Assignment 4 – Introduction to Semantic Systems

Teodor Chakarov – 12141198

We have been given the following data:

- 1000_keywords, 1000_links, 1000_movies_metadata as CSV files
- 1000_movies_metadata and 1000_credits as JSON files
- Start ttl file with the base structure of our ontology model

1. Portege software

In order to create the missing data properties and the object properties which are given in the Assignment document, I used the Protégé software.

As Classes I created the IMDBResource and Cast

As Data and class properties I had to create all of the given once which were in black color. Based on the information in the brackets, I had to decide whether that should be class or object property.

As Data propery:

- idFilm
- isAdultFilm
- homepage
- description
- keyword
- tagline
- revenue
- originalTitle
- hasSpokenLanguage
- hasProductionCountry
- hasCastCharacter
- imdbid
- url
- vote_average
- vote_count

As Class property:

- hasIMDBResource
- hasCast
- hasCastActor

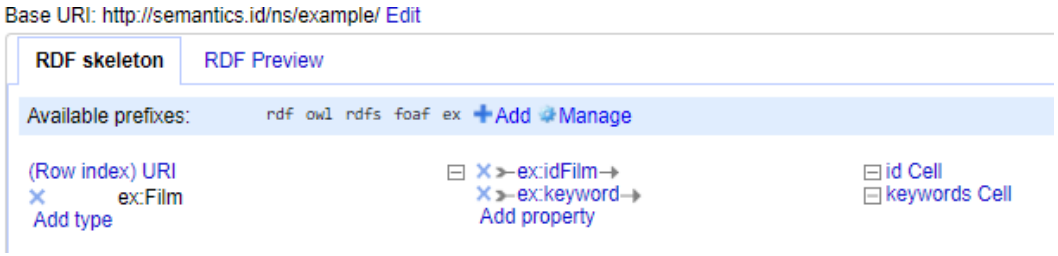
2. Open Refine

I used Open Refine to map the csv data with our ontology model. I created separate projects for each CSV file and I did my RDF skeletons as follows:

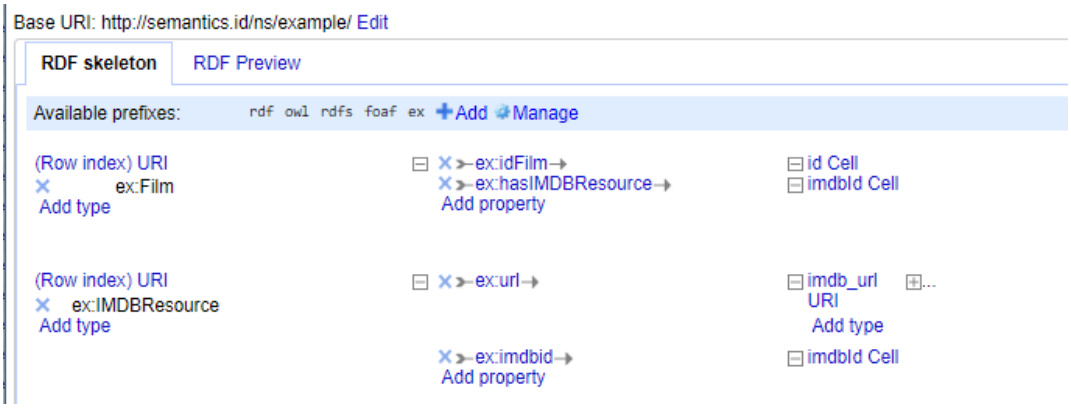
I had to add prefix as ex with the following url in order to get the recommendations

