. Disseny de l'ontologia

### 3.1. Metodologia seguida per crear l'ontologia

Per crear una ontologia es poden utilitzar diferents tipus de metodologies. En aquest treball s'ha utilitzat la metodologia **Ontology Development 101** proposada per *Noy & McGuinness* a l'any 2001 a la Universitat de Stanford<sup>7</sup>.

Aquesta metodologia està formada per diferents fases, però s'ha de tenir en compte que:

- No existeix un mode correcte de modelar un domini.
- La millor opció dependrà de l'aplicació concreta. El desenvolupament d'una ontologia és un procés iteratiu.
- Els conceptes en l'ontologia han de ser propers als objectes (físics o lògics) i les relacions en el domini d'interès.

Les fases que componen la metodologia **Ontology Development 101** són:

- Fase 1. Determinar el domini i la cobertura de l'ontologia
- Fase 2. Considerar la reutilització d'ontologies existents.
- Fase 3. Enumerar els termes importants en l'ontologia
- Fase 4. Definir les classes i la seva jerarquia
- Fase 5. Definir les propietats de les classes.
- Fase 6. Definir les característiques de les propietats
- Fase 7. Crear instàncies.

A continuació, es realitza l'estudi de l'ontologia en cadascuna de les fases que componen aquesta metodologia.

#### **Fase 1. Determinar el domini i la cobertura de l'ontologia**

Es vol desenvolupar una ontologia capaç de trobar punts wifi en les ciutats de New York i Buenos Aires.

L'ontologia ha d'incloure els diferents llocs de les ciutats on és possible la connexió a un punt wifi. A més s'haurien de conèixer els detalls dels punts wifi com ara les coordenades de geolocalització del punt, l'adreça, el tipus de connexió, gratuïta o no, la ubicació del punt wifi, per exemple si és un quiosc o un establiment públic, com ara biblioteques.

L'ontologia podria ser utilitzada en altres aplicacions relacionades amb connexions wifi de les ciutats.

---

<sup>7</sup> Ontology Development 101: A Guide to Creating Your First Ontology: <http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.pdf>, consultat 25/10/2017.

## **Fase 2. Considerar la reutilització d'ontologies existents.**

**Swoogle**<sup>8</sup> és un projecte d'investigació del grup "eBiquity" del departament d'Informàtica i Enginyeria Elèctrica, a la Universitat de Maryland. És un programa de codi obert baix llicència Creative Commons. La seva interfície és la següent:



Figura 5: Interfície cercador ontologies Swoogle

Swoogle és un motor de recuperació basat en l'anàlisi de la semàntica. Aquest buscador recupera, processa, analitza i indexa documents que estan disponibles online. La seva interfície pot resultar familiar ja que és similar a la del buscador *Google*.

Per a buscar ontologies relacionades amb l'ontologia que es realitza a aquest treball, s'ha cercat amb Swoogle la paraula "wifi". El resultat ha sigut de diverses coincidències. Concretament es troba una ontologia amb aquest terme, ja que els altres enllaços no estan actius al moment de la cerca<sup>9</sup>. Aquesta ontologia que s'ha trobat no es tracta d'una ontologia específica, per tant, es procedirà a crear una nova ontologia des de 0.

## **Fase 3. Enumerar els termes importants en l'ontologia**

A l'ontologia es parlarà sobre:

- Punts wifi, tipus dels punts wifi, ubicació dels punts wifi, llocs on es troben aquests punts.
- Ciutats, zones de la ciutat, coordenades de geolocalització (latitud i longitud), adreça.

---

<sup>8</sup> Swoogle: <http://swoogle.umbc.edu/2006/>, consultat 1/11/2017.

<sup>9</sup> Consultat 1/11/2017.

Una vegada creada l'ontologia les consultes que es realitzarien respondrien les següents preguntes:

- Quina és la ciutat on es troba el punt wifi?
- Quants punts wifi hi ha en una ciutat?
- Quin és el tipus del punt wifi?
- Quins punts wifi són gratuïts?
- Quines són les coordenades d'un punt wifi concret?

#### **Fase 4. Definir les classes i la seva jerarquia**

Per definir la jerarquia de les classes d'una ontologia es pot prendre diferents aproximacions:

- **De dalt a baix:** Es defineixen els conceptes més generals i es van especialitzant.
- **De baix a dalt:** Es defineixen les classes més específiques i s'agrupen segons propietats comunes.
- **Combinació d'ambdues:** Es defineixen els conceptes més importants i s'especialitza i generalitza per completar l'ontologia.

Cap d'aquests mètodes és essencialment millor i dependrà en gran mesura del domini de l'ontologia.

Per a aquest treball s'ha escollit un procés de *dalt a baix*. Per tant la jerarquia quedaria de la següent forma:

- Punt wifi
- Localització
  - Ciutat
  - Zona
  - Adreça
- Lloc
- Coordenades

La jerarquia de les classes de l'ontologia que s'està estudiant seria la següent:

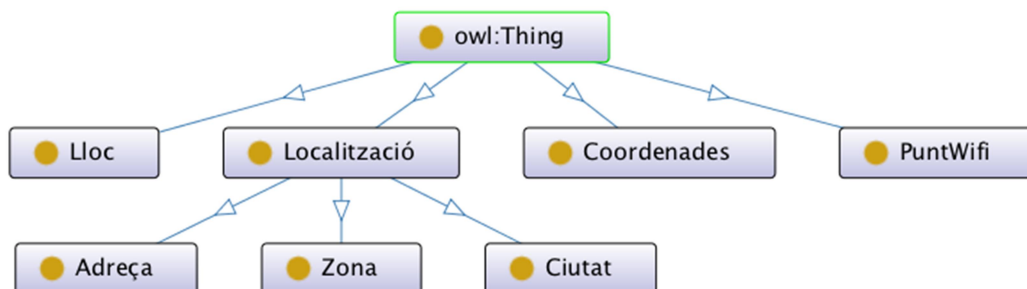


Figura 6: Jerarquia de classes de l'ontologia

La imatge de la figura 6, s'ha realitzat amb Ontograf<sup>10</sup>. Ontograf és un plugin del programa Protégé .

La classe *Thing*<sup>11</sup> és una classe de tots els individus i és una superclasse de totes les classes del llenguatge OWL. Les classes de l'ontologia, es crearan a partir d'aquesta classe.

### **Fase 5. Definir les propietats de les classes.**

A aquesta fase s'ha de descriure l'estructura interna de les classes. Es realitzarà una llista de les característiques i en quines classes estaran. Es poden tenir diferents tipus de propietats:

- **Propietats descriptives:** normalment són qualitats de les classes.
- **Propietats identificadores:** solen ser noms que identifiquen a les classes.
- **Relacions** amb instàncies a altra classe

A l'ontologia que s'està creant es tindrien les següents propietats:

- Propietats identificadores: identificador del punt wifi, nom de la ciutat, nom de la zona, nom del lloc, coordenades geolocalització i nom de l'adreça.
- Propietats descriptives: codi postal de l'adreça, estat del punt wifi, nom del punt wifi, tipus de punt wifi.

Relacions:

- Una coordenada de geolocalització correspon a un punt wifi.
- Un punt wifi disposa d'una coordenada de geolocalització.
- Un punt wifi està a una localització.
- Una localització té diferents punts wifi.
- Un punt wifi està a un lloc concret.
- Un lloc ofereix punts wifi.
- Una ciutat està formada per zones.
- Una zona pertany a una ciutat.
- Una adreça està situada a una ciutat.
- Una ciutat posseeix diferents adreces.

---

<sup>10</sup> Ontograf: <https://protegewiki.stanford.edu/wiki/OntoGraf> consultat 10/12/2017.

<sup>11</sup> Lenguaje de Ontologías Web (OWL): <https://www.w3.org/2007/09/OWL-Overview-es.html#ClassEs>, consultat 27/10/2017.



### **Fase 6. Definir les característiques de les propietats**

A aquesta fase s'han de definir les següents característiques:

- Cardinalitat
- Tipus dels valors
- Valors per defecte
- Obligatorietat
- A les relacions definir el rang

Per tant, a l'ontologia d'aquest treball es tenen les següents característiques:

- Tipus dels valors. Els valors de les propietats d'aquesta ontologia són de tipus string i sencer. En concret:
  - Valors tipus *string*: estat del punt wifi, coordenades geolocalització (és un valor format per dos nombres que indiquen la longitud i latitud), nom de l'adreça, nom de la ciutat, nom del lloc, nom del punt, nom de la zona, tipus del punt wifi.
  - Valors tipus *integer*: codi postal de l'adreça, identificador del punt wifi.
- Respecte a les relacions:
  - Un punt wifi sols té una localització.
  - Una ciutat pot tenir diferents punts wifi.
  - Una ciutat té diferents zones.
  - Una zona sols pertany a una ciutat.

### **Fase 7. Crear instàncies.**

Les instàncies que formaran part d'aquesta ontologia es crearan mitjançant la importació de fitxers amb format csv que s'han obtingut del repositori *Kaggle*.

## **3.2. Eines per a la creació d'ontologies**

Una vegada fet el disseny de l'ontologia és possible fer el seu desenvolupament amb diferents eines i formats. Les eines permeten la codificació d'una determinada ontologia basant-se en un determinat llenguatge. A aquest treball, s'ha fet un estudi de les eines que es poden utilitzar per al desenvolupament d'ontologies.

Entre les solucions estudiades es pot destacar:

- Protégé
- WebProtégé
- Llibreries Apache Jena

### 3.2.1. Protégé

Protégé és una eina per al desenvolupament d'ontologies i sistemes basats en el coneixement creada a la Universitat de Stanford. Protégé és una eina de codi obert i multiplataforma.

