



LEARNING, APPLYING, MULTIPLYING BIG DATA ANALYTICS

---

**Horizon 2020 Grant Agreement No 809965**  
**Contract start date: July 1st 2018, Duration: 36 months**

## **LAMBDA Lecture**

### **Upotreba alata GraphDB**

Due date of deliverable: 30/06/2021  
Actual submission date: 30/06/2021

Revision: Version 1.0



This project has received funding from the European Union's Horizon 2020 Research and Innovation programme, H2020-WIDESPREAD-2016-2017 Spreading Excellence and Widening Participation under grant agreement No 809965.



Workpackage	WP 3 Cooperation for Teacher and Student Training
Responsible for WP	Institute for Computer Science - University of Bonn
Deliverable Lead	Institute Mihajlo Pupin (Valentina Janev)

© Copyright the LAMBDA Consortium. The LAMBDA Consortium comprises:

Institute Mihajlo Pupin ( <b>PUPIN</b> )	Co-ordinator	Serbia
Fraunhofer Institute for Intelligent Analysis and Information Systems ( <b>Fraunhofer/IAIS</b> )	Contractor	Germany
Institute for Computer Science - University of Bonn ( <b>UBO</b> )	Contractor	Germany
Department of Computer Science - University of Oxford ( <b>UOXF</b> )	Contractor	UK

**Disclaimer:**

The information in this document reflects only the authors' views and the European Community is not liable for any use that may be made of the information contained therein. The information in this document is provided "as is" without guarantee or warranty of any kind, express or implied, including but not limited to the fitness of the information for a particular purpose. The user thereof uses the information at his/her sole risk and liability.



## Executive Summary

U ovoj lekciji biće opisani principi funkcionisanja semantičkog veba. Definisaće se značenja ontologija, taksonomija i ostalih izraza koji se često koriste. Praktična upotreba ovih tehnologija biće pokazana korišćenjem alata GraphDB kompanije Ontotext. Kroz primere u alatu GraphDB biće objašnjeno kako se pišu upiti za pretraživanje informacija u bazama podataka korišćenjem jezika SPARQL. Grafovi znanja i povezani podaci biće objašnjeni korišćenjem ovog alata.



## Table of Contents

Executive Summary .....	3
Table of Contents .....	4
1. Ontologije I taksonomije.....	6
Grafovi znanja .....	6
RDF i RDFS .....	7
2. Alat GraphDB.....	7
3. Sparql i grafovi znanja u GraphDB.....	11
Reference .....	14



# 1. Ontologije I taksonomije

Ontologija u filozofskom značenju predstavlja postojanje nečega u stvarnosti. U semantičkom vebru, ontologija se bavi istraživanjem kategorija stvari koje dokazano postoje ili se pretpostavlja da postoje u nekom domenu. Ontologija se bavi definisanjem pomenutih stvari i relacijama između njih. Ontologije postoje kako bi informacije dostupne na internetu bile dostupne za procesiranje na računarima.

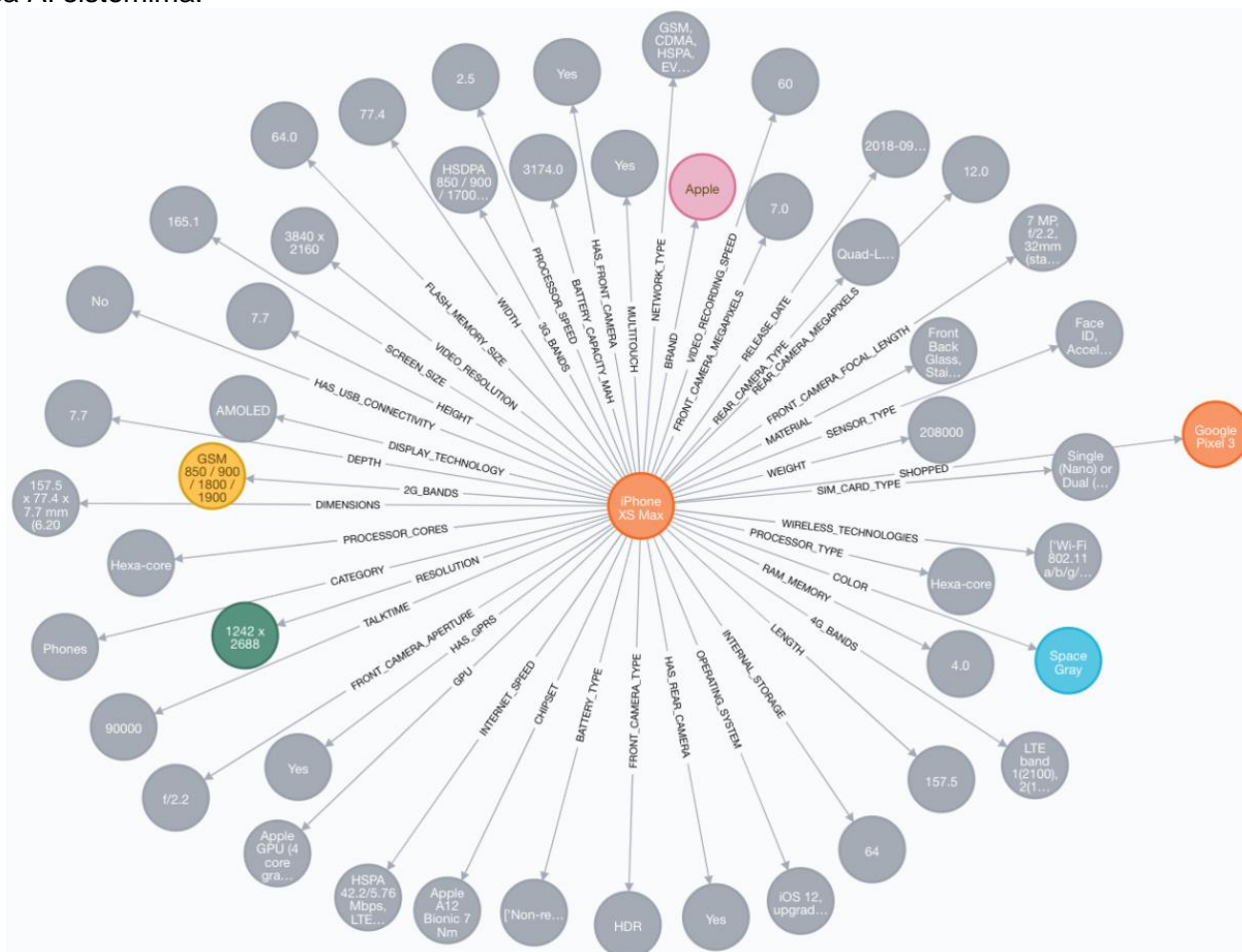
Taksonomije formalizuju veze između koncepata i određuju koji izrazi se odnose na koji koncept. Taksonomije se mogu koristiti za klasifikovanje dokumenata u kategorije, gde klasifikacija može biti prema temi ili tipu, na primer da li je u pitanju email ili neki izveštaj. Model ontologije  $O$  predstavljen je tripletom:

