RWTH AACHEN UNIVERSITY

Big Data in Medical Informatics
Assignment #3

Due on January 19, 2017

Prof. Dr. Stefan Decker

Dr. Oya Deniz Beyan

Submitted by:

Sanchit Alekh

MSc. Software Systems Engineering

Matrikelnummer : 359831

Problem 1

Solution

The following prefixes apply to all of the queries listed below:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX dbr: <http://dbpedia.org/resource/>
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX mov: <http://www.movieontology.org/2009/10/01/movieontology.owl#>
PREFIX mve: <http://www.movieontology.org/2009/11/09/>
PREFIX dbp: <http://dbpedia.org/property/>
```

(a)

```
SELECT ?movie
WHERE {
?movie mov:hasDirector mov:Director1;
mov:isTranslatedTo mov:Spanish.
}
```

(b)

```
SELECT ?movie ?award
WHERE {
?movie mov:hasActor mov:Actor2.
OPTIONAL {?movie mov:isAwardedWith ?award}
}
```

(c)

```
SELECT ?movie
WHERE {
?movie mov:hasDirector mov:Director1;
MINUS {?movie mov:hasActor mov:Actor2.}
}
```

(d)

```
SELECT DISTINCT ?actor
WHERE {
?movie mov:belongsToGenre mov:Fun.
?movie mov:hasActor ?actor.
}
```

Problem 1 continued on next page...

(e)
```
SELECT ?someother
WHERE {
?movie mov:hasDirector mov:Director1.
?movie mov:hasActor ?actor.

?someother mov:hasActor ?actor.
FILTER NOT EXISTS {?someother mov:hasDirector mov:Director1.}
}
```

(f)
```
SELECT ?movie
WHERE {
{?movie mov:isAwardedWith mov:Oscar_Award.}
UNION
{?movie mov:isAwardedWith mov:Golden_Globe_Award.}
}
```

(g)
```
SELECT (AVG(?runtime) AS ?avgruntime)
WHERE {
?movie mov:runtime ?runtime.
}
```

(h)
```
SELECT ?genre (AVG(?runtime) AS ?avgruntime)
WHERE {
?movie mov:runtime ?runtime.
?movie mov:belongsToGenre ?genre.
}
GROUP BY ?genre
```

(i)
```
ASK
{
?movie mov:hasDirector mov:Director1.
?movie mov:hasActor mov:Actor3.
}
```

Problem 2

Solution

The following prefixes apply to all of the queries listed below:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX wp: <http://vocabularies.wikipathways.org/wp#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX openlinks: <http://www.openlinksw.com/schemas/virtrdf#>
PREFIX ctd_vocab: <http://bio2rdf.org/ctd_vocabulary:>
PREFIX omim: <http://bio2rdf.org/omim_vocabulary:>
PREFIX db: <http://bio2rdf.org/drugbank_vocabulary:>
```

- **Part 1:**

```
SELECT ?links ?label ?description
WHERE
{?a rdf:type ctd_vocab:Gene-Disease-Association.
 ?a ctd_vocab:gene <http://bio2rdf.org/ncbigene:675>.
 ?a ctd_vocab:disease ?links.
 ?d rdfs:label ?label.
 ?d omim:clinical-features ?description.
}
LIMIT 10
```

- **Part 3:**

```
SELECT ?drugName (COUNT(?d) as ?drugCount)
WHERE
{?a rdf:type db:Enzyme-Relation.
 ?a db:drug ?d.
 ?d rdfs:label ?drugName.}
```

```
GROUP BY ?drugName
ORDER BY DESC(?drugCount)
```

- **Part 6:**

```
SELECT ?spname (COUNT(?metabolite) AS ?count)
WHERE {
  ?metabolite rdf:type wp:Metabolite.
  ?metabolite dcterms:isPartOf ?species.
  ?species wp:organismName ?spname.
}
GROUP BY ?spname
ORDER BY DESC(?count)
```