Protégé és un entorn amb suport complet per a l'OWL 2 Web Ontology Language i connexions directes en memòria a raonadores de lògica de descripció com HermiT i Pellet.

Les aplicacions desenvolupades amb Protégé són emprades en resolució de problemes i decisions en dominis particulars.

Protégé admet la creació i edició d'una o més ontologies en un únic espai de treball a través d'una interfície d'usuari completament personalitzable. Les eines de visualització permeten la navegació interactiva de les relacions ontològiques. Algunes operacions que permet realitzar Protégé són: canvi de nom d'entitats, importació d'ontologies, els axiomes es poden exportar a altres ontologies, a més de moltes altres opcions de redisseny.

### 3.2.2. WebProtégé

WebProtégé és un entorn gratuït de desenvolupament d'ontologies de codi obert per a la Web. WebProtégé està desenvolupant activament l'equip Protégé al Stanford Center. WebProtégé és lliure i de codi obert sota llicència BSD 2-clause. WebProtégé proporciona les següents funcions:

- Suport per editar ontologies OWL 2.
- Una interfície d'edició simple per defecte, que proporciona accés a construccions OWL d'ús comú.
- Seguiment de canvis complet i historial de revisions.
- Eines de col·laboració, com ara compartir, permisos, notes i discussions, rellotges i notificacions per correu electrònic.
- Interfície d'usuari personalitzable.
- Formes web personalitzables per a l'edició específica de l'aplicació/domini.
- Suport per editar ontologies OBO.
- Múltiples formats per carregar i descarregar ontologies (formats compatibles: RDF/XML, Tortuga, OWL/XML, OBO i altres).

### 3.2.3. Apache Jena

Apache Jena és un *framework* de Java lliure i de codi obert per construir aplicacions web semàntiques i Linked Data. Jena va ser desenvolupada originalment per investigadors en HP Labs, començant a Bristol, Regne Unit, l'any 2000. Jena sempre ha estat un projecte de codi obert i s'ha utilitzat àmpliament en una àmplia varietat d'aplicacions web i demostradors semàntiques. En 2009 HP va cedir el projecte a la fundació Apache que va decidir adoptar-lo en 2010.

Jena proporciona una àmplia biblioteca de Java per ajudar els desenvolupadors a desenvolupar codi que maneja RDF, RDFS, RDFa, OWL i SPARQL d'acord amb les recomanacions publicades per W3C.

La seva arquitectura inclou:

- API per treballar (llegir, processar, escriure) ontologies RDF i OWL
- Motor d'inferència per raonar sobre ontologies RDF i OWL
- Estratègies d'emmagatzematge flexible per emmagatzemar tripletes RDF en memòria o fitxer
- Motor de queries compatible amb especificació SPARQL

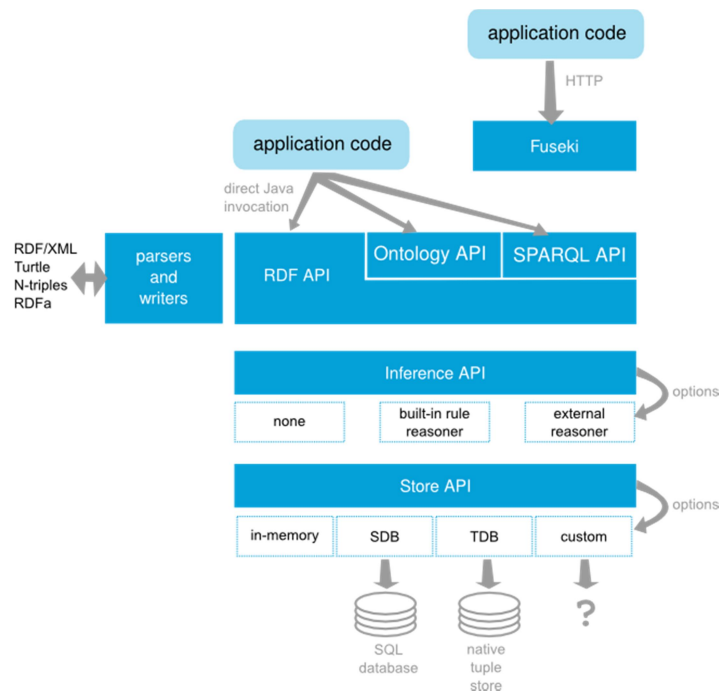


Figura 7: Arquitectura Framework Apache Jena

Per fer ús d'aquestes llibreries es pot utilitzar l'IDE Eclipse. Eclipse permet afegir llibreries externes. Tant per poder utilitzar Eclipse com Jena és necessari instal·lar Java JDK, en el moment de la creació del treball la versió és la 1.8.x.

### 3.2.4. Elecció de l'eina per al desenvolupament de l'ontologia

Una vegada estudiades les eines per al desenvolupament de l'ontologia, s'ha d'escollir quina s'utilitzarà al treball.

La llibreria Apache Jena s'utilitza principalment per treballar amb grafs RDF. Jena és capaç de treballar amb ontologies OWL perquè el llenguatge RDF pot ser utilitzat com a sintaxi per representar OWL.

Protégé utilitza l'API OWL. Aquesta API es centra amb el llenguatge OWL. Usant l'API OWL, es tracta una ontologia com a conjunt d'axiomes en lloc de com un conjunt de triples que passen a codificar algun axioma.

Cada llenguatge té les seves característiques específiques. OWL té més constructors que RDF, però RDF és més adequat per a treballar amb el motor de consulta SPARQL.

Per a la realització d'aquest treball es considera una millor opció escollir **Protégé** com a eina per al desenvolupament de l'ontologia, ja que treballa amb l'API OWL i permet treballar directament sobre les estructures i conceptes que són rellevants amb aquest llenguatge.

Per a la càrrega dels arxius CSV amb les dades dels punts wifi de les ciutats s'utilitzarà l'eina **WebProtégé**, ja que permet diferents formes de càrrega i descàrrega d'arxius.

## 4. Implementació de l'ontologia

### 4.1. Instal·lació de les eines necessàries

#### 4.1.1. Protégé

Aquest treball s'ha realitzat sobre un sistema operatiu Windows 10. La versió de Protégé que s'ha decidit instal·lar és la 4.3. Actualment Protégé està per la versió 5, però hi ha problemes de compatibilitat entre Protégé 5 i Windows 10, per tant s'ha optat per instal·lar **Protégé 4.3**.

#### Requisits

Per a la instal·lació de Protégé s'ha de tenir al sistema la màquina virtual de Java. Si no es té instal·lada la màquina, la pàgina web de Protégé ofereix la possibilitat de descàrrega del programari amb la màquina virtual de Java.

#### Plugins

Protégé és un framework al qual se li poden instal·lar plugins resultat d'altres projectes. Si es vol consultar qualsevol plugin i la versió de Protégé que suporta es pot fer a la wiki de Protégé<sup>12</sup>. Per a la realització de l'ontologia s'han instal·lat els següents plugins:

- OntogGraf<sup>13</sup>

OntoGraf dona suport per a la navegació interactiva de les relacions de les ontologies OWL. Es permeten diversos dissenys per organitzar automàticament l'estructura de l'ontologia. S'admeten diferents relacions: subclasse, propietats d'objectes individuals, domini / rang i equivalència. Les relacions i els tipus de nodes es poden filtrar per ajudar-vos a crear la visualització que desitgeu.

#### Raonador semàntic

Un raonador semàntic, és un programari capaç d'inferir conseqüències lògiques d'un conjunt de fets o axiomes assertats. Les regles d'inferència normalment s'especifiquen per mitjà d'un llenguatge d'ontologia i, sovint, un llenguatge lògic de descripció.

El raonador que s'ha utilitzat a aquesta ontologia és:

- Hermit<sup>14</sup>: Hermit és un raonador d'ontologies escrit usant el Llenguatge d'Ontologia Web (OWL). Donat un fitxer OWL, Hermit pot determinar si l'ontologia és o no coherent, identificar relacions de subsunció entre classes i molt més.

---

<sup>12</sup> Protégé Wiki, [https://protegewiki.stanford.edu/wiki/Protege\\_Plugin\\_Library](https://protegewiki.stanford.edu/wiki/Protege_Plugin_Library), consultat: 04/12/2017.

<sup>13</sup> Protégé OntoGraf Wiki, <https://protegewiki.stanford.edu/wiki/OntoGraf>, consultat: 06/12/2017.

<sup>14</sup> Hermit, <http://www.hermit-reasoner.com/>, consultat 20/12/2017.

### 4.1.2. WebProtégé

WebProtégé és una aplicació web que s'executa en un contenidor de servlets, com ara Tomcat. El fitxer d'instal·lació de WebProtégé és un fitxer de tipus *war* que es pot implementar fàcilment en un contenidor de servlets.

#### Requisits

Per a la instal·lació de WebProtégé es necessita:

- Tomcat (o algun altre contenidor de servlets).
- MongoDB: s'utilitza per emmagatzemar diferents dades de configuració de WebProtégé.
- Arxiu *webprotege.war*.
- I un directori de dades on WebProtégé emmagatzemarà totes les seves dades.

### 4.1.3. Recursos

Per a realitzar l'ontologia d'aquest treball s'han descarregat els fitxers d'un repositori Open Data anomenat Kaggle.

Les dades dels punts d'accés wifi de New York es poden trobar a:

<https://www.kaggle.com/new-york-city/nyc-public-wifi/data>

Les dades dels punts d'accés wifi de New York es poden trobar a:

<https://www.kaggle.com/new-york-city/nyc-public-wifi/data>

## 4.2. Creació de l'ontologia

Al punt 3 s'ha descrit la metodologia seguida per a crear l'ontologia i també s'han definit les classes, propietats i relacions que ha de tenir. Per realitzar aquest estudi s'han tingut en compte els fitxers de dades csv dels punts wifi de les ciutats de New York i Buenos Aires que s'han descarregat del repositori Open Data Kaggle.

