Web Science & Engineering

Marios Marinos, Student number : 5353106

September 30, 2020

1 Open Questions

a) Firstly, the term ontology was introduced in philosophy, in the nineteenth century. Around the 1980s the term ontology was used from the AI community and eventually David Powers introduced the term of ontology to Computer Science on one of his papers around 1990. Now, ontology in Computer Science includes a representation, formal naming and definition of the categories, properties and relations between the concepts. Therefore, an ontology is a way of showing the properties of a subject area and how they are related, by defining a set of concepts and categories that represent the subject.

b) Knowledge base (KB) is a broader term than ontology. Similar to an ontology, a KB is represented in a knowledge representation (KR) formalism, which allows automatic inference. It could include multiple axioms, definitions, rules, facts, statements, and any other primitives. In contrast to ontologies, however, KB's are not intended to represent a shared or consensual conceptualisation. Thus, ontologies are a specific sort of a KB. Many KB's can be split into ontology and instance data parts, in a way analogous to the splitting of schemata and concrete data in databases.

Core differences between SPARQL and SQL.

c)
- SPARQL is a query language used to query RDF data stores. While SPARQL may initially look like SQL, you can observe that there are important differences because the data is graph-based so queries match graph patterns instead SQL's relational matching operations. So the syntax is similar but SPARQL queries graph data and SQL queries relational data in tables.
- In contrast to SQL, SPARQL queries are not constrained to working within one database: federated queries can access multiple data stores (endpoints). This is technically possible because SPARQL is more than just a query language. It is also an HTTP-based transport protocol, where any SPARQL endpoint can be accessed via a standardized transport layer. RDF results can be returned in several data-interchange formats and RDF entities are identified by Universal Resource Identifiers (URIs).

Core similarities between SPARQL and SQL.

- Both languages have many key words in common and also aggregate function names. In addition they have similar structure in their queries.
- SQL semantics revolve around joining tables together and then looking through every row to see if the contents of row fields meet specified conditions. If one thinks of a collection of triples containing the same predicate as a (distributed) table named by the triplet predicate and containing 2 columns, the triplet subject and object, then the "." operator in SparQL queries is similar to a join, in which shared SparQL variables within triple patterns essentially define a join condition specifying equality.

d) Firstly, Knowledge graph embeddings (KGE's) are low-dimensional representations of the entities and relations in a knowledge graph. They provide a generalizable context about the overall KG that can be used to infer relations. KGE models define different score functions that measure the distance of two entities relative to its relation type in the low-dimensional embedding space. These score functions are used to train the KGE models so that the entities connected by relations are close to each other while the entities that are not connected are far away. KGE can be really useful as they can be used for clustering, recommendations systems and also building knowledge bases.

2 WikiData Hands-on

2.1 Getting Started

No deliverable.

2.2 What to do during the weekend

The SPARQL statement I used is this :

```
#defaultView:Map
SELECT ?geo ?item WHERE {
    ?item wdt:P31/wdt:P279* wd:Q33506;
          wdt:P17 wd:Q55;
          wdt:P625 ?geo.
}
```

Figure 1: Delft Map

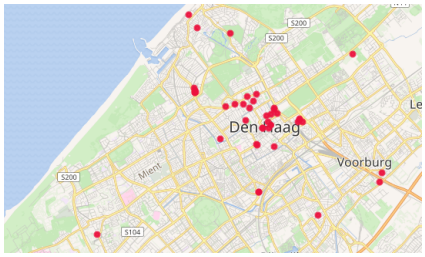

Figure 2: Den Haag Map

2.3 Safe money on tickets

The SPARQL statement I used is this :

```
#defaultView:ImageGrid
SELECT ?title ?image ?artist_nameLabel WHERE {
  ?painting wdt:P195 wd:Q190804;
            wdt:P31 wd:Q3305213;
            wdt:P18 ?image;
            wdt:P1476 ?title;
            wdt:P170 ?artist_name.
# filter by language to english in order to discard duplicates.
  FILTER(LANG(?title) = "en").
  SERVICE wikibase:label { bd:serviceParam
   wikibase:language "[AUTO_LANGUAGE],en". }
 } ORDER BY ?artist_nameLabel
```