$$O = \langle X, R, F \rangle,$$

gde je  $X$  konačni skup klasa (koncepta),  $R$  je konačni skup atributa koji obezbeđuju veze između entiteta u nekom domenu, dok je  $F$  predstavljen konačni skup interpretacija funkcija koje su definisane za entitet.

## Grafovi znanja

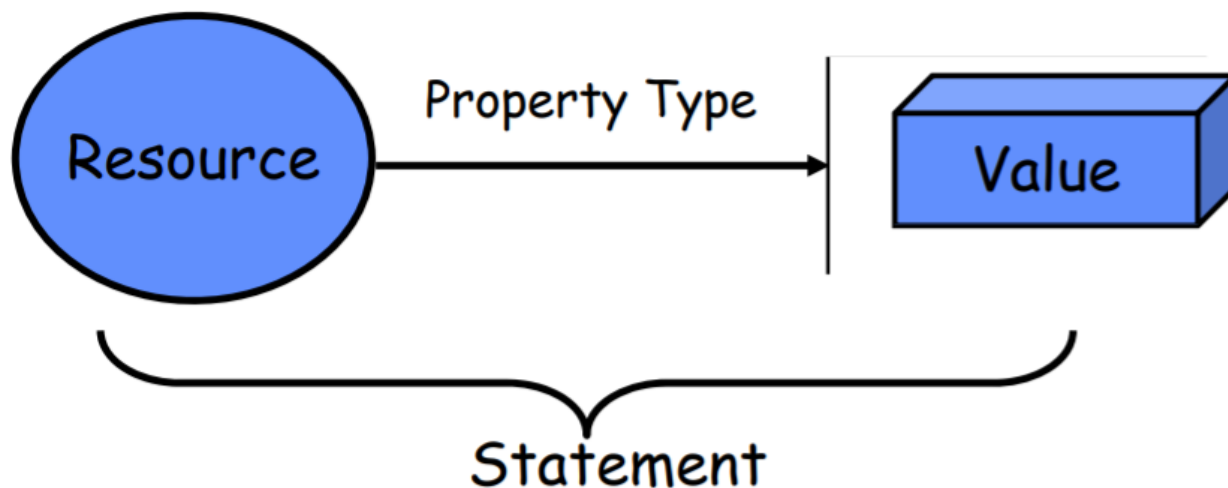
Graf znanja predstavlja podatke međusobno povezane u nekom domenu. Svaki podatak predstavljen je kao čvor koji ima veze ka ostalim podacima tj. čvorovima. Na ovaj način izbačena je upotreba klasičnih baza podataka gde su podaci predstavljani u tabelama i vezani ključevima preko kojih se pristupa vezanim informacijama. Ovi grafovi su od koristi kako ljudima, tako i računarima. Sve velike kompanije već duže vreme poseduju i rade na usavršavanju sopstvenih grafova znanja. Google koristi svoj graf znanja za upotrebu Google Asistenta i Google Home uređaja u kombinaciji sa AI sistemima.