Segons s'ha decidit al punt 3.2.4 d'aquest treball, per a la creació de l'ontologia s'utilitza l'eina de creació d'ontologies Protégé.

La interfície de Protégé es mostra a la figura 8. La seva interfície esta formada per una col·lecció de quadres que l'usuari pot administrar segons li convinga.

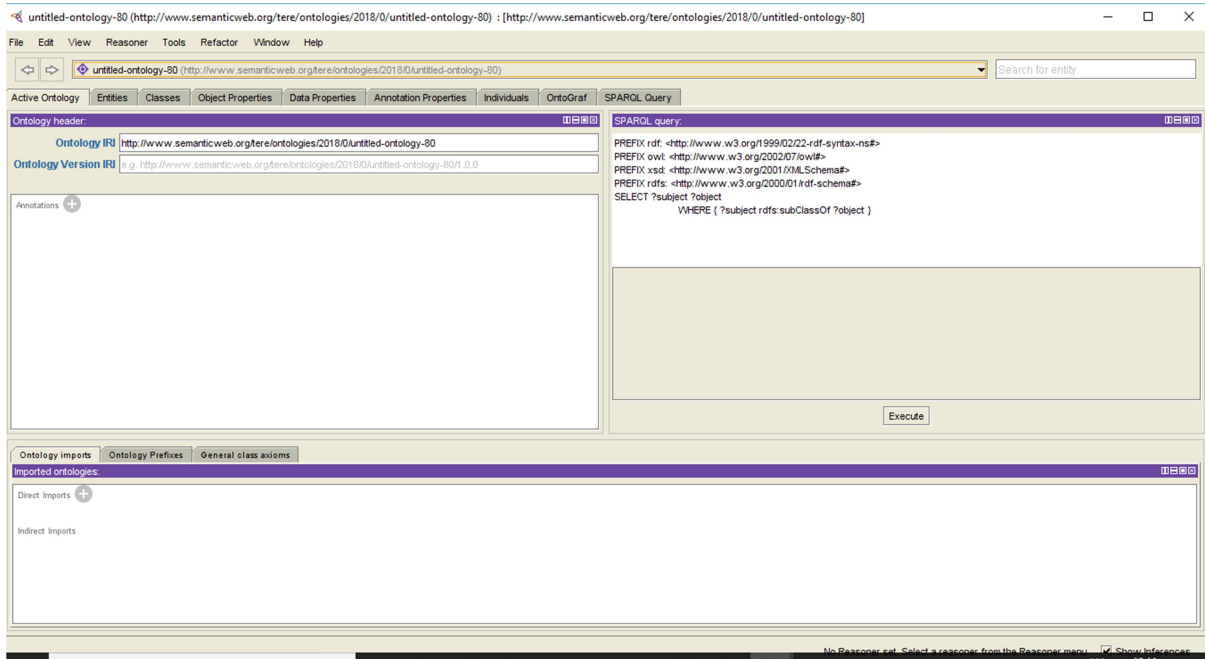


Figura 8: Interfície Protégé

Per poder elegir l'acció que es vol realitzar en l'ontologia es pot fer mitjançant la barra de pestanyes que proporciona Protégé:



Figura 9: Barra de navegació de Protégé

### 4.2.1. Classes

Són sets que contenen individus. Es tracta de representacions concretes de conceptes, que es descriuen mitjançant declaracions que indiquen els requeriments per ser membre d'aquestes. Les classes que es creïn han de ser subclasses de la classe Thing, que al programa Protégé ja està creada.

#### Classe PuntWifi

| Camps classe PuntWifi |         |
|-----------------------|---------|
| Camp                  | Tipus   |
| id                    | integer |
| estat                 | string  |
| nomPunt               | string  |
| tipus                 | string  |

Taula 3: Camps classe PuntWifi

#### Classe Adreça

| Camps classe Adreça |         |
|---------------------|---------|
| Camp                | Tipus   |
| nomAdreça           | integer |

Taula 4: Camps classe Adreça

### Classe Coordenades

| Camps classe Coordenades |         |
|--------------------------|---------|
| Camp                     | Tipus   |
| idCoordenada             | integer |
| Geolocalització          | string  |

Taula 5: Camps classe Coordenades

### Classe Ciutat

| Camps classe Ciutat |         |
|---------------------|---------|
| Camp                | Tipus   |
| nomCiutat           | integer |

Taula 6: Camps classe Ciutat

### Classe Zona

| Camps classe Zona |         |
|-------------------|---------|
| Camp              | Tipus   |
| nomZona           | integer |

Taula 7: Camps classe Zona

### Classe Lloc

| Camps classe Zona |         |
|-------------------|---------|
| Camp              | Tipus   |
| nomLloc           | integer |

Taula 8: Camps classe Zona

El resultat de la creació de les classes és el següent:

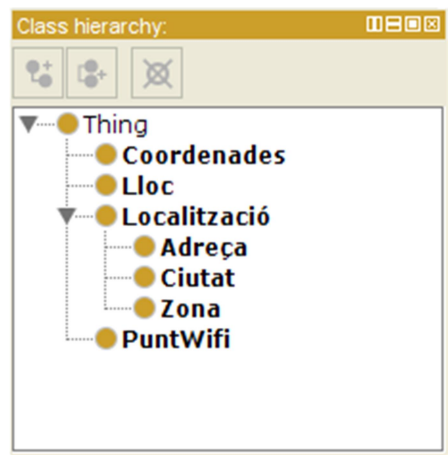


Figura 10: Classes creades amb Protégé

Les classes Adreça, Ciutat i Zona són subclasses de Localització, per tant, s'ha d'indicar aquesta característica en l'ontologia.



## 4.2.2. Propietats

Les propietats indiquen relacions entre individus o entre un individu i una dada, i es detallen habitualment a nivell de classe. Es poden crear subpropietats i existeixen tant per a un conjunt d'individus (és a dir, propietat o atribut de classe) com per a individus concrets (propietat o atribut d'instància).

Una propietat té un domini (domain) i un rang (range). En el domini és el conjunt d'individus d'una classe que es relaciona amb el conjunt d'individus d'una altra classe, el rang. Hi ha tres tipus de propietats:

- **Propietats d'anotació (annotation properties):** utilitzades per afegir informació a classes, individus i propietats d'objecte o dada. Entre aquestes propietats, destaquen els labels i els comentaris.
- **Propietats tipus dada (datatype restrictions):** vinculen individus i valors (dades). Cada propietat està assignada a una classe i té un rang que correspon al tipus de dades ha de tenir.

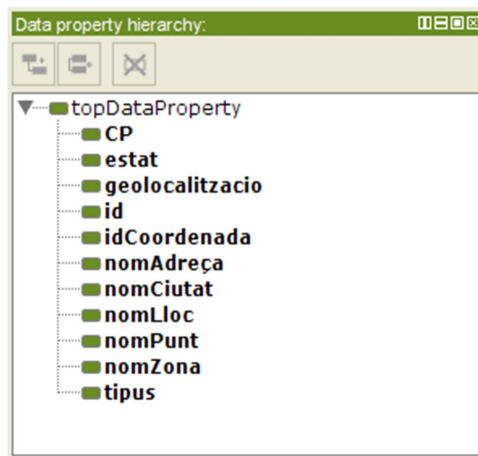


Figura 11: Data Properties amb Protégé

Per exemple, la propietat `id` té com a domini la classe `PuntWifi` i el rang és `integer`.

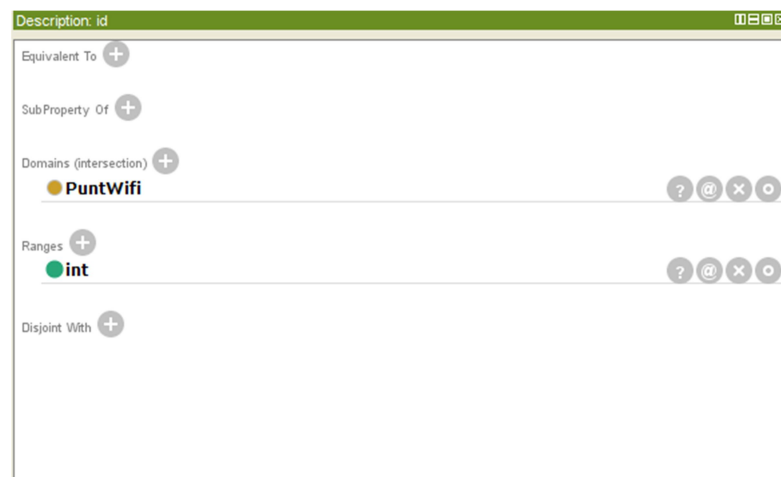


Figura 12: Descripció de la propietat `id`

- **Propietats tipus objecte (object type restrictions):** Les propietats de tipus objecte són relacions binàries entre individus.