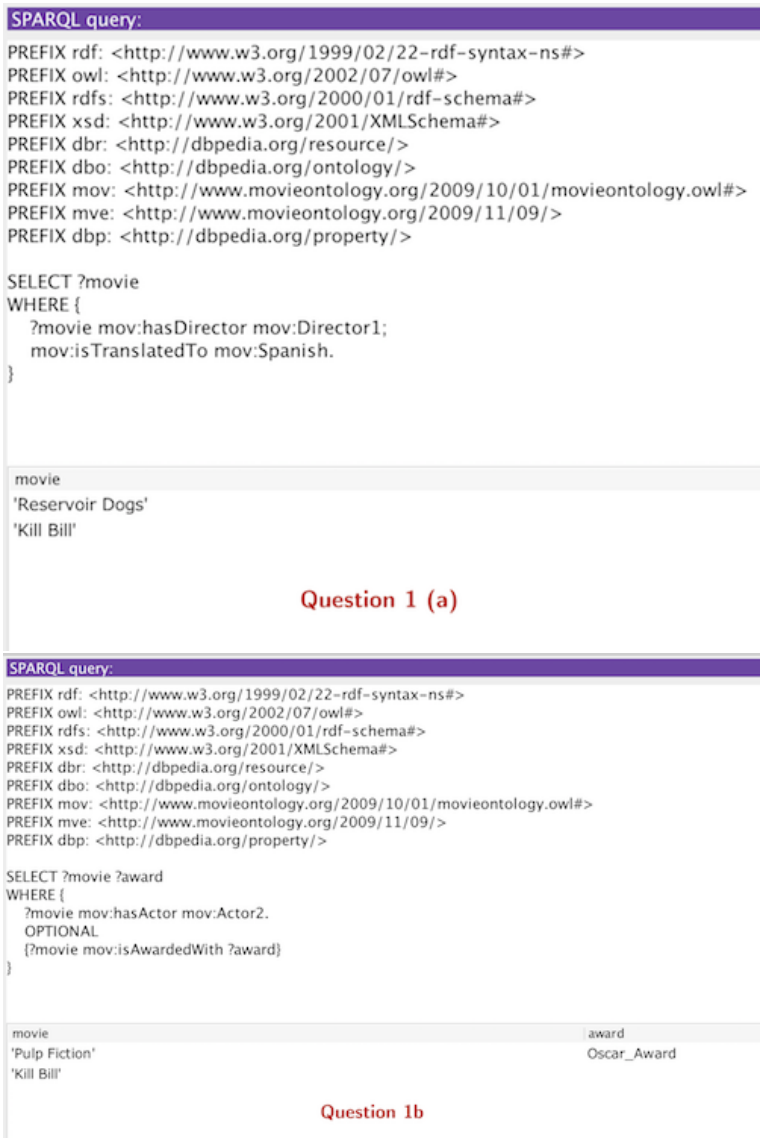
- **Part 9: (excluding Interaction)**

```
SELECT ?name ?complex ?pathway
WHERE {
  ?protein rdf:type wp:Protein.
  ?complex rdf:type wp:Complex.
  ?protein rdfs:label ?name.
  ?protein dcterms:isPartOf ?pathway.
  FILTER regex(?name, "^cyp", "i")
  FILTER NOT EXISTS{?pathway rdf:type wp:Interaction.}
}
```

- **Part 9: (including Interaction)**

```
SELECT COUNT(?name) as ?count
WHERE {
  ?protein rdf:type wp:Protein.
  ?complex rdf:type wp:Complex.
  ?protein rdfs:label ?name.
  ?protein dcterms:isPartOf ?pathway.
```

```
  FILTER regex(?name, "^cyp", "i")
}
```



SPARQL query:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX dbr: <http://dbpedia.org/resource/>
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX mov: <http://www.movieontology.org/2009/10/01/movieontology.owl#>
PREFIX mve: <http://www.movieontology.org/2009/11/09/>
PREFIX dbp: <http://dbpedia.org/property/>

SELECT ?movie
WHERE {
   ?movie mov:hasDirector mov:Director1;
   mov:isTranslatedTo mov:Spanish.
}
```

movie
'Reservoir Dogs'
'Kill Bill'

Question 1 (a)

SPARQL query:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX dbr: <http://dbpedia.org/resource/>
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX mov: <http://www.movieontology.org/2009/10/01/movieontology.owl#>
PREFIX mve: <http://www.movieontology.org/2009/11/09/>
PREFIX dbp: <http://dbpedia.org/property/>

SELECT ?movie ?award
WHERE {
   ?movie mov:hasActor mov:Actor2.
   OPTIONAL
   {?movie mov:isAwardedWith ?award}
}
```

movie	award
'Pulp Fiction'	Oscar_Award
'Kill Bill'	

Question 1b

Figure 1: Snapshots of the Query Responses - 1

SPARQL query:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX dbr: <http://dbpedia.org/resource/>
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX mov: <http://www.movieontology.org/2009/10/01/movieontology.owl#>
PREFIX mve: <http://www.movieontology.org/2009/11/09/>
PREFIX dbp: <http://dbpedia.org/property/>