Slika 1 – Primer grafa znanja

## RDF i RDFS

RDF (Resource Description Framework) je deo W3C standarda koji se koristi za opisivanje meta podataka. Za predstavljanje podataka koristi se struktura XML – a. Korišćenjem XML – a omogućujemo mašinama da razumeju značenje podataka i informacija. Takođe, moguće je izvršiti pretragu nad ovim podacima korišćenjem upita u Sparql jeziku.



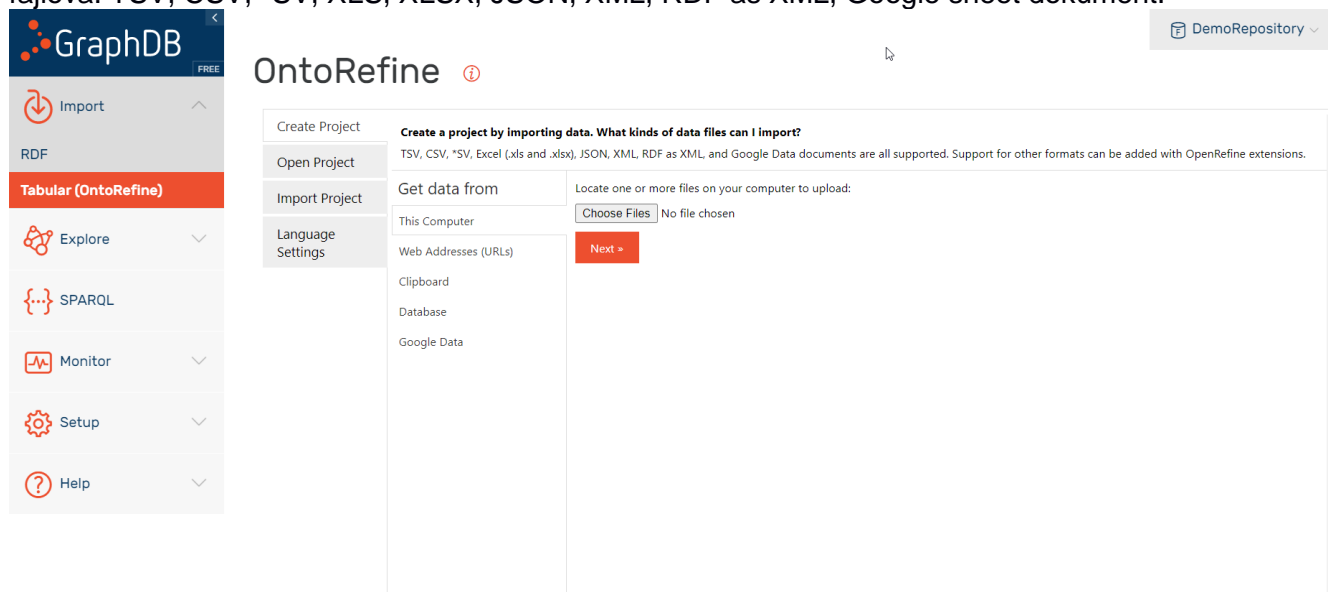
Slika 2 – Primer definisanja resursa u RDF

RDF – S (Schema) je nadogradnja na RDF i omogućuje definisanje hijerarhije klasa i veza među resursima.

## 2. Alat GraphDB

Alat GraphDB koristi se za izradu grafova znanja. Omogućuje pretragu i vizuelizaciju podataka korišćenjem Sparql jezika.

Podaci se mogu dodati korišćenjem opcije OntoRefine koja podržava upload gotovo svih formata fajlova: TSV, CSV, \*SV, XLS, XLSX, JSON, XML, RDF as XML, Google sheet dokument.



Slika 3 – OntoRefine import fajlova



# OntoRefine ?

Create Project << Start Over Configure Parsing Options Project name Netherlands\_restaurants.csv Tags Create Project >

	Trcid	Title	Shortdescription	Longdescription	Calendarsummary	TitleEN	ShortdescriptionEN	LongdescriptionEN
1.	669d7d82-8962-4e88-b2e1-7b8706633aa0	Smits Noord-Zuid Hollandsch Koffiehuis	Het Smits Koffiehuis ontleent haar ontstaan aan de stoomtram die de verbinding onderhield met Amsterdam naar het noorden van de provincie en is in 1919 gebouwd. Nu is er een restaurant en een koffiebar. Ook is hier een informatiekantoor van Amsterdam Marketing gehuisvest.			Smits Noord-Zuid Hollandsch Koffiehuis	The Smits Koffiehuis dates back to 1919. This charming building served as the departure and arrival point for a steam tram that once connected Amsterdam to the northern parts of the Noord Holland province. In addition to the restaurant and café, this beautiful landmark in front of Central Station also houses the Tourist Information Office and a GVB (public transport) office.	