Les propietats que s'han creat en l'ontologia es mostren en la següent taula.

| Propietats tipus objecte |              |              |             |
|--------------------------|--------------|--------------|-------------|
| ObjectProperty           | Domini       | Rang         | Inversa     |
| corresponA               | Coordenades  | PuntWifi     | disposa     |
| disposa                  | PuntWifi     | Coordenades  | corresponA  |
| esta                     | PuntWifi     | Localització | té          |
| estaA                    | PuntWifi     | Lloc         | ofereix     |
| estaFormada              | Ciutat       | Zona         | pertanyA    |
| estaSituada              | Adreça       | Ciutat       | poseeix     |
| ofereix                  | Lloc         | PuntWifi     | estaA       |
| pertanyA                 | Zona         | Ciutat       | estaFormada |
| poseeix                  | Ciutat       | Adreça       | estaSituada |
| té                       | Localització | PuntWifi     | esta        |

Taula 9: Propietats tipus Object Property

El resultat de la creació de les propietats de tipus objecte amb el programari Protégé és el següent:

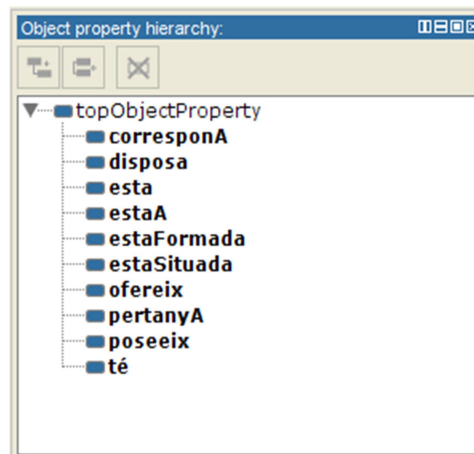


Figura 13: Object Properties amb Protégé

Aquestes propietats poden tenir diferents característiques. En concret, aquests objectes tenen la propietat inversa: si una propietat vincula A i B, la seva propietat inversa relacionarà B i A. Per exemple:

- Una ciutat *estaFormada* per zones i una zona *pertanyA* una ciutat.
- La relació *estaFormada* té com inversa *pertanyA*. En aquest cas domini i rang *estaFormada* seran rang i domini, respectivament, de *pertanyA*.

A continuació, es mostra un exemple amb Protégé:



Figura 14: Descripció de la relació estaFormada

Una vegada creades les classes i les propietats podem consultar els diferents nodes amb el plugin Ontograf de Protégé. El resultat seria el següent:

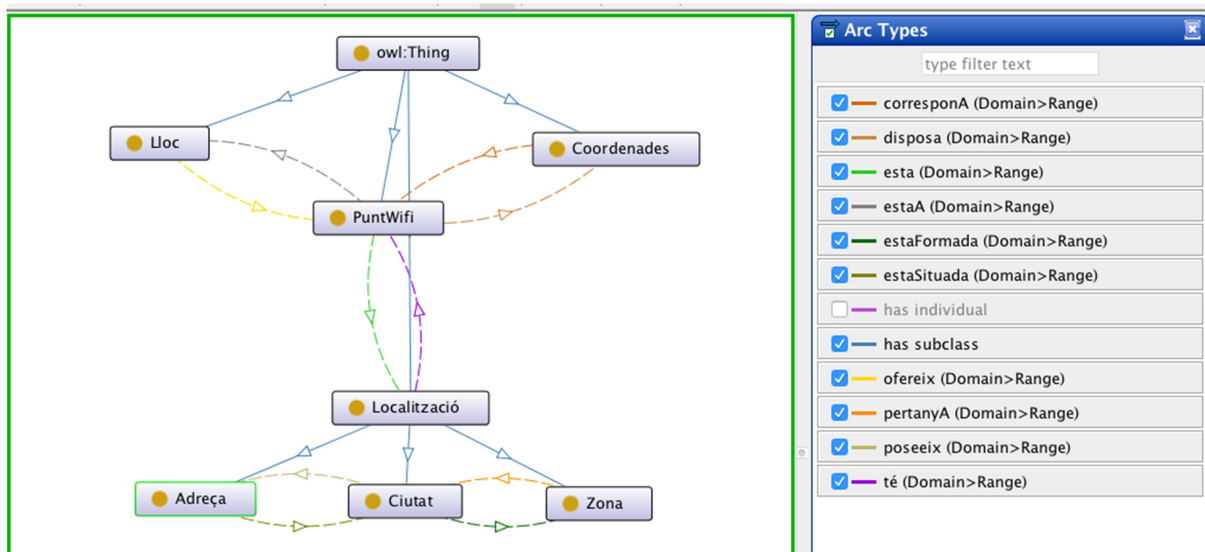


Figura 15: Diagrama ontologia

### 4.2.3. Instàncies

Les instàncies representen objectes del món real, en el domini en què estem treballant. Habitualment estan vinculats a una classe mitjançant una relació.

Per a crear les instàncies s'utilitza WebProtégé que ofereix la possibilitat d'importar arxius CSV. L'ontologia s'ha creat amb Protégé, per tant, ara aquesta ontologia s'ha d'importar a WebProtégé.

Quan s'importa l'ontologia, l'aparença és la següent:

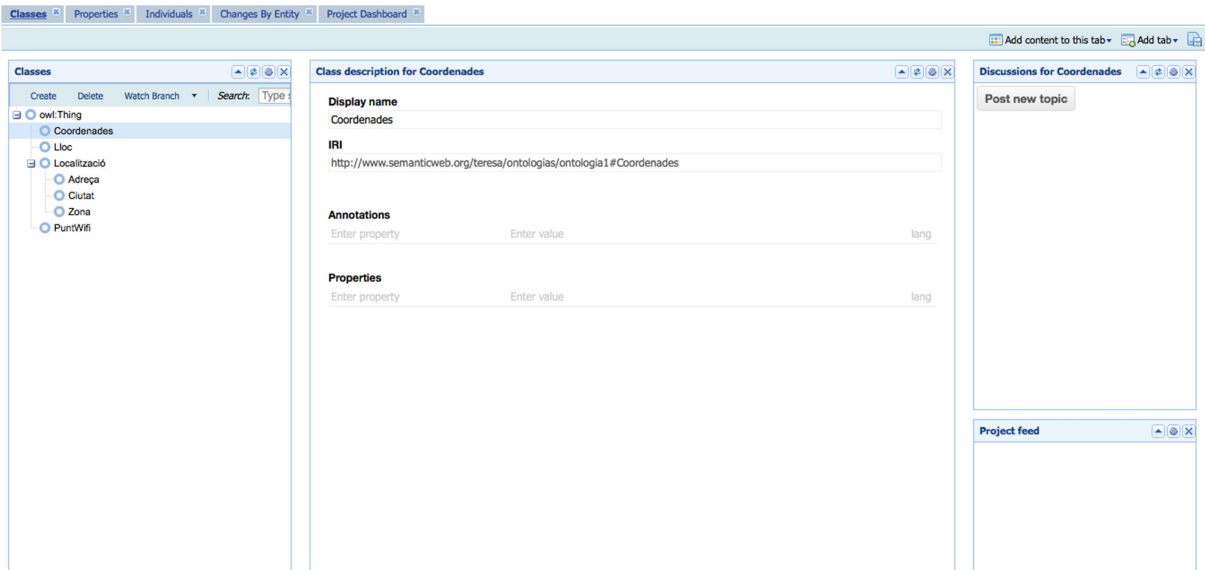


Figura 16: Pantalla Principal Ontologia WebProtégé

Es pot observar que estan les classes i subclasses que s'han creat en Protégé. Una vegada importada, ja es poden crear les instàncies.

Per crear les instàncies s'ha de seleccionar la classe, la qual es volen importar les instàncies, **pulsar el botó *shift* + seleccionar *Create amb el botó esquerra del ratolí***

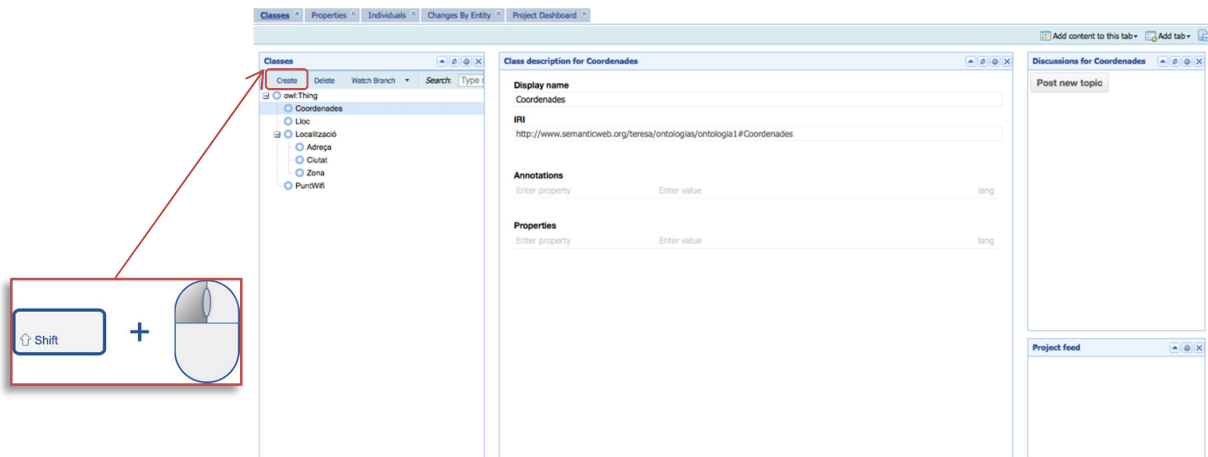


Figura 17: Creació instàncies de les classes

Sortirà una finestra on s'ha de seleccionar el fitxer de dades CSV que s'ha d'importar i es selecciona el fitxer que correspongui<sup>15</sup>.