SELECT ?movie
WHERE {
   ?movie mov:hasDirector mov:Director1;
   MINUS {?movie mov:hasActor mov:Actor2.}
}
```

movie
'Reservoir Dogs'

Question 1(c)

SPARQL query:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX dbr: <http://dbpedia.org/resource/>
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX mov: <http://www.movieontology.org/2009/10/01/movieontology.owl#>
PREFIX mve: <http://www.movieontology.org/2009/11/09/>
PREFIX dbp: <http://dbpedia.org/property/>

SELECT DISTINCT ?actor
WHERE {
   ?movie mov:belongsToGenre mov:Fun.
   ?movie mov:hasActor ?actor.
}
```

actor
Actor3
Actor1
Actor4

Question 1 (d)

Figure 2: Snapshots of the Query Responses - 2

SPARQL query:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX dbr: <http://dbpedia.org/resource/>
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX mov: <http://www.movieontology.org/2009/10/01/movieontology.owl#>
PREFIX mve: <http://www.movieontology.org/2009/11/09/>
PREFIX dbp: <http://dbpedia.org/property/>

SELECT ?someother
WHERE {
   ?movie mov:hasDirector mov:Director1.
   ?movie mov:hasActor ?actor.

   ?someother mov:hasActor ?actor.
   FILTER NOT EXISTS {?someother mov:hasDirector mov:Director1.}
}
```

someother
Hairspray

Question 1 (e)

SPARQL query:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX dbr: <http://dbpedia.org/resource/>
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX mov: <http://www.movieontology.org/2009/10/01/movieontology.owl#>
PREFIX mve: <http://www.movieontology.org/2009/11/09/>
PREFIX dbp: <http://dbpedia.org/property/>

SELECT ?movie
WHERE {
   {?movie mov:isAwardedWith mov:Oscar_Award.}
   UNION
   {?movie mov:isAwardedWith mov:Golden_Globe_Award.}
}
```

movie
'Pulp Fiction'
Hairspray
'The Fabulous Baker Boys'

Question 1 (f)

Figure 3: Snapshots of the Query Responses - 3

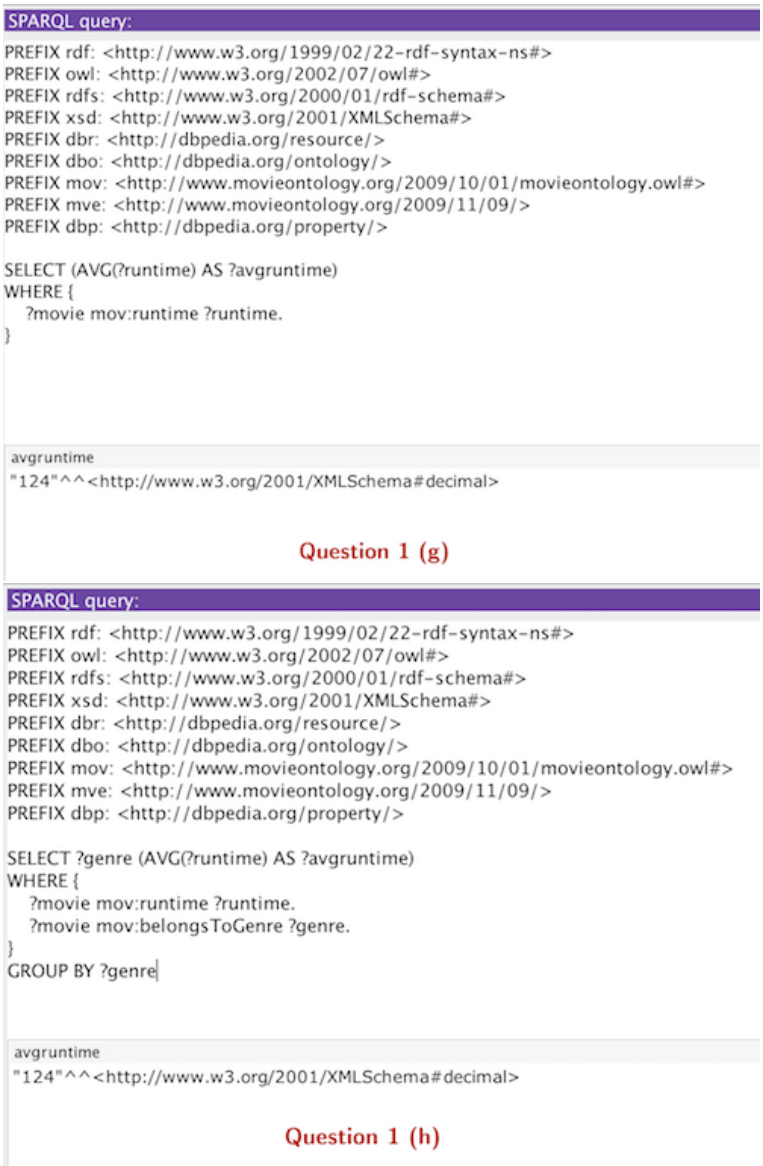


Figure 4: Snapshots of the Query Responses - 4

SPARQL query:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX dbr: <http://dbpedia.org/resource/>
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX mov: <http://www.movieontology.org/2009/10/01/movieontology.owl#>
PREFIX mve: <http://www.movieontology.org/2009/11/09/>
PREFIX dbp: <http://dbpedia.org/property/>