Parse data as Character encoding Update Preview

**CSV / TSV / separator-based files**

- Line-based text files
- Fixed-width field text files
- PC-Axis text files
- JSON files
- MARC files
- JSON-LD files
- RDF/N3 files

Columns are separated by

- commas (CSV)
- tabs (TSV)
- custom: ;

Escape special characters with \

- Column names (comma separated):

Options:

- Ignore first 0 line(s) at beginning of file
- Parse next 1 line(s) as column headers
- Discard initial 0 row(s) of data
- Load at most 0 row(s) of data
- Use character " " to enclose cells containing column separators
- Parse cell text into numbers, dates, ...
- Store blank rows
- Store blank cells as nulls

Preferences: Help, About

Slika 4 – Import CSV fajla

Pri importu podataka iz CSV fajla možemo definisati opcije kao što su pozicija header – a, koji separator se koristi i druge.

SQL Exporter

Content Download

Field Name	SQL Type	Size	Allow Null	Default
<input checked="" type="checkbox"/> Trcid	VARCHAR		<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/> Title	VARCHAR		<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/> Shortdescription	VARCHAR		<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/> Longdescription	VARCHAR		<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/> Calendarsummary	VARCHAR		<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/> TitleEN	VARCHAR		<input checked="" type="checkbox"/>	

Select All De-select All

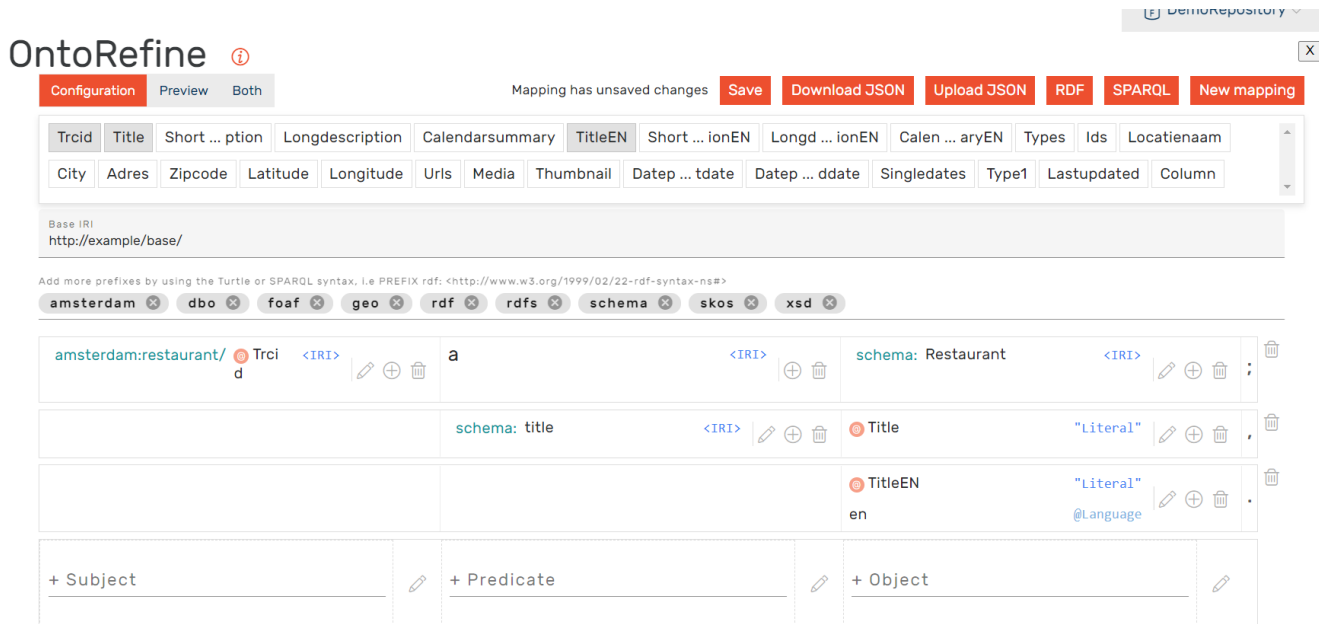
Output empty row (i.e. all cells null)  Ignore facets and filters and export all rows  Trim Column Names

Cancel

Slika 5 – Sql export

Podaci se nakon obrade mogu eksportovati u SQL bazu podataka.





Slika 6 – Kreiranje RDF tripleta

Pre samog kreiranja RDF tripleta, potrebno je dodati prefikse koje ćemo koristiti. U ovom slučaju, pored automatski dodatih foaf, dbo, geo, rdf, skos i xsd, potrebno je kreirati prefikse amsterdam, dbo i schema.

