M164 CS2 Knowledge Technologies

Fall 2026-2027

Homework I

Out: November 2, 2026

Due: November 30, 2026 at 23:59. There will be no extensions to this deadline.

Total marks: 30%

Exercise 1 (Wikidata)

One of the large knowledge bases or ontologies we discussed in the introductory lecture is Wikidata (https://www.wikidata.org/wiki/Wikidata:Main_Page). In this exercise you will become familiar with Wikidata by examining its contents, posing SPARQL queries and adding information to it. More specifically, you have to do the following:

- Become familiar with Wikidata by browsing its web site. Take the tutorials and understand how Wikidata represents knowledge (i.e., its knowledge model). Maybe you would also like to read the following papers: https://cacm.acm.org/magazines/2014/10/178785-wikidata/fulltext and https://link.springer.com/content/pdf/10.1007%2F978-3-319-11964-9_4.pdf.
- Choose your favourite Greek entity (a football team, a mountain, a politician, a university etc.) and add information about it in Wikidata (the entity may already exist). You do not need to add lots of information although this is encouraged, but enough so that you appreciate how this is done. In your answer explain to us what you added to Wikidata and give us relevant pointers. If you enjoy this, become a Wikidata contributor!
- Read about the SPARQL query service of Wikidata. Then use this service to pose the following queries:
    - Find all the prime ministers of Greece known to Wikidata. Output their name, the party or parties they have been members of and the university (-ies) that they have graduated from (assuming that they have a university degree ☺).
    - Find all the Greek universities known to Wikidata. Output their name, the city that they are located in and the number of Greek authors that have graduated from them (order answers by this number).

Exercise 2 (Querying the Greek administrative geography dataset using SPARQL)

Our group has initiated the development of a linked open data portal of interest to Greece. In the context of this effort, we have developed an ontology and a corresponding dataset for the new administrative system of Greece known as the Kallikratis plan (http://en.wikipedia.org/wiki/Administrative_divisions_of_Greece). This exercise involves posing SPARQL 1.1 queries against this ontology and dataset.

First, we ask you to understand the Kallikratis ontology `gag-ontology.rdf` given at the Web page of the exercises for the course (http://cgi.di.uoa.gr/~pms509/). See also the brief documentation available there.

Then consider the dataset for Kallikratis given at the same Web page. Load the dataset in GraphDB and use SPARQL 1.1 to express the following queries:

- Give the official name and population of each municipality (δήμος) of Greece.
- For each municipality (δήμος) of Greece, give its official name, the official name of the regional unit (περιφερειακή ενότητα) it belongs to, and the official name of each municipal unit (δημοτική ενότητα) in it. Organize your answer by municipality.
- For each municipality of the region Crete with population more than 5,000 people, give its official name and its population.
- For each municipality of Crete for which we have no seat (έδρα) information in the dataset, give its official name.
- For each municipality of Crete, give its official name and all the administrative divisions of Greece that it belongs to according to Kallikratis. Your query should be the simplest one possible, and it should not use any explicit knowledge of how many levels of administration are imposed by Kallikratis.
- For each region of Greece, give its official name, how many regional units belong to it, the official name of each regional unit (περιφερειακή ενότητα) that belongs to it, and how many municipalities belong to that regional unit.
- Check the consistency of the dataset regarding stated populations: the sum of the populations of all administrative units A of level L must be equal to the population of the administrative unit B of level L+1 to which all administrative units A belong to. (You have to write one query only.)
- Give the decentralized administrations (αποκεντρωμένες διοικήσεις) of Greece that consist of more than two regional units. (You cannot use SPARQL 1.1 aggregate operators to express this query.)

Exercise 3 (http://schema.org)

As we have discussed in class, http://schema.org is a major effort from the top search engine companies (Google, Bing, Yahoo and Yandex) to help web designers annotate their pages with structured information which can then be used by search engines for better indexing of these web pages.

You can read about this effort at http://schema.org/ .

As you can see http://schema.org/ provides an ontology for annotating Web pages. This exercise asks you to write queries that navigate this ontology and are evaluated using RDFS reasoning. First browse the ontology starting from the page https://schema.org/docs/schemas.html. You should also read about the data model and other information about this ontology at http://schema.org/docs/documents.html. Then download the recent version of the core ontology https://schema.org/docs/developers.html as triples, store it in a GraphDB repository that supports inferencing, and use SPARQL 1.1 to express the following queries:

- Find all subclasses of class Place (note that http://schema.org/ prefers to use the equivalent term "type" for "class").
- Find all the superclasses of class Place.
- Find all properties defined for the class Place together with all the properties inherited from its superclasses.
- Find all classes that are subclasses of class Thing and are found in at most 2 levels of subclass relationships away from Thing.
- Finally, express the above queries on the ontology and dataset but without the use of inferencing.

Deliverables

For this homework, you will submit through e-class a zip file with the following:

- A pdf report (the first page should include your name and ID.) with the following content:
  - All the SPARQL queries and sample results for exercises 1, 2, 3 along with any documentation/remarks if needed.
  - For exercise 1 provide links for whatever you added to Wikidata.

- A txt file with the SPARQL queries for exercises 1, 2 and 3 separated by an empty line.