

Introduction to Data Science

Homework 4 Knowledge Bases - YAGO

20 points (+ up to 20 bonus points)

1. Consider YAGO Web interfaces that you may use to browse:
 - a. <https://yago-knowledge.org/sparql>
 - b. <https://yago-knowledge.org>
 - c. <https://yago-knowledge.org/schema>
2. Select two object *types* that you will be working with from the following list or other types you may find in YAGO:
 - Books, Cars, Movies, Location, Universities, Restaurants, Patents, Cricket players, Jobs, Different medical conditions (such as influenza, diabetes, other), Cartoon characters, Clothing brands/businesses, Oscar winning actors and actresses
3. Query YAGO to fetch as many objects of the type you selected.
 - a. One of the ways to query YAGO is by *specific properties* for an object type. For example, for *Songs* you may notice by browsing YAGO there are properties such as “lyricist”, “genre”, “composer”. You can use these properties and fetch *Songs*. It is advisable to browse objects from YAGO interface to understand the properties and then query.

One of the possible *SPARQL* queries to fetch *Songs* from YAGO may look like below. It uses a property “lyricist”.

```
PREFIX vcard:<http://schema.org/lyricist>
SELECT ?y
WHERE
{ ?y vcard: ?g }
```

However, notice this is just an example. It fetches just *Songs* having this property, which may or may not be a complete list.

Your task is to design a query that in your opinion provides as complete dataset as possible. Also, of course, please, do not select *Songs* as one of your objects.

- b. Execute your SPARQL query on YAGO using the query interface: <https://yago-knowledge.org/sparql>

For *Songs* query above, for example, it will return the following list of songs (shown a subset below):

	y
1	http://yago-knowledge.org/resource/Sabotage_(song)
2	http://yago-knowledge.org/resource/Animals_(Maroon_5_song)
3	http://yago-knowledge.org/resource/Svalutation_Q3506203
4	http://yago-knowledge.org/resource/Nie_spoczniemy_Q20972184
5	http://yago-knowledge.org/resource/Ballada_o_pancernych_Q9164651

...

However, again, this is only an example, there are many ways that one can use to query YAGO for objects.

4. (Bonus) The query above that you write fetches only the object names. Write a query(ies) to get as many attributes for your types of objects selected. For example, for Songs, a query (just for one specific song Miss_Sunshine) below will return the attributes that this song has.

Query:

```
PREFIX vcard:<http://yago-knowledge.org/resource/Miss_Sunshine>
SELECT *
WHERE
{ vcard: ?g ?y
}
```

Result set:

	g	y
1	http://schema.org/isPartOf	http://yago-knowledge.org/resource/Sunshine_(R.I.O._album)
2	http://schema.org/genre	http://yago-knowledge.org/resource/Electronic_dance_music
3	http://schema.org/datePublished	"2011-04-13" ^{^^xsd:date}

g

y

4 <http://schema.org/lyricist>

<http://yago-knowledge.org/resource/Yanou>

5 <http://schema.org/alternateName>

"Miss Sunshine (cantar de R.I.O.)"@ast

You can see this query uses just one specific song – Miss_Sunshine, so it may not return all possible attributes other songs may have. Your task is to find and fetch as many different attributes for the types of objects that you selected as possible.

What to turn-in

Part 1: Turn in a PDF file having the following items:

1. (20 points = 10 per object) SPARQL queries you wrote for two objects of your choice.
2. (Bonus, 10 points per object) SPARQL queries you wrote to get attributes for up to 2 objects of your choice (may be different or the same).
3. Result sets of your queries (compressed). Two CSV (text) files – one per object for the main part. Two CSV (text) files for the second (bonus) part. All result sets compressed with ZIP or another compressor.
4. Do not forget to write your name at the beginning of the file.

Good luck!