ASK
{
   ?movie mov:hasDirector mov:Director1.
   ?movie mov:hasActor mov:Actor3.
}
```

Result

False

Question 1 (i)

1	http://bio2rdf.org/omim:109730	"ANIRIDIA; AN [omim:106210]"@en

Question 2_1

et al. (2001)} to screen 14 subjects with aniridia for PAX6 mutations and to use MRI to look for alterations in brain structure. Olfactory capacity was also tested. Because PAX6 governs cellular proliferation and migration of 'later-born' neurons, and because subtle but significant malformation may be detectable only by quantitative MRI, they measured regional brain volumes. Interhemispheric communication is through 2 major pathways, the anterior commissure and the corpus callosum. To determine whether callosal hypoplasia, found in both homozygous small eye mice and the human compound heterozygote ({25:Glaser et al., 1994}), also occurs in human heterozygotes, {62:Sisodiya et al. (2001)} measured callosal area and found significant reduction in the study group compared with controls. Two subjects with hypoplastic olfactory bulbs had mild or moderate hyposmia. Of the remaining 12 subjects with visually normal olfactory bulbs, only 1 had normal olfaction. Some subjects were previously aware of notably reduced olfaction. The authors noted that anosmia had been reported anecdotally in an aniridia subject ({42:Martha et al., 1995}). Absence of the anterior commissure without callosal agenesis had not been reported as a malformative sequence in humans. In a similar study of 24 subjects with ocular abnormalities and PAX6 mutations, including the 14 patients reported by {62:Sisodiya et al. (2001)}, {44:Mitchell et al. (2003)} found absence of the pineal gland in 13 subjects and absence of the AC in 12. The authors noted that neither of these findings had been reported in Pax6 mutant mouse models. Because some of the sporadic cases of aniridia are caused by large chromosomal deletions, which may include the Wilms tumor gene ({607102}), such patients may have an increased risk of developing Wilms tumor (WT1; {607102}). Based on the unique registration of both cancer and aniridia cases in Denmark, {28:Gronskov et al. (2001)} were able to make an accurate risk estimate for Wilms tumor in sporadic aniridia. They found that patients with sporadic aniridia had a relative risk of 67 (CI, 8.1-241) of developing Wilms tumor. Among patients investigated for mutations, Wilms tumor developed in only 2 patients of 5

Figure 5: Snapshots of the Query Responses - 5

Query x +

http://drugbank.bio2rdf.org/sparql

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX openlinks: <http://www.openlinksw.com/schemas/virtrdf#>
PREFIX ctd_vocab: <http://bio2rdf.org/ctd_vocabulary:>
PREFIX omim: <http://bio2rdf.org/omim_vocabulary:>
PREFIX db: <http://bio2rdf.org/drugbank_vocabulary:>

SELECT ?drugName (COUNT(?d) as ?drugCount)
WHERE
{?a rdf:type db:Enzyme-Relation.
 ?a db:drug ?d.
 ?d rdfs:label ?drugName.}
GROUP BY ?drugName
ORDER BY DESC(?drugCount)
LIMIT 100
```

Question 2_3

Table | Raw Response | Pivot Table | Google Chart | Geo

Showing 1 to 100 of 100 entries (in 0.802 seconds)

	drugName	drugCount
1	"L-Glutamine [drugbank:DB00130]"@en	"19"^^xsd:integer
2	"Tamoxifen [drugbank:DB00675]"@en	"19"^^xsd:integer
3	"Troglitazone [drugbank:DB00197]"@en	"19"^^xsd:integer
4	"Valproic Acid [drugbank:DB00313]"@en	"18"^^xsd:integer

Query x +

http://sparql.wikipathways.org/

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX wp: <http://vocabularies.wikipathways.org/wp#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>

