WSE HOMEWORK - SEMANTIC WEB

This first assignment must be solved within a week and needs to be handed in by **Friday, Oct 4**.

Your deliverable needs to be uploaded to **Brightspace** and is a **text document**, covering the aspects requested by each exercise section. Each task specifies what needs to be included in the deliverable.
Your deliverable will be graded on a 1..10 scale.

1 OPEN QUESTIONS

Deliverable: Briefly answer the following questions in your text document.

a) What is an "ontology" in computer science?
b) What is a "knowledge base"?
c) What are the core differences and similarities between SPARQL and SQL?
d) Why are "knowledge graph embeddings" useful?

2 WIKIDATA HANDS-ON

This exercise is designed to motivate you playing around and experiencing querying Linked Open Data Knowledge Bases like Wikidata using SPARQL. Please refer to additional web resources like references and tutorials to learn more about both Wikidata and SPARQL (the Wikidata SPARQL tutorial is a good start for this https://www.wikidata.org/wiki/Wikidata:SPARQL_tutorial, and https://www.w3.org/TR/sparql11-query/).

Also, read / quickread this: https://dl.acm.org/citation.cfm?doid=2661061.2629489

2.1 GETTING STARTED

Inspect the following two wikidata pages:
https://www.wikidata.org/wiki/Q221092
https://www.wikidata.org/wiki/Q36600

Use the Wikidata query service at https://query.wikidata.org/.

Run the following query:

```
#find the Mauristhuis.
SELECT ?item ?label WHERE {
  VALUES ?label {"Mauritshuis"@en}
  ?item rdfs:label ?label.
}
```

Hint: Most strings in Wikidata are multilanguage, thus the `@en`.
Think about: Did you receive the result you expected from the query? If no, what would you change?
Deliverable: Nothing. This is just a warm-up.

2.2 WHAT TO DO DURING THE WEEKEND

Now, to something more useful: Use the Wikidata query endpoint to plot a map of *all* museums in the Netherlands.

Hint: Most real-life locations (like museums) have the property `wdt:625` (labelled "`coordinate location`"`@en`) encoding geo locations. These can be used in the map plot view (which can either be manually shown or toggled using a comment directive). See example below:

```
#defaultView:Map
SELECT ?geo ?item WHERE {
   VALUES ?label {"Mauritshuis"@en}
   ?item rdfs:label ?label;
         wdt:P625 ?geo.
}
```

Secret Hint: Feel free to ignore secret hints if you want the challenge. You cursor select / copy&paste the text below (it's grey text on grey background…):

Deliverable: Provide your SPARQL statement. Also give a screenshot of your mapped results (please zoom in the map screenshot such that only Delft / Den Haag area is visible and readable).

2.3 SAFE MONEY ON TICKETS

Create a SPARQL Query retrieving the image, name, and name of artist for each painting in the Rijksmuseum Amsterdam. Order the result by artistname. Use the image grid visualization (`#defaultView:ImageGrid`).

Hint: Read https://www.wikidata.org/wiki/Wikidata:SPARQL_tutorial#Our_first_query how to easily work with English labels.

Deliverable: Provide your SPARQL statement. Also give a screenshot of the first few of your results.

2.4 WHO IS A BUSY PAINTER?

Modify your query above to create a table (not image grid) covering a statistic of name of painter, and the number of paintings they made which are in the Rijksmuseum, ordered by number of painting. Who is the most productive artist in the Rijksmuseum?

Deliverable: Provide your SPARQL statement. Also give a screenshot of the first dozen or so rows of your results. Also, how many Rembrandts does the Rijksmuseum have?

2.5 WHERE DID ALL THE VERMEERS GO?

One of the most famous Delft citizens was Johannes Vermeer (`wdt:Q41264`). Unfortunately, none of his works remain in Delft nowadays – find out where they are!

Hint: Wikidata has information on 37 paintings of Johannes Vermeer.

Please make the following Wikidata queries:

2.5.1 PAINTINGS AND COUNTRIES

Create a table with all paintings of Johannes Vermeer including their English *name, e.g.* "Girl with a Pearl Earring"*,* and the *country, e.g.* "Netherlands" they are in, ordered alphabetically by the paintings name .

Hint: Wikidata is not always as nice and clean as one would like, you will encounter some issues e.g., with "A Young Woman Seated at the Virginals" and "Woman Holding a Balance". Do one of the following
a) feel free to ignore the issue … (but think about why this issue exists)
b) fix the issue, so that you only have 37 rows in your table with the *latest known country*. This is the much more challenging solution for extra points or if you cannot stand dirty results. (this will keep you occupied for a while, I assume).
Secret Hint for b:
Ok, my solution involves (https://en.wikibooks.org/wiki/SPARQL/WIKIDATA_Qualifiers,_References_and_Ranks), several subqueries, aggregates, conditionals, binds, and optionals.
Deliverable: Provide your SPARQL statements. Also copy-paste your result table if it has less than 50 rows. Otherwise, copy the first 50 rows plus tell the number of total rows.

2.5.2 PAINTINGS AND CITIES

Create a table with all painting of Johannes Vermeer including their English *name,* and the *city* and *country* they are in right now.

Hint: Reuse your solution from the last exercise. Note that the some results like “Manhatten” (New York City would be the right one here) or “Mitte” are not really cities. Try fixing this as far as your patience allows…
Hint: “The Astronomer” is in Paris, “The Girl with the Wine Glass” in Brunswick, and “The Procuress” in Dresden.
Think about: Why is this seemingly simple query again so hard? What went wrong here? What should happen to fix the data properly?
Deliverable: Provide your SPARQL statement BUT MARK WHICH PARTS ARE NEW FOR DEALING WITH CITIES. Also copy-paste your result table if it has less than 50 rows. Otherwise, copy the first 50 rows plus tell the number of total rows.