ex	http://semantics.id/ns/example/

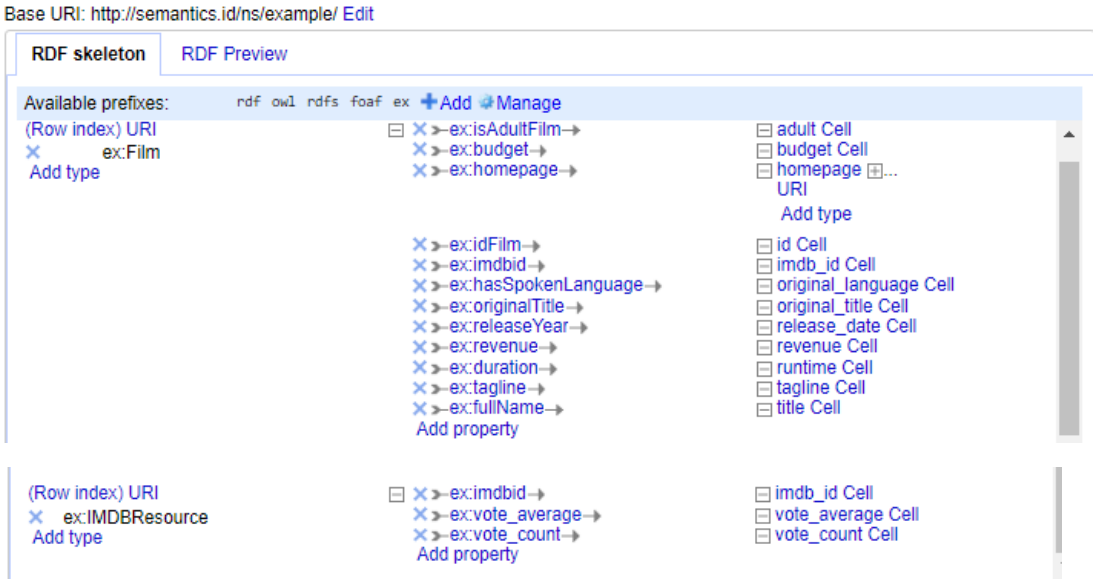
❖ 1000_keywords



❖ 1000_links



❖ 1000_movies_metadata



After I Mapped the data, for each project I extracted the ttl file as well as OpenRefine project as tar.gz file. You can find the data in folder **openRefineData**

3. RLF JSON data to RDF

I used https://github.com/carml/carml for extracting the data from the JSON file. I had to create for both JSON files a configuration file for the mappings of the keys form Jthe data and the ontology. In the CARML folder zou can find the two configuration files.

1000_credits is mapping the film ids and hasCast poperty. Then for the cast is mapping character which is string, hasCastActor which is class propery. And as final actor is mapped with the name of the class Actor and gender.

1000_movies_metadata is mapping the film original title as string, hasSpokenLanguage as sting, hasProductionCountry to production_countries… and hasGenre to genres

As output file I get the rml_ontology.ttl file

4. GraphDB

In order to work with the ontology I created, I used GraphDB. I imported the 4 ttl files that I exported from Open Refine for each of the CSVs, rml_ontology.ttl form CARML folder and also used the ontology ttl file.

I executed the following queries:

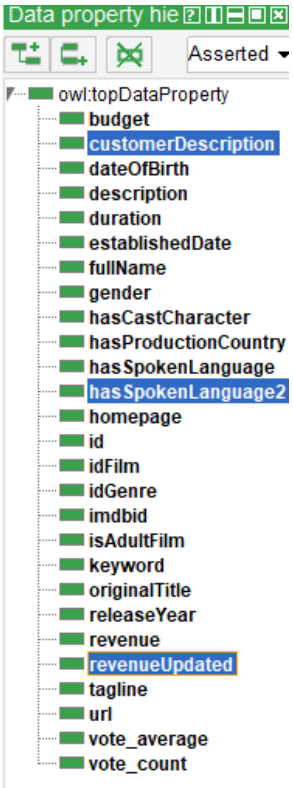
Q1- I created second data propery which is string and belongs to class Film. Its named hasSpokenLanguage2 and it is created to the movie with name Dracula. It stores the information about second language.

Q2 – Its creting revenueUpdatet data property. It is mapped to film which is 1h long and is 10 as integer.

Q3 – The last query creates the customerDescription data property. Which belongs to Film class and is string. It finds movie named Pulp Fiction and appends customer review on it.

After I execute the queries I exported the new graph model as ttl file, named final_ontology.ttl

When it is imported in Protégé we can see the new properties we created…




```
select * where {
        ?film a ex:Film .
  ?film ex:fullName "Pulp Fiction"
```

}

With reverse query we can see the film:31 gives us

:budget	"8000000"^^xsd:double
:customerDescription	"Best Movie in the history of Movies!"

After Q3