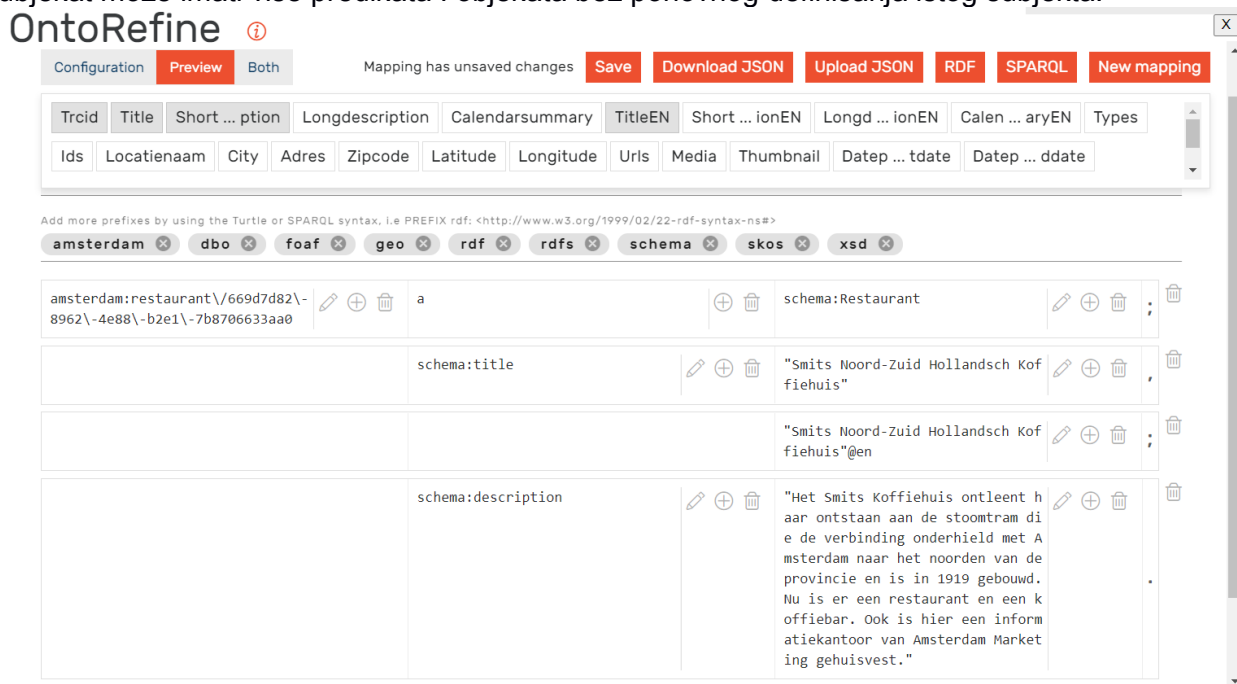
Prefiks se definiše na sledeći način:

PREFIX amsterdam: <https://data/amsterdam/nl/resource/> ,

PREFIX dbo: <http://dbpedia.org/ontology/> ,

PREFIX schema: <http://www.schema.org/>

RDF triplet opisan je pomoću subjekta, predikata i objekta. Korišćenjem Turtle sintakse, jedan subjekat može imati više predikata i objekata bez ponovnog definisanja istog subjekta.