Figura 18: Càrrega d'arxius CSV amb WebProtégé

Una vegada seleccionat el fitxer ens mostra la següent pantalla:

### Import CSV File

Import rows as:

Display name column:

Property value mapping

| Column   | Property name and column type            |
|----------|--|
| Column 1 | Enter property name <input type="text"/> |
| Column 2 | Enter property name <input type="text"/> |
| Column 3 | Enter property name <input type="text"/> |
| Column 4 | Enter property name <input type="text"/> |
| Column 5 | Enter property name <input type="text"/> |
| Column 6 | Enter property name <input type="text"/> |
| Column 7 | Enter property name <input type="text"/> |

| Column 1 | Column 2 | Column 3 | Column 4             |
|----------|----------|----------|----------------------|
| OBJECTID | Borough  | Type     | Provider             |
| 998      | MN       | Free     | LinkNYC - Citybridge |
| 999      | MN       | Free     | LinkNYC - Citybridge |
| 1000     | MN       | Free     | LinkNYC - Citybridge |

OK Cancel

Figura 19: Importació de dades amb WebProtégé

<sup>15</sup> A partir d'aquest moment tots els exemples del treball es veuran amb el fitxer de dades dels punts WIFI de New York.

Les dades del CSV es poden importar en forma de classes o de instàncies (també anomenats individuals). Per a aquest treball, s'han importat les dades corresponents a cada classe en forma d'instàncies.

Quan es vol crear una nova instància s'ha de seleccionar la columna en la qual està la dada que interressi emmagatzemada i el nom de la propietat corresponent amb l'ontologia. A més d'indicar la propietat, també s'ha d'indicar el rang.

Per exemple, si es vol importar les instàncies corresponents a la classe PuntWifi d'aquesta ontologia:

### Import CSV File

Import rows as: Individual ← 1

Display name column: Column 1 ← 2

Property value mapping

| Column   | Property name and column type |
|----------|-------------------------------|
| Column 1 | id Integer ← 3.1              |
| Column 2 | Enter property name           |
| Column 3 | tipus String ← 3.2            |
| Column 4 | Enter property name           |
| Column 5 | Enter property name           |
| Column 6 | Enter property name           |
| Column 7 | Enter property name           |

↑ 3

| id       | Column 2 | tipus | Column 4             |
|----------|----------|-------|----------------------|
| OBJECTID | Borough  | Type  | Provider             |
| 998      | MN       | Free  | LinkNYC - Citybridge |
| 999      | MN       | Free  | LinkNYC - Citybridge |
| 1000     | MN       | Free  | LinkNYC - Citybridge |

OK Cancel

Figura 20: Exemple creació instàncies.

Els passos per a la creació de les instàncies són els següents<sup>16</sup>:

1. Primer s'indica que cada fila que s'importi serà una instància (*Import row as*). Es representa amb el número 1 a la figura 20.
2. S'ha de seleccionar per quina columna es vol anomenar cada instància (*Display Name Column*), en aquest cas la columna 1. Es representa amb el número 2 a la figura 20.

<sup>16</sup> Els números que apareixen a la figura 20 corresponen amb els passos que s'indiquen a aquest llistat.

3. S'assigna cada propietat amb el valor de la columna CSV. Es representa amb el número 3 a la figura 20. Per assignar cadascuna de les propietats es realitza de la següent forma:
  - 3.1. La primera columna del fitxer dels punts wifi de New York és l'identificador del punt wifi, doncs la propietat id de la classe PuntWifi tindrà aquest valor i és de tipus *integer*. Es representa amb el número 3.1 a la figura 20.
  - 3.2. La columna 3 del fitxer indica el tipus de punt wifi, per tant se li assigna la propietat tipus i el rang, en aquest cas *string*. Es representa amb el número 3.2 a la figura 20.

Aquest procés s'ha de realitzar per cadascuna de les classes de l'ontologia.

Una vegada creades totes les instàncies, es pot exportar de nou l'ontologia per seguir treballant amb Protégé o es pot continuar treballant amb WebProtégé. Com WebProtégé és una eina relativament nova i encara està en fase d'experimentació s'ha optat per seguir fent el treball d'aquesta ontologia amb Protégé.

Si es realitza la importació de l'ontologia a Protégé, es pot comprovar que estan les instàncies creades de cadascuna de les classes.

Per exemple, la classe zona té les següents instàncies:

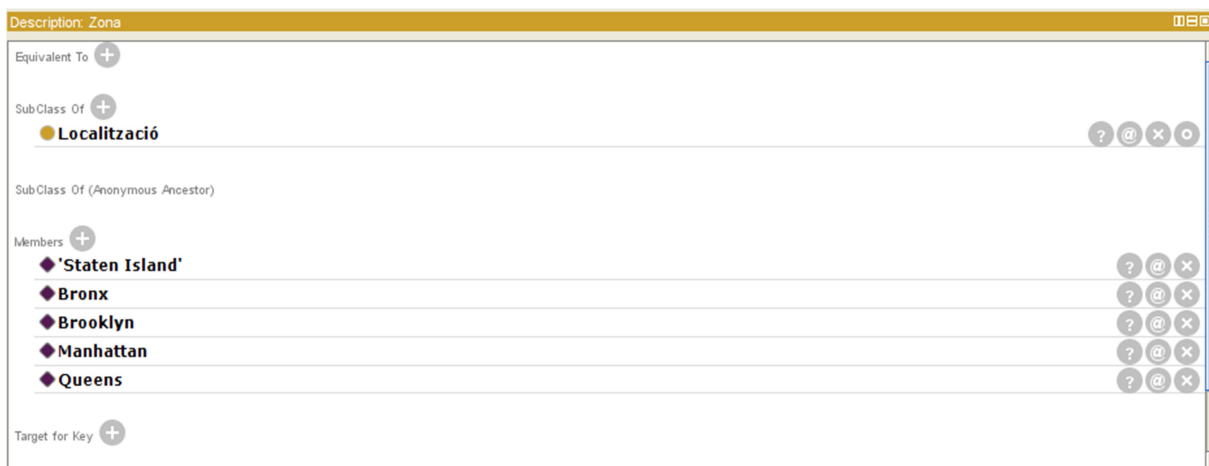


Figura 21: Exemple instàncies de la classe Zona

Cadascuna de les instàncies té les seves propietats de tipus objecte i data, i a més, té assignades les classes corresponents.

Per exemple, la instància anomenada "World Trade Center" és una instància de la classe Adreça i té els valors que es poden veure a la figura 21.

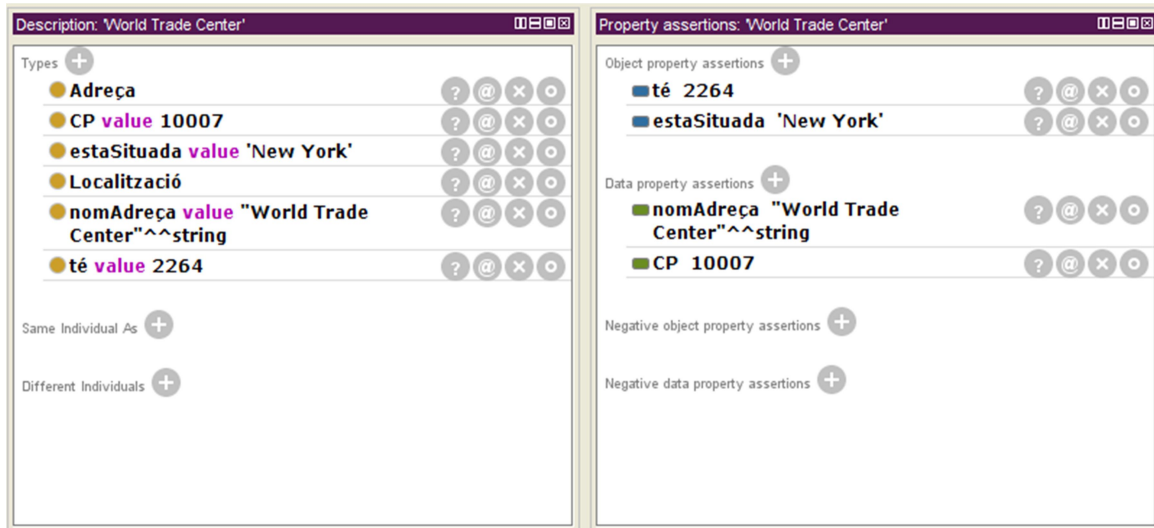


Figura 22: Exemple instància adreça "World Trade Center"

En importar l'ontologia a Protégé, potser no apareguin els Object Properties corresponents amb la importació de dades que s'ha fet a Webprotégé. Açò és un bug de Protégé respecte a formats del llenguatge OWL que es produeix quan es fan importacions d'ontologies a Protégé. La solució perquè es mostrin totes les propietats està en fer ús del Raonador (*Reasoner*) per comprovar que l'ontologia efectivament està ben creada.

Al menú de la part superior de Protégé hi ha una opció anomenada "**Reasoner**". S'ha de seleccionar el Raonador perquè comprovi l'ontologia. Per defecte, Protégé porta incorporat el raonador anomenat Hermit.

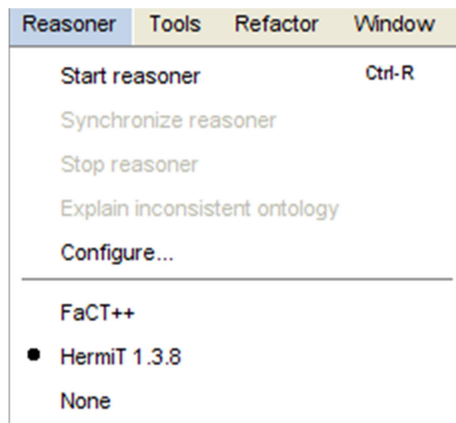


Figura 23: Activar Raonador Hermit

Per a comprovar les inferències realitzades pel raonador, sha de tenir activada l'opció "Show Inference" que està a la part inferior de la interfícies de Protégé.