SELECT ?spname (COUNT(?metabolite) AS ?count)
WHERE {
  ?metabolite rdf:type wp:Metabolite.
  ?metabolite dcterms:isPartOf ?species.
  ?species wp:organismName ?spname.
}
GROUP BY ?spname
ORDER BY DESC(?count)
```

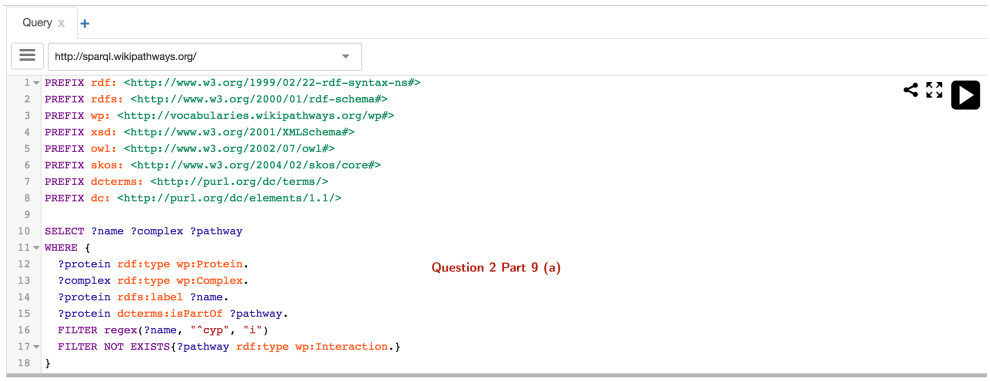
Table | Raw Response | Pivot Table | Google Chart | Geo

Showing 1 to 18 of 18 entries (in 0.11 seconds)

	spname	count
1	"Homo sapiens"^^xsd:string	"6951"^^xsd:integer
2	"Bos taurus"^^xsd:string	"1689"^^xsd:integer

Question 2 Part 6

Figure 6: Snapshots of the Query Responses - 6



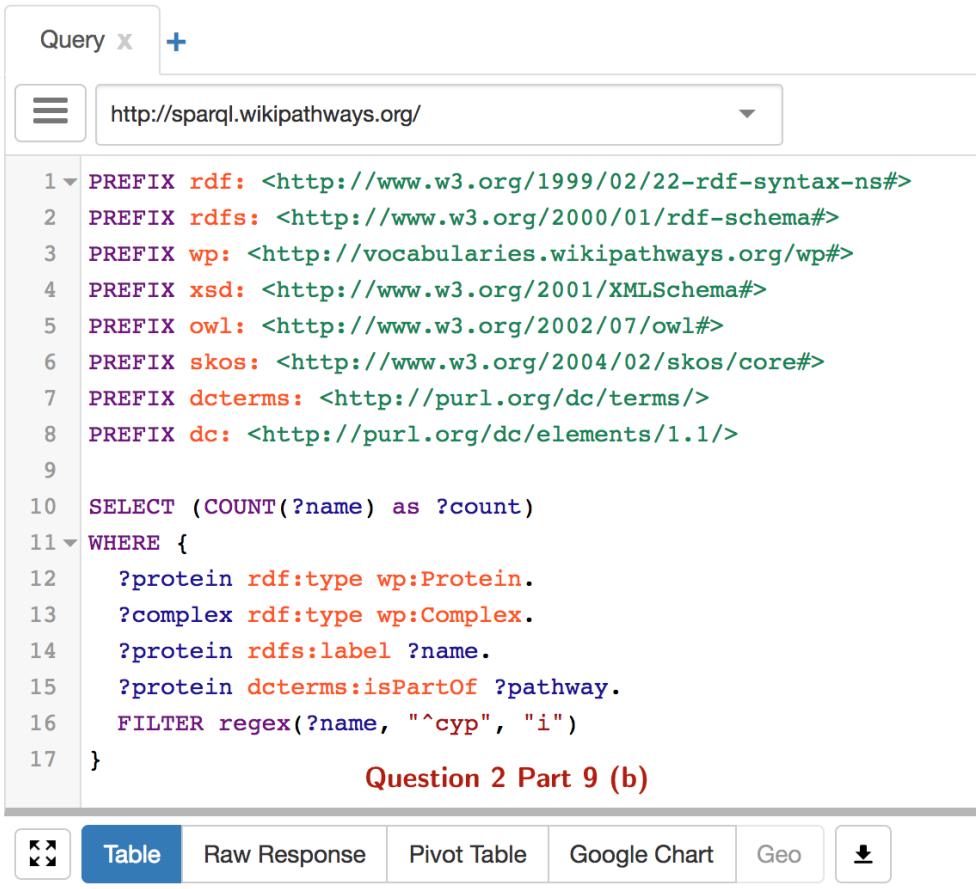


Figure 7: Snapshots of the Query Responses - 7