Slika 7 – Preview opcija

Korišćenjem opcije preview možemo videti koje vrednosti dobijamo za subjekat, predikat i objekat. Ova opcija je korisna kako bi proverili da li smo negde napravili greški pri definisanju prefiksa.



```
@base <http://example/base/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix geo: <http://www.opengis.net/ont/geosparql#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix dbo: <http://dbpedia.org/ontology/> .
@prefix amsterdam: <https://data/amsterdam.nl/resource/> .
@prefix schema: <http://www.schema.org/> .

<https://data/amsterdam.nl/resource/restaurant/669d7d82-8962-4e88-b2e1-7b8706633aa0>
  a <http://www.schema.org/Restaurant>;
  <http://www.schema.org/title> "Smits Noord-Zuid Hollandsch Koffiehuis", "Smits Noord-Zuid Hollandsch Koffiehuis"@en;
  <http://www.schema.org/description> "Het Smits Koffiehuis ontleent haar ontstaan aan de stoomtram die de verbinding onderhield met Amsterdam naar het noorden van de provincie en is in

<https://data/amsterdam.nl/resource/restaurant/8d9eb314-4433-477e-84b7-35a3552b6f8e>
  a <http://www.schema.org/Restaurant>;
  <http://www.schema.org/title> "Afrikaans Restaurant Kilimanjaro", "African Restaurant Kilimanjaro"@en;
  <http://www.schema.org/description> "Een Afrikaans (Senegal, Ghana, Zuid-Afrika e.a.) restaurant tegenover NEMO, dichtbij het Waterlooplein. De menukaart is voor degenen die niet van

<https://data/amsterdam.nl/resource/restaurant/ca0e88a6-5cd2-4dc0-8d6a-4273f84721bd>
  a <http://www.schema.org/Restaurant>;
  <http://www.schema.org/title> "Akbar Indian Restaurant", "Akbar Indian Restaurant"@en;
  <http://www.schema.org/description> "Een chique restaurant, ingericht in Moghul stijl, waarin de gast het gevoel krijgt dat hij in India aan het eten is. De keuken serveert gerechten

<https://data/amsterdam.nl/resource/restaurant/adcd917e-a472-4aed-aace-999c95e97006>
  a <http://www.schema.org/Restaurant>;
  <http://www.schema.org/title> "Alfonso's Mexican Restaurant", "Alfonso's Mexican Restaurant"@en;
  <http://www.schema.org/description> "In de Utrechtsestraat in Amsterdam, op slechts vijf minuten loopafstand van het Rembrandtplein, de Kleine Komedie en Carré, tref je Alfonso's: eer

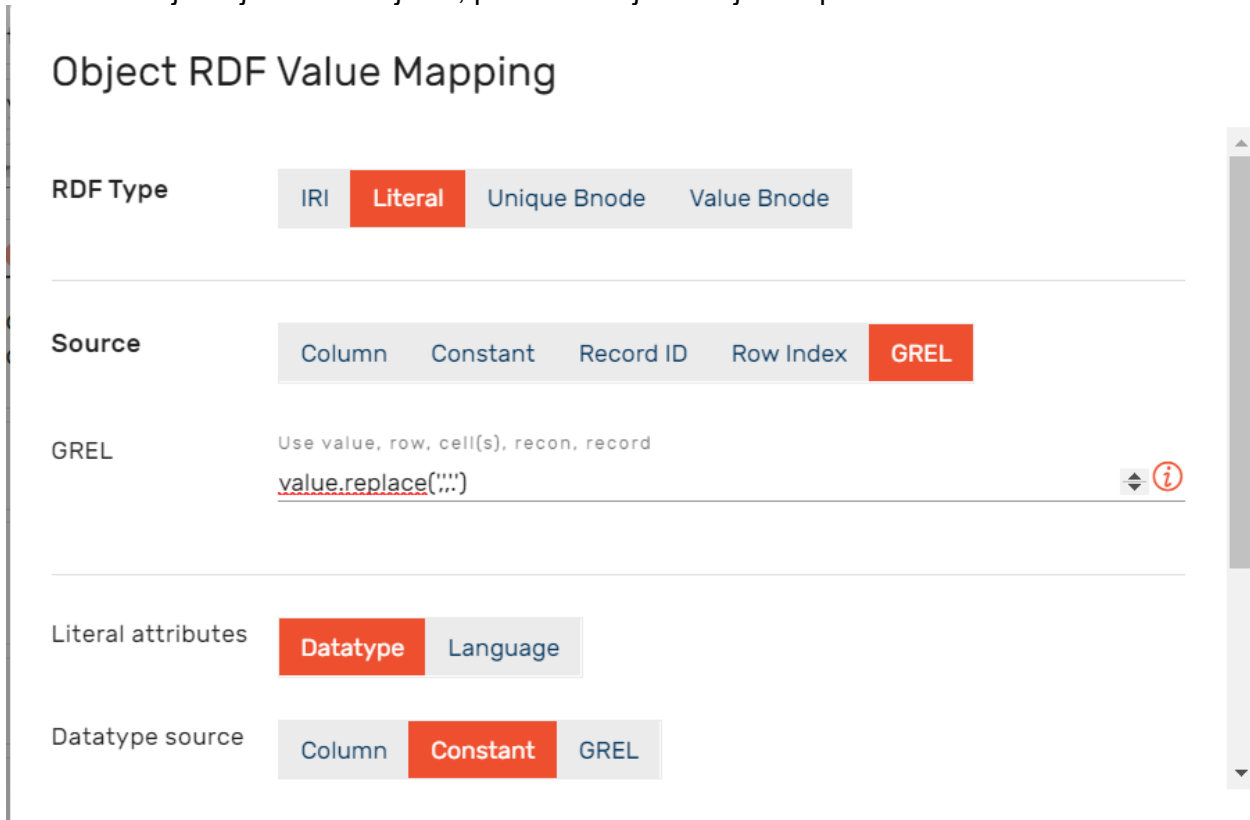
<https://data/amsterdam.nl/resource/restaurant/8c83496c-0b8b-41f2-987c-21ddb78b2978>
  a <http://www.schema.org/Restaurant>;
  <http://www.schema.org/title> "Café het Paleis", "Café het Paleis"@en;
  <http://www.schema.org/description> "Het café-restaurant Paleis, vlak achter het Paleis op de Dam, is al jaren favoriet vanwege de betaalbare gerechten, maar het is ook een plek om ev

<https://data/amsterdam.nl/resource/restaurant/56b1dd7d-4ac3-4e24-826e-cb8b55617f42>
  a <http://www.schema.org/Restaurant>;
  <http://www.schema.org/title> "Amstel Bar & Brasserie", "Amstel Bar & Brasserie"@en;
  <http://www.schema.org/description> "De Amstel Bar & Brasserie is een informele plek voor lunch, diner of late maaltijd na het theaterbezoek. Ook voor cocktails kun je hier terecht. I

<https://data/amsterdam.nl/resource/restaurant/2760a2d8-0404-45df-aa6e-fd88c602d5b9>
  a <http://www.schema.org/Restaurant>;
```

Slika 8 – RDF export

OntoRefine omogućuje da izvezemo podatke u formi RDF tripleta. Na slici iznad možemo videti primer RDF fajla koji sadrži subjekat, predikat i objekat koji smo podesili.



Slika 9 – GREL opcija

Ukoliko imamo neke kolone koje sadrže neke karaktere koji nisu dozvoljeni, možemo ih zameniti GREL funkcijom:

*value.replace(',', '.')*

U ovom slučaju, menjamo zarez sa tačkom.