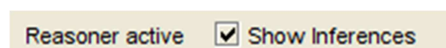


Figura 24: Show Inference



A l'activar aquest raonador es pot comprovar que l'ontologia creada és correcta. Les propietats de tipus objecte que no hi eren abans d'aplicar el raonador ara si apareixen.

Per desar els axiomes que ha creat el raonador s'ha de seleccionar del menú "**File**" l'opció "**Export inferred axioms as ontology**".

### 4.3. CONSULTES SPARQL

Un cop poblada l'ontologia i comprovat que no hi ha cap inconsistència és hora d'extreure el coneixement representat i comprovar que l'ontologia dona resposta a algunes qüestions que s'havien plantejat al punt 3 d'aquest treball.

Per extreure aquesta informació s'utilitza el llenguatge de consultes SPARQL. SPARQL està present a l'eina Protégé a través de la pestanya anomenada "*SPARQL Query*".

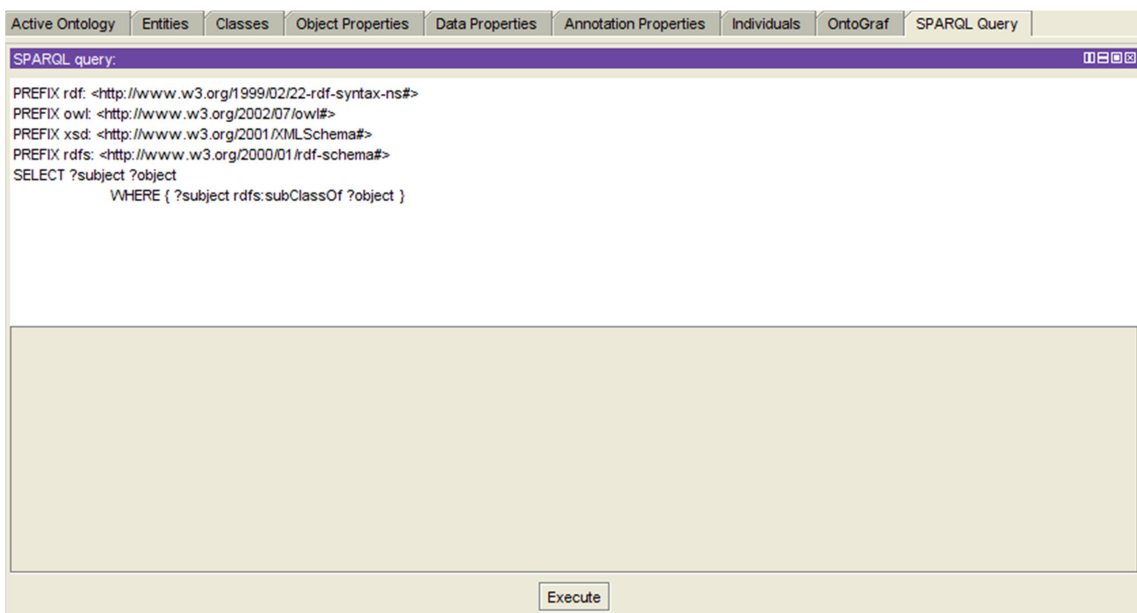


Figura 25: Pestanya SPARQL Query en Protégé 4.3

### 4.3.1. Consultes a l'ontologia

A continuació es mostren una sèrie de consultes que s'han realitzat sobre l'ontologia creada.

- **Mostra les ciutats i les seves zones:**

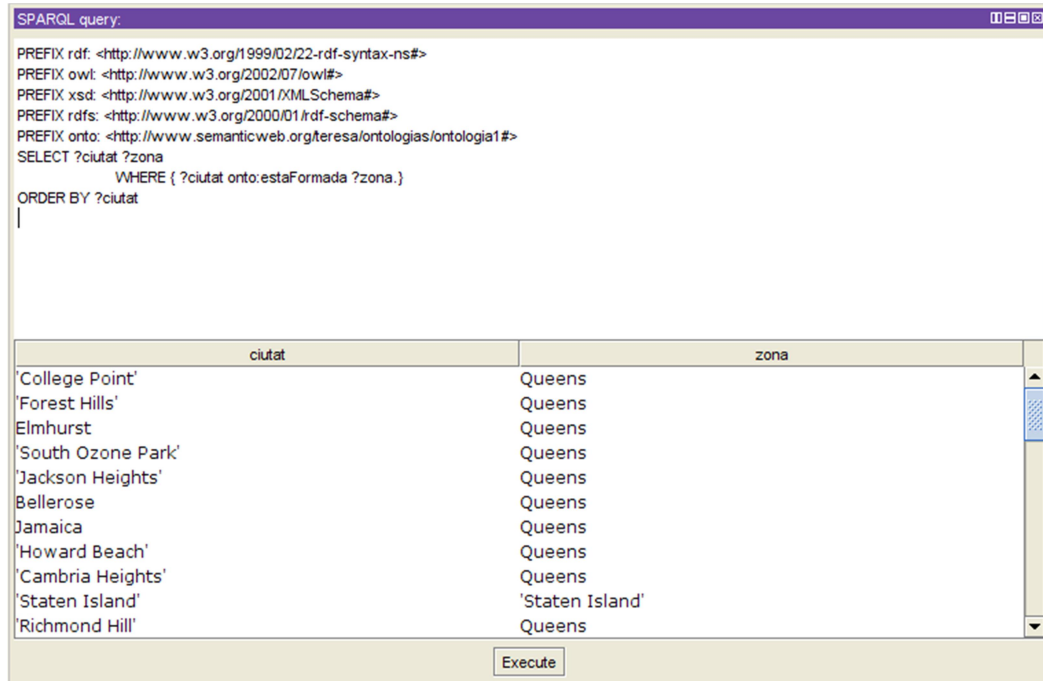


Figura 26: Resultat consulta "Mostra les ciutats i les seves zones"

Es poden utilitzar modificadors<sup>17</sup> de seqüència de solució per poder tractar les dades de la solució, ja que es genera una solució desordenada.

Un modificador de seqüència de solució pot ser algun dels següents:

- **Modificador d'ordre (order by):** ordenar les solucions
- **Modificador de projecció:** escollir certes variables
- **Modificador distinct:** garantir que les solucions a la seqüència són úniques
- **Modificador Reduced:** permetre l'eliminació d'algunes solucions no úniques
- **Modificador Offset:** controlar on comencen les solucions en la seqüència global de solucions
- **Modificador Limit:** restringir el nombre de solucions.

<sup>17</sup> SPARQL Lenguaje de consulta para RDF : <http://skos.um.es/TR/rdf-sparql-query/#solutionModifiers>, consultat 26/12/2017.

A continuació es mostra un exemple de la mateixa solució d'abans però usant els modificadors `offset` i `límit`:

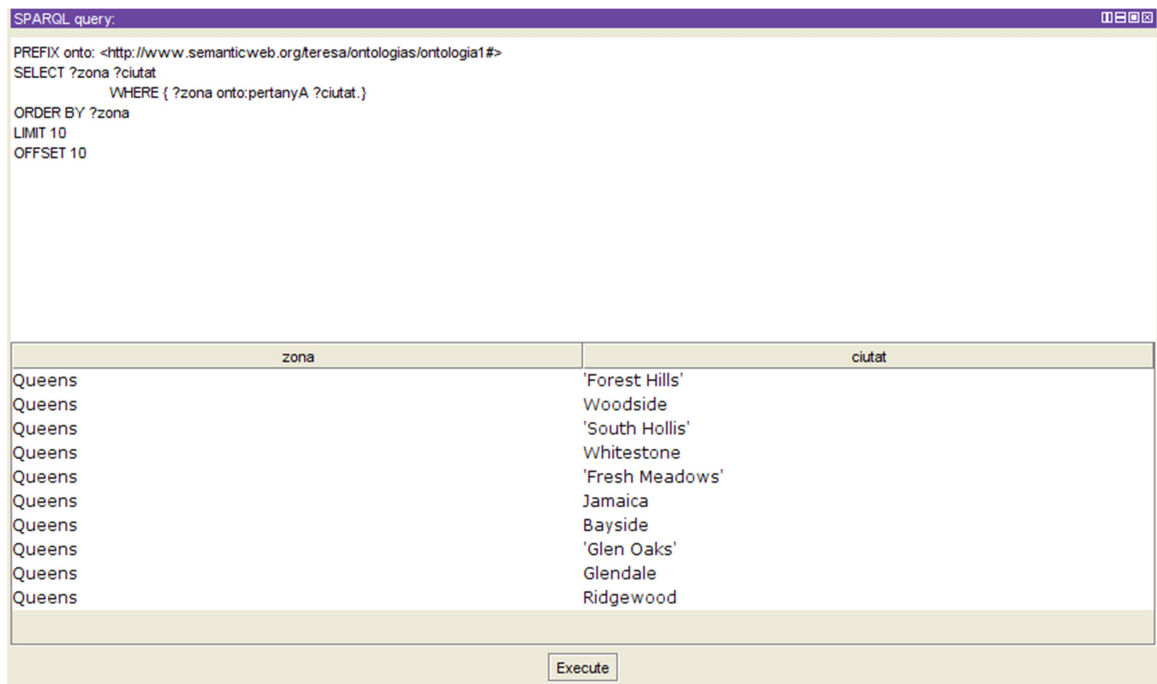


Figura 27: Resultat consulta amb modificadors “Mostra les ciutats i les seves zones”

El modificador *OFFSET* produeix que les solucions generades comencin després d'un nombre específic de solucions.

La clàusula *LIMIT* estableix un límit superior al nombre de solucions retornades. Si el nombre de solucions reals és més gran que el límit, llavors com a molt seran retornades el nombre límit de solucions.