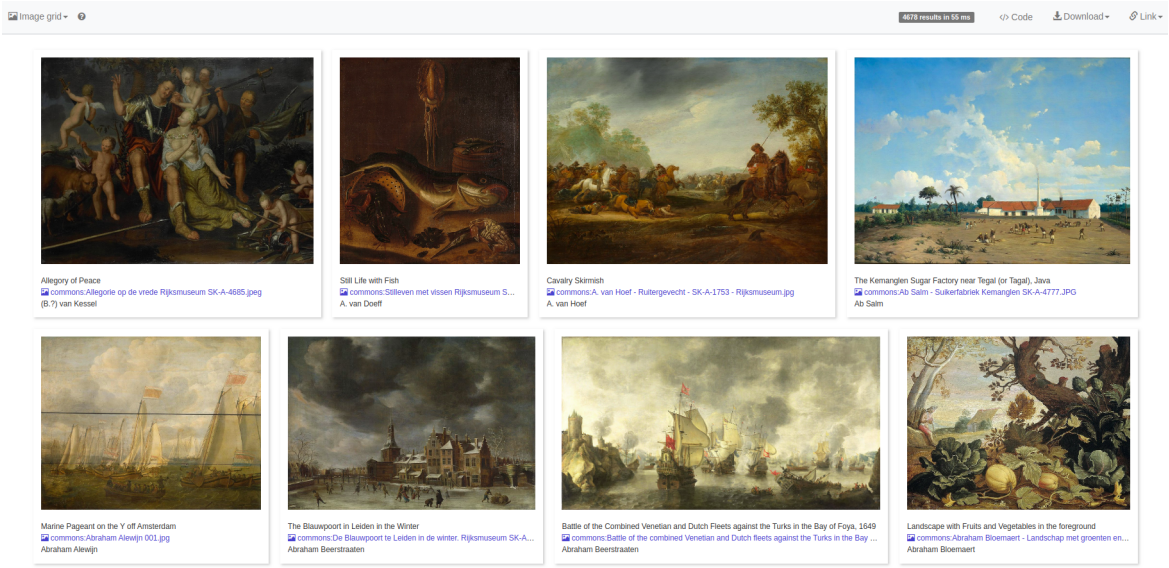

Figure 3: Results from above SPARQL code.

2.4 Who is a busy painter?

The SPARQL statement I used is this :

```
SELECT ?artist_name ?artist_nameLabel (COUNT(?painting) AS ?count)
WHERE
{
  ?painting wdt:P195 wd:Q190804;
            wdt:P31 wd:Q3305213;
            wdt:P18 ?image;
            wdt:P1476 ?title;
            wdt:P170 ?artist_name.
            FILTER(LANG(?title) = "en").
# filter by language to en in order to discard duplicates.
  SERVICE wikibase:label { bd:serviceParam
  wikibase:language "[AUTO_LANGUAGE],en". }
 }
GROUP BY ?artist_name ?artist_nameLabel
ORDER BY DESC(?count)
```

Search

artist_name	artist_nameLabel	count
wd:Q4233718	anonymous	526
wd:Q956812	Jean Baptiste Vanmour	61
wd:Q289441	George Hendrik Breitner	54
wd:Q4795396	Arnoud van Halen	45
wd:Q978158	Jacob Maris	44
wd:Q6149476	Jan Maurits Quinkhard	39
wd:Q864092	Michiel van Mierevelt	34
wd:Q768520	Pieter van der Werff	34
wd:Q864059	Jan van Ravesteyn	32
wd:Q5598	Rembrandt	32
wd:Q528460	Jozef Israëls	28
wd:Q217715	Caspar Netscher	27
wd:Q374039	Ferdinand Bol	25
wd:Q2627927	Cornelis Troost	25
wd:Q533958	Gerard de Lairesse	23
wd:Q205863	Jan Steen	23

Figure 4: Count of paintings in descending order.

From the above results we can conclude that there are plenty of paintings in Rijksmuseum Amsterdam for which we don't know the artist. So, the most productive artist is on 2^{nd} place and its Jean Baptiste Vanmour with 61 paintings as we can see. Finally, the Rembrandts Rijksmuseum Amsterdam owns are 32 (shown in Figre 4.)

2.5 Where did all the vermeers go?

The SPARQL statement I used is this :

```
SELECT ?paintingLabel ?locationLabel
WHERE{
   ?painting wdt:P170 wd:Q41264;
             wdt:P276 ?location.
   SERVICE wikibase:label { bd:serviceParam
   wikibase:language "[AUTO_LANGUAGE],en". }
  }
```

paintingLabel	locationLabel
Diana and Her Companions	Mauritshuis
The Geographer	Städel Museum
The Geographer	Villa San Donato
The Procuress	Gemäldegalerie Alte Meister
Saint Praxedis	National Museum of Western Art
Woman with a Pearl Necklace	Gemäldegalerie
Woman in Blue Reading a Letter	Rijksmuseum
The Milkmaid	Netherlands
The Milkmaid	Rijksmuseum
Girl with a Pearl Earring	Netherlands
Girl with a Pearl Earring	Mauritshuis

Figure 5: The first 11 paintings of Johannes Vermeer.

2.5.1 Paintings and countries.

```
  SELECT ?paintingLabel ?countryLabel ?max ?count
WHERE{
  {
      SELECT ?painting ?paintingLabel ?country (COUNT(?painting) as ?count) (MAX(?endtime)
          ?painting wdt:P170 wd:Q41264;
                    wdt:P276 ?location.
          ?painting p:P276 ?check.
          ?location wdt:P17 ?country
            OPTIONAL {?check pq:P582 ?endtime.}

        }GROUP BY ?painting ?paintingLabel ?country
    }FILTER(?count >= 1)

   SERVICE wikibase:label { bd:serviceParam
    wikibase:language "[AUTO_LANGUAGE],en". }
 }
  ORDER BY ?paintingLabel
```

Tried the above query and got this results. I know that is not correct of course, but it's my try.

paintingLabel	countryLabel	max	count
A Girl Asleep	United States of America		1
A Lady Writing a Letter	United States of America		1
A Young Woman Seated at the Virginals	Russia	13 January 2019	20
Christ in the House of Martha and Mary	United Kingdom		1
Diana and Her Companions	Netherlands		1
Girl Interrupted at her Music	United States of America		1
Girl Reading a Letter at an Open Window	Germany		1
Girl with a Flute	United States of America		1
Girl with a Pearl Earring	Kingdom of the Netherlands		2
Girl with a Pearl Earring	Netherlands		2

Figure 6: The first 11 paintings of Johannes Vermeer with country.