Moguće je mapirati i slike. U ovom slučaju, koristimo schema: image prefix ya predikat, dok kao objekat dodajemo kolonu u CSV fajlu koja sadrži URL do slike. Kao tip objekta, potrebno je izabrati IRI (Internationalized Resource Identifier).



```
8 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
9 PREFIX dbo: <http://dbpedia.org/ontology/>
10 PREFIX amsterdam: <https://data/amsterdam/nl/resource/>
11 PREFIX schema: <http://www.schema.org>
12
13 CONSTRUCT {
14   ?s1 a schema:Restaurant ;
15     schema:title ?o_title, ?o_title_2 ;
16     schema:description ?o_description ;
17     schema:latitude ?o_latitude ;
18     schema:longitude ?o_longitude ;
19     amsterdam:zipcode ?o_zipcode ;
20     schema:image ?o_image .
```

Table Raw Response Pivot Table Google Chart Download as Visual

Filter query results Showing results from 1 to 1,000 Query took 0.7s, minutes ago.

	subject	predicate	object
1	<a href="https://data/amsterdam/nl/resource/restaurant/669d8962-4e88-b2e1-7b8706633aa0">https://data/amsterdam/nl/resource/restaurant/669d8962-4e88-b2e1-7b8706633aa0</a>	<a href="http://www.schema.org/rdf:type">http://www.schema.org/rdf:type</a>	<a href="http://www.schema.org/Restaurant">http://www.schema.org/Restaurant</a>
2	<a href="https://data/amsterdam/nl/resource/restaurant/669d8962-4e88-b2e1-7b8706633aa0">https://data/amsterdam/nl/resource/restaurant/669d8962-4e88-b2e1-7b8706633aa0</a>	<a href="http://www.schema.org/title">http://www.schema.org/title</a>	"Smits Noord-Zuid Hollandsch Koffiehuis"
3	<a href="https://data/amsterdam/nl/resource/restaurant/669d8962-4e88-b2e1-7b8706633aa0">https://data/amsterdam/nl/resource/restaurant/669d8962-4e88-b2e1-7b8706633aa0</a>	<a href="http://www.schema.org/title">http://www.schema.org/title</a>	"Smits Noord-Zuid Hollandsch Koffiehuis"@en
4	<a href="https://data/amsterdam/nl/resource/restaurant/669d8962-4e88-b2e1-7b8706633aa0">https://data/amsterdam/nl/resource/restaurant/669d8962-4e88-b2e1-7b8706633aa0</a>	<a href="http://www.schema.org/description">http://www.schema.org/description</a>	"Het Smits Koffiehuis ontleent haar ontstaan aan de stoomtram die de verbinding onderhield" keyboard shortcuts

Slika 10 – Sparql upit

Na slici iznad možemo videti kako izgledaju rezultati kada se pokrene Sparql upit nad importovanim RDF tripletima.

### 3. Sparql i grafovi znanja u GraphDB

Sparql upite moguće je pokretati odabirom stavke Sparql u glavnom meniju.

The screenshot shows the GraphDB interface. On the left is a navigation menu with options: Import, Explore, SPARQL (highlighted), Monitor, Setup, and Help. The main area is titled 'SPARQL Query & Update' and contains a query editor with the following code:

```
1 PREFIX pub: <http://ontology.ontotext.com/taxonomy/>
2 SELECT ?person ?occupation WHERE {
3   ?x pub:occupation ?o .
4   ?x a pub:Person .
5   ?x pub:preferredLabel ?person .
6   {
7     ?o pub:hasValue ?value .
8     ?value pub:preferredLabel ?occupation .
9   } UNION {
10    ?o pub:hasValue ?occupation .
11    filter (isLiteral(?occupation)) .
12  }
13 }
```

Below the editor are tabs for 'Table', 'Raw Response', 'Pivot Table', and 'Google Chart', and a 'Download as' button. A status bar at the bottom indicates 'Showing results from 1 to 189 of 189. Query took 0.1s, minutes ago.'

Slika 11 – Primer Sparql upita

Na slici iznad prikazan je Sparql upit koji vraća sve ljude iz baze i njihova zanimanja.



	person	occupation
1	"Ben Bernanke"@en	"economist"@en
2	"Ben Bernanke"@en	"politician"@en
3	"Maurice R. Greenberg"@en	"business executive"@en
4	"Jeff Bezos"@en	"computer scientist"@en
5	"Alberto Giacometti"@en	"artist"@en
6	"Larry King"@en	"actor"@en
7	"Larry King"@en	"television presenter"@en
8	"Larry King"@en	"writer"@en
9	"Larry King"@en	"radio host"@en
10	"Larry King"@en	"voice actor"@en
11	"Larry King"@en	"journalist"@en