El resultat de realitzar aquesta consulta mostra 10 solucions (*modificador limit*), a partir de la solució 10 (*modificador offset*).

- **Mostra les zones de la ciutat de New York**

Per realitzar aquesta consulta s'ha utilitzat els filtres del llenguatge SPARQL. La concordança amb un patró de graf produeix una seqüència de solució, on cada solució té un conjunt de vincles de variables amb termes RDF. Els filtres (**FILTERS**) SPARQL restringeixen les solucions a aquelles per a les quals l'expressió del filtre s'avalua com a veritable (*true*).

Per realitzar la consulta, s'ha utilitzat un filtre per cercar mitjançant la classe **rdfs:label**.

RDFS és una extensió semàntica de RDF. És un llenguatge primitiu d'ontologies que proporciona els elements bàsics per a la descripció de vocabularis. Un arxiu RDFS és un arxiu RDF, és a dir, es tracta d'un arxiu amb la mateixa sintaxi i la mateixa estructura que la que s'usa en RDF. La sintaxi està basada en XML.

RDFS té diferents tipus de classes. En concret, a aquesta consulta s'utilitza rdfs: label, que és una instància de rdf: Property que s'usa per a proporcionar una versió clarament comprensible del nom d'un recurs.

La consulta quedaria de la següent forma per a mostrar el lloc wifi de la ciutat de New York:

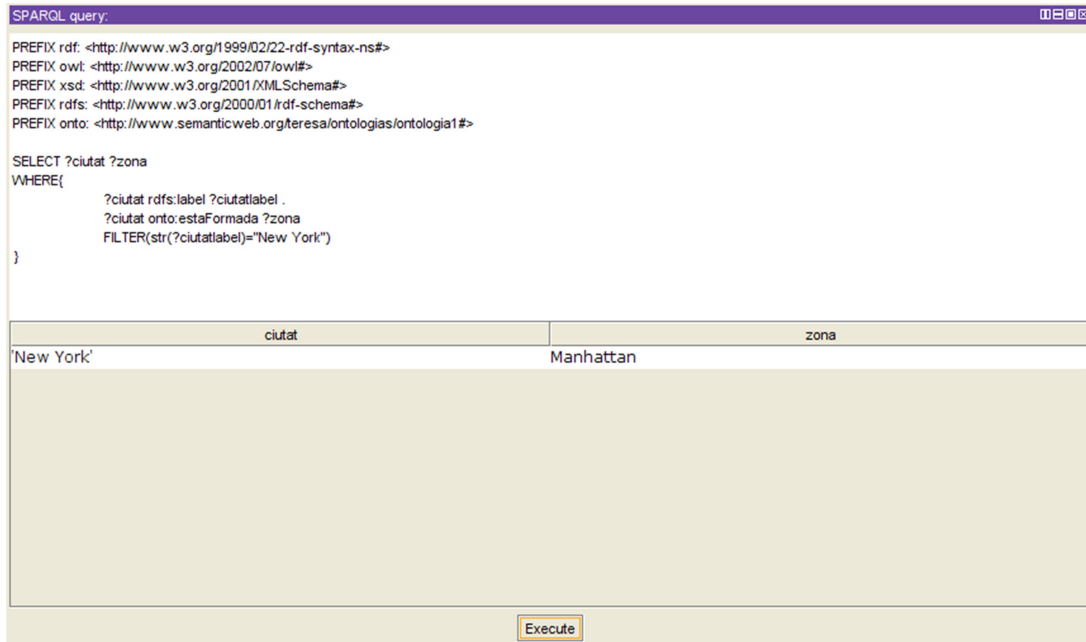


Figura 28: Resultat consulta "Mostra les zones de la ciutat de New York"

- **Mostra la localització del punt wifi amb l'identificador 705**

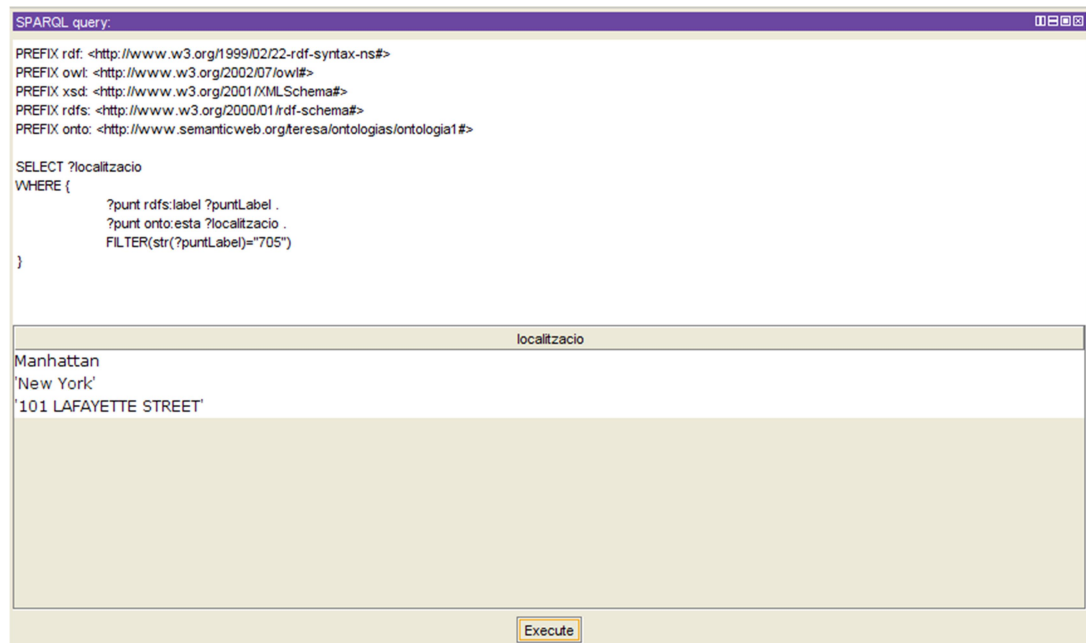


Figura 29: Resultat consulta "Mostra la localització del punt wifi amb l'identificador 705"

En realitzar la consulta de la localització, es pot observar que la primera fila mostra la Zona en la que està localitzat, la segona mostra la Ciutat i l'última mostra l'Adreça. Açò es produeix perquè Zona, Ciutat i Adreça són subclasses de la classe Localització.

- **Mostra el número de punts wifi d'una ciutat, en concret de "Staten Island".**

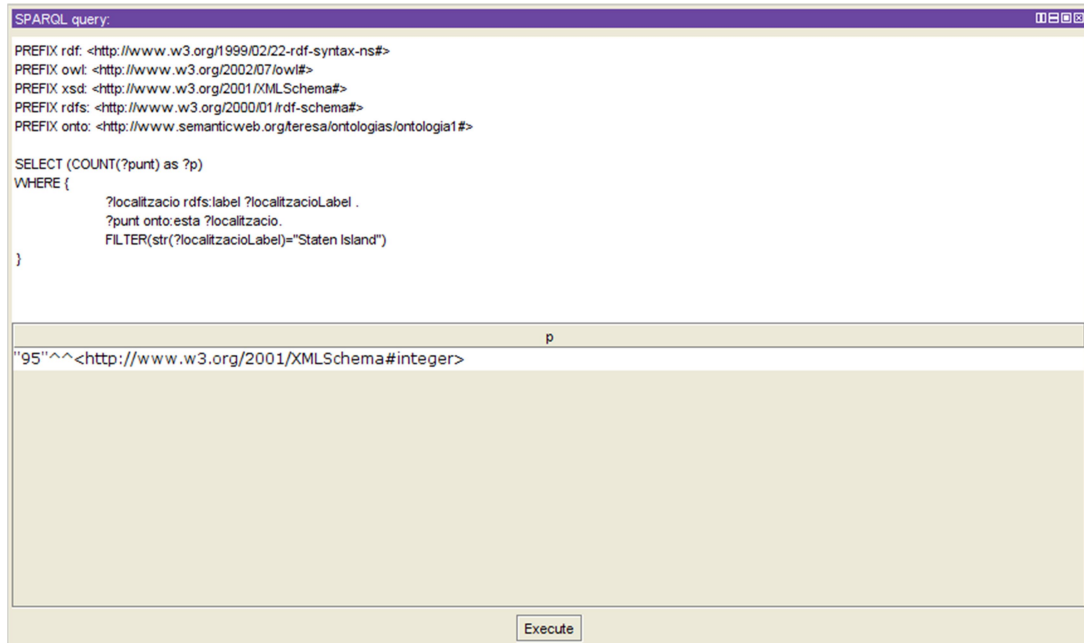


Figura 30: Resultat consulta "Mostra el número de punts wifi d'una ciutat, en concret de "Staten Island""

