Assignment 5 SPARQL queries

UniProt SPARQL Endpoint: http://sparql.uniprot.org/sparql

Q1: 1 POINT How many protein records are in UniProt?

```
PREFIX up: <http://purl.uniprot.org/core/>

SELECT (COUNT (?numprotein) AS ?pcount)

WHERE
{
   ?numprotein a up:Protein .
}
```

Answer:

Q2: 1 POINT How many Arabidopsis thaliana protein records are in UniProt?

```
PREFIX up:<http://purl.uniprot.org/core/>
PREFIX taxon:<http://purl.uniprot.org/taxonomy/>

SELECT (COUNT(DISTINCT ?thalianaprotein) AS ?pthalianacount)
WHERE
{
   ?thalianaprotein a up:Protein .
   ?thalianaprotein up:organism taxon:3702 .
}
```

Answer:

Q3: 1 POINT retrieve pictures of Arabidopsis thaliana from UniProt?

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX up: <http://purl.uniprot.org/core/>

SELECT ?pictures

WHERE
{
  ?taxon a up:Taxon.
  ?taxon up:scientificName "Arabidopsis thaliana".
  ?taxon foaf:depiction ?pictures .

}
```

Answer:

Q4: 1 POINT: What is the description of the enzyme activity of UniProt Protein Q9SZZ8.

```
PREFIX uniprotkb:<http://purl.uniprot.org/uniprot/>
PREFIX skos:<http://www.w3.org/2004/02/skos/core#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX up:<http://purl.uniprot.org/core/>


SELECT ?enzyme_name ?enzyme_reaction ?activity
WHERE
{
   uniprotkb:Q9SZZ8 a up:Protein ;
            up:enzyme ?enzyme .

   ?enzyme skos:prefLabel ?enzyme_name .
   ?enzyme up:activity ?activity .
   ?activity rdfs:label ?enzyme_reaction .
}
```

Answer:

Results

Sparql XML | Sparql JSON | CSV | Show query | Share

enzyme_name	enzyme_reaction	activity
"Beta-carotene 3-hydroxylase"xsd:string	"Beta-carotene + 4 reduced ferredoxin [iron-sulfur] cluster + 2 H(+) + 2 O(2) = zeaxanthin + 4 oxidized ferredoxin [iron-sulfur] cluster + 2 H(2)O."xsd:string	http://purl.uniprot.org/enzyme/1.14.15.24#SIPF8A63F68B2741FFE

Q5: 1 POINT: Retrieve the proteins ids, and date of submission, for proteins that have been added to UniProt this year (HINT Google for "SPARQL FILTER by date").

```
PREFIX up:<http://purl.uniprot.org/core/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

SELECT ?prot_id ?date

WHERE
{
   ?protein a up:Protein .
   ?protein up:created ?date .
   BIND (SUBSTR(STR(?protein),33) AS ?prot_id) . #the prot_id starts in the 33 position
   FILTER (?date >= "2021-01-01"^^xsd:date) .

}
```

Answer:

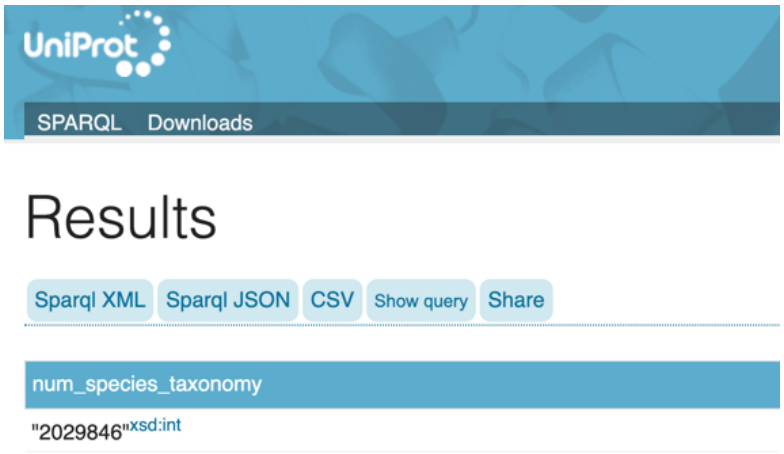
prot_id	date
"A0A1H7ADE3"xsd:string	"2021-06-02"xsd:date
"A0A1V1AIL4"xsd:string	"2021-06-02"xsd:date
"A0A2Z0L603"xsd:string	"2021-06-02"xsd:date
"A0A4J5GG53"xsd:string	"2021-04-07"xsd:date
"A0A6G8SU52"xsd:string	"2021-02-10"xsd:date
"A0A6G8SU69"xsd:string	"2021-02-10"xsd:date
"A0A7C9JLR7"xsd:string	"2021-02-10"xsd:date
"A0A7C9JMZ7"xsd:string	"2021-02-10"xsd:date
"A0A7C9KUQ4"xsd:string	"2021-02-10"xsd:date
"A0A7D4HP61"xsd:string	"2021-02-10"xsd:date
"A0A7D6A5N9"xsd:string	"2021-06-02"xsd:date
"A0A7D6FMY9"xsd:string	"2021-02-10"xsd:date
"A0A7D6VKU9"xsd:string	"2021-02-10"xsd:date
"A0A7D6VKZ9"xsd:string	"2021-02-10"xsd:date
"A0A7D7EJU1"xsd:string	"2021-02-10"xsd:date
"A0A7D7HYH9"xsd:string	"2021-02-10"xsd:date
"A0A7G5HK20"xsd:string	"2021-02-10"xsd:date
"A0A7G6B4J7"xsd:string	"2021-02-10"xsd:date
"A0A7G6T9F2"xsd:string	"2021-02-10"xsd:date
"A0A7G7EDL3"xsd:string	"2021-02-10"xsd:date
"A0A7G8TLN3"xsd:string	"2021-02-10"xsd:date
"A0A7H0XTK9"xsd:string	"2021-02-10"xsd:date
"A0A7H0ZDX6"xsd:string	"2021-02-10"xsd:date
"A0A7H1SVD2"xsd:string	"2021-02-10"xsd:date
"A0A7H4JPY6"xsd:string	"2021-02-10"xsd:date

Q6: 1 POINT How many species are in the UniProt taxonomy?