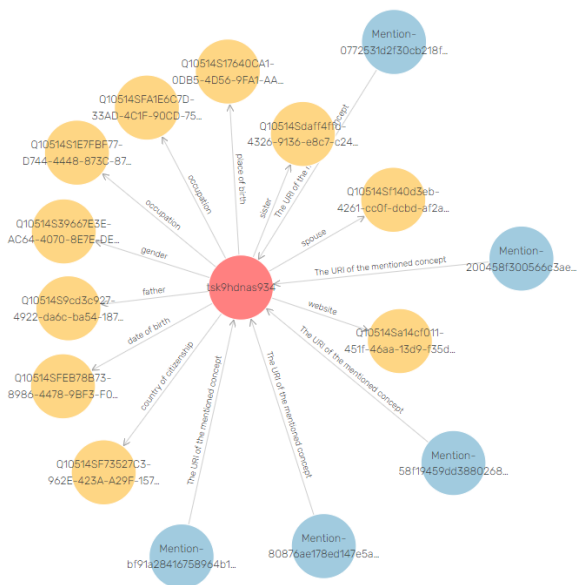
Slika 12 – Rezultati upita

Sve ljude koji postoje u bazi možemo izlistati sledećim upitom:

```
PREFIX pub: < http://ontology.ontotext.com/taxonomy/>
PREFIX publishing: < http://ontology.ontotext.com/publishing# >
SELECT ? x ? label WHERE {
    ? x a pub: Person .
    ? x pub: preferredLabel ? label . }
```

Izlistane podatke možemo vizuelizovati grafovima znanja. Grafovi znanja u GraphDB prikazani su kao čvorovi. Svaki čvor povezan je sa ostalima u zavisnosti od toga koliko podataka je povezano i uneto u bazu.

### Visual graph



[tsk9hdnas934](#)

tsk9hdnas934

Types:

**pub:Athlete**

RDF rank:

0

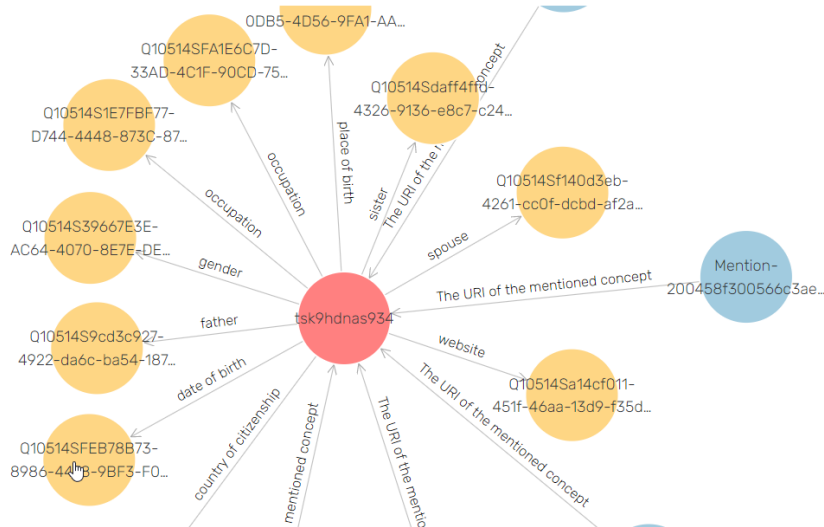
pub:preferredLabel  
**Fernando Alonso**

Slika 13 – Graf znanja

Na slici iznad prikazani su nodovi, tj. čvorovi vezani za resurs koji opisuje Fernanda Alonsa. Vidimo da postoji dosta vezanih nodova koji sadrže informacije o mestu rođenja osobe, roditeljima osobe, zanimanju itd. Do svakog čvora postoji veza sa značenjem. Na primer, praćenjem putanje koja je



nazvana „website“, dolazimo do podataka o web stranici koja je vezana za ovu osobu. Praćenjem veze „dateOfBirth“ dolazimo podatka o datumu rođenja ove osobe.



**[Q10514SFEB78B73-8986-4478-9BF3-F03F3A747CFB](#)**

[Q10514SFEB78B73-8986-4478-9BF3-F03F3A747CFB](#)

Types: *No types*

RDF rank: **0**

pub:hasValue  
1981-07-29

Slika 14 – Veza sa resursom koji sadrži informaciju o datumu rođenja

Relacije među klasama možemo videti Explore/Class relationships meniju. Tu možemo selektovati za koje klase želimo da vidimo veze. Prikazana je i informacija o broju ulaznih i/ili izlaznih linkova za klasu.

Class	Links
pub-old:TextMention	17K
pub-old:Document	6K
pub:Thing	5K
pub:Concept	2K
pub:Agent	762
pub:Location	592
pub:PopulatedPlace	523
pub:Organization	464
pub:Company	302
pub:Person	298
pub:TextRelation	188

Slika 15 – Relacije među klasama



## Reference

1. Ontotext GraphDB website, <https://graphdb.ontotext.com/documentation/standard/>
2. Elin K. Jacob (2003.), *Ontologies and the Semantic Web*
3. Intechopen website, <https://www.intechopen.com/books/cloud-computing-security-concepts-and-practice/semantic-web-and-interactive-knowledge-graphs-as-an-educational-technology>