- **Mostra les coordenades del punt wifi amb id 822**

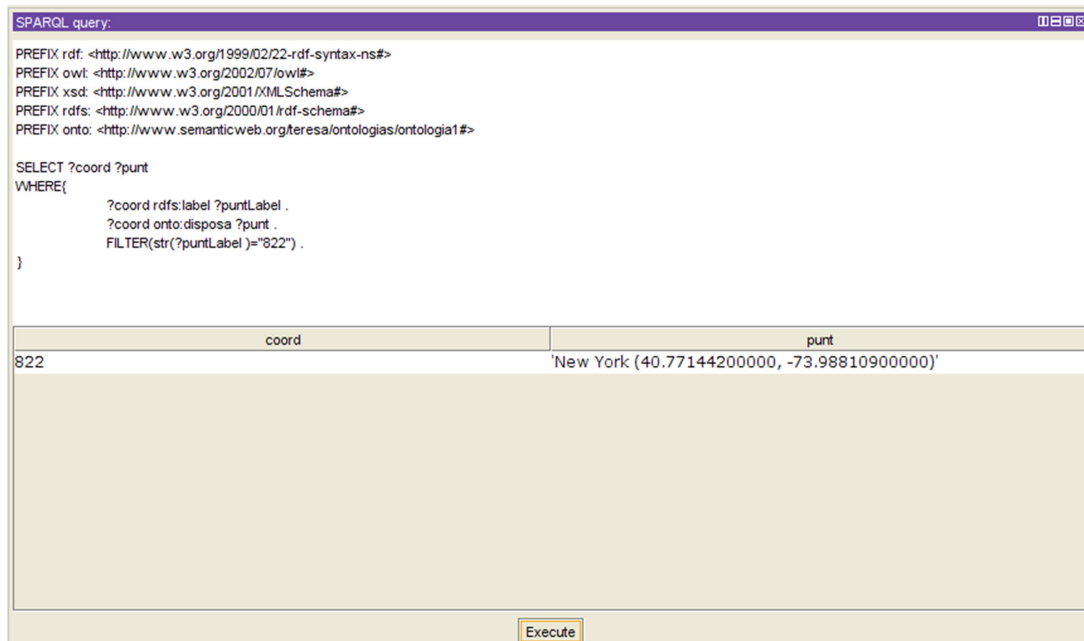


Figura 31: Resultat consulta "Mostra les coordenades del punt wifi amb id 822"

## 5. Conclusions i línies de futur

### 5.1. Conclusions

Vaig escollir la web semàntica com a tema principal del treball final de grau perquè he treballat amb el llenguatge XML i també he desenvolupat diferents aplicacions web. Però he pogut comprovar que la web semàntica va més enllà que conèixer el funcionament de la web i XML.

Per realitzar el treball he hagut de cercar molta informació perquè, ja que els meus coneixements del tema d'ontologies eren pràcticament nuls. M'ha suposat un esforç treballar tots els termes d'aquesta especialitat, ja que la informació que hi ha sobre aquest tema és poca o relativament nova.

Considero que aquest treball és un aprenentatge en el món de web semàntica i de l'ontologia, ha sigut un continu descobriment d'eines i llenguatges que desconeixia i que m'han semblat útils i he pogut comprovar les altes possibilitats que té el llenguatge XML. Cal destacar l'esforç del W3C per poder crear estàndards per poder representar la informació.

Des de la planificació del disseny de l'ontologia vaig tenir en compte que l'eina de creació d'ontologies Protégé seria la que utilitzaria per a la seva implementació. Una vegada analitzades les dades i fet el disseny de l'ontologia és quan van començar els problemes. Vaig crear l'ontologia amb Protégé però em vaig adonar que la versió de Protégé no permetia la importació de dades en format CSV. Vaig investigar molt sobre com poder importar dades en Protégé 5.2, però l'única forma que vaig trobar va ser utilitzar una altra eina de la mateixa companyia anomenada WebProtégé per poder importar les dades dels fitxers CSV en forma d'instàncies.

El fitxer descarregat de la ciutat de Buenos Aires no està en format Unicode, per tant, sols s'ha portat a terme l'ontologia amb les dades de la ciutat de New York.

Una vegada vaig poblar l'ontologia amb WebProtégé em vaig trobar un altre problema, les propietats de tipus Objecte no s'importaven d'una forma correcta a Protégé. Vaig haver d'investigar, una altra vegada, el perquè d'aquest nou inconvenient. Pel que sembla els formats OWL quan s'importa una ontologia en Protégé des de Webprotégé donen problemes.

Al veure que tenia molts inconvenients amb el mètode que havia escollit, vaig intentar realitzar l'ontologia amb un l'altre mètode que havia estudiat, utilitzant la llibreria Apache Jena. Per tant, vaig tenir que començar de nou amb l'estudi de la creació d'ontologia i importació de fitxers CSV amb Jena. Per falta de temps, i de no tenir informació sobre com utilitzar aquesta llibreria, vaig decidir reprendre el treball tal com l'havia començat, amb Protégé.

Gràcies al Raonador de Protégé vaig poder solucionar el tema de les propietats de tipus Objecte que, he comentat abans, i les instàncies es van crear amb totes les seves propietats.

Una vegada implementada l'ontologia i feta la seva població, vaig passar a realitzar les consultes SPARQL. Com he comentat anteriorment, el treball l'he realitzat sobre el sistema operatiu Windows 10, en realitzar les consultes amb Protégé, el programa es quedava aturat i no deixava accedir a la pestanya corresponent de les consultes SPARQL. Buscant informació sobre aquest cas, vaig descobrir

que l'última versió de Protégé està tenint problemes amb aquest sistema operatiu, per tant vaig haver d'instal·lar la versió 4.3 de Protégé en la que he pogut acabar el treball sense més problemes.

Pel que fa a la planificació i seguiment del treball no ha estat encertat 100%, ja que per motius laborals he tingut alguns canvis de tasques laborals al final del projecte i he hagut d'invertir molt temps per poder finalitzar-lo. A més he tingut que fer desplaçaments en els últims mesos per motius laborals estant incomunicada i sense poder avançar com calia.

En conclusió, aquest treball ha sigut un continu aprenentatge en el que he pogut comprovar la gran capacitat de la web semàntica i de les ontologies.

## **5.2. Línies de futur**

Les línies de futur d'aquest projecte estarien enfocades a afegir més punts wifi d'altres ciutats. Per raons de temps no s'ha creat les instàncies de la ciutat de Buenos Aires, ja que el fitxer no estava amb el format correcte per a l'aplicació Webprotégé. Tenint en compte que el disseny i la implementació de l'ontologia està fet, sols s'hauria de fer importacions de les ciutats que es volguessin afegir.

## Bibliografia

**Tim Barners-Lee, James Hendler, Ora Lassila.** "The Semantic Web". Scientific American. Maig 2001.

**W3C.** "Guía Breve de Web Semántica". [data de consulta: 01/10/2017].

<<https://www.w3c.es/Divulgacion/GuiasBreves/WebSemantica>>

**R. Studer, V.R. Benjamins, D. Fensel.** (1998). "Knowledge Engineering: Principles and Methods". [data de consulta: 28/10/2017]

<<http://www.inf.furb.br/~paulofernando/downloads/EGC/fundamentos%20de%20ec/KnowledgeEngineering-PrinciplesAndMethods.pdf>>

**Gruber, Thomas R.** (1993). "A Translation Approach to Portable Ontology Specifications". Knowledge Systems Laboratory. [article en línia] [data de consulta: 02/10/2017]

<<http://tomgruber.org/writing/ontolingua-kaj-1993.pdf>>

**Information System Group.** "Hermit". [data de consulta: 20/12/2017].

<<http://www.hermit-reasoner.com/>>

**World Wide Web Foundation.** (2017). "Open Data Barometer". [data de consulta: 27/10/2017]

<<http://opendatabarometer.org>>

**W3C.** (2004). "Lenguaje de Ontologías Web (OWL) Vista General". [data de consulta: 02/10/2017].

<<https://www.w3.org/2007/09/OWL-Overview-es.html>>

**Noy & McGuinness.** (2001). "Ontology Development 101: A Guide to Creating Your

First Ontology". [data de consulta: 25/10/2017].

<<http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.pdf>>

**Stanford University.** (2016). "Protégé". [data de consulta: 03/10/2017]. <<https://protege.stanford.edu>>

**Protégé Wiki.** (2013). "Protégé Plugin Library" [data de consulta: 04/12/2017]

<[https://protegewiki.stanford.edu/wiki/Protege\\_Plugin\\_Library](https://protegewiki.stanford.edu/wiki/Protege_Plugin_Library)>

**Protégé OntoGraf Wiki.** (2010). "Ontograf" [data de consulta: 06/12/2017]

<<https://protegewiki.stanford.edu/wiki/OntoGraf>>

**Protégé WebProtégé Wiki.** (2016). "WebProtégé" [data de consulta: 13/12/2017]

<<https://protegewiki.stanford.edu/wiki/WebProtege>>

**Swoogle Semantic Web Search.** (2006). "Swoogle" [data de consulta: 1/11/2017]

<<http://swoogle.umbc.edu/2006/>>



**Apache Jena.** (2017). "Apache Jena" [data de consulta: 26/12/2017] <<https://jena.apache.org>>

**W3C.** (2008) "SPARQL Lenguaje de consulta para RDF" [data de consulta: 26/12/2017]

<<http://skos.um.es/TR/rdf-sparql-query/>, consultat 26/12/2017>

# Annexos

## Annex A: Glossari

**API:** Application Programming Interface, "Interfície de Programació d'Aplicacions"

**CSV:** Comma-Separated Values, "Valors separats per columnes".

**ODB:** Open Data Barometer, "Baròmetre de Dades Obertes".

**Open Data:** Open Data, "Dades Obertes".

**OWL:** Web Ontology Language, "Llenguatge d'Ontologies Web".

**RDF:** Resource Description Framework, "Marc de Descripció de Recursos".

**SGML:** Standard Generalized Markup Language, "Llenguatge d'Etiquetatge Generalitzat Estàndard"

**SPARQL:** SPARQL Protocol and RDF Query Language, "Llenguatge de consultes per a grafs RDF".

**W3C:** World Wide Web Consortium, "Consorci World Wide Web".

**Wifi:** Connexió de dispositius electrònics sense fils.

**WWW:** World Wide Web, "Sistema de distribució de documents d'hipertext o hipermitjans interconnectats i accessibles via Internet".

**XML:** Extensible Markup Language, "Llenguatge de Marques Extensible".

## Annex B: Lliurables del projecte

Amb l'entrega de la memòria, s'adjunta el fitxer de l'ontologia anomenat: "**ontologiaWifi**".