```
PREFIX up: <http://purl.uniprot.org/core/>

SELECT (COUNT (DISTINCT ?num) AS ?num_species_taxonomy)
WHERE
{
   ?num a up:Taxon .
   ?num up:rank up:Species .
}
```

Answer:



Q7: 2 POINT How many species have at least one protein record? (this might take a long time to execute, so do this one last!).

```
PREFIX up:<http://purl.uniprot.org/core/>

SELECT (COUNT(DISTINCT ?num) AS ?species_atl_oneprot)
WHERE
{
   ?protein a up:Protein .
   ?protein up:organism ?num .
   ?num a up:Taxon .
   ?num up:rank up:Species .
}
```

Answer:

Q8: 3 points: find the AGI codes and gene names for all Arabidopsis thaliana proteins that have a protein function annotation description that mentions "pattern formation".

```
PREFIX up: <http://purl.uniprot.org/core/>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>



SELECT DISTINCT ?AGI_name ?name

WHERE
{
   ?protein a up:Protein ;
         up:organism ?taxon_id ;
      up:encodedBy ?g ;
      up:annotation ?function_annot .
         ?taxon_id a up:Taxon ;
         up:scientificName "Arabidopsis thaliana" .

   ?g skos:prefLabel ?name .
   ?g up:locusName ?AGI_name .
   ?protein up:annotation ?annotation .
      ?annotation rdfs:comment ?f_annot .
      FILTER REGEX (?f_annot, "pattern formation", "i") .
}
```

Answer:

Results

Sparql XML | Sparql JSON | CSV | Show query | Share

AGI_name	name
"At3g54220"xsd:string	"SCR"xsd:string
"At1g13980"xsd:string	"GN"xsd:string
"At5g40260"xsd:string	"SWEET8"xsd:string
"At4g21750"xsd:string	"ATML1"xsd:string
"At1g69670"xsd:string	"CUL3B"xsd:string
"At1g63700"xsd:string	"YDA"xsd:string
"At2g46710"xsd:string	"ROPGAP3"xsd:string
"At1g26830"xsd:string	"CUL3A"xsd:string
"At1g55325"xsd:string	"MED13"xsd:string
"At3g09090"xsd:string	"DEX1"xsd:string
"At4g37650"xsd:string	"SHR"xsd:string
"At5g55250"xsd:string	"IAMT1"xsd:string
"At3g02130"xsd:string	"RPK2"xsd:string
"At2g42580"xsd:string	"TTL3"xsd:string
"At1g69270"xsd:string	"RPK1"xsd:string
"At5g02010"xsd:string	"ROPGEF7"xsd:string
"At1g66470"xsd:string	"RHD6"xsd:string
"At5g37800"xsd:string	"RSL1"xsd:string
"At1g49770"xsd:string	"BHLH95"xsd:string

SPARQL Endpoint: https://rdf.metanetx.org/sparql

Q9: 4 POINTS: what is the MetaNetX Reaction identifier (starts with "mnxr") for the UniProt Protein uniprotkb:Q18A79.

```
PREFIX mtnx: <https://rdf.metanetx.org/schema/>
PREFIX up: <http://purl.uniprot.org/uniprot/>

SELECT DISTINCT ?id

WHERE{
   ?pept mtnx:peptXref up:Q18A79 .
   ?cata a mtnx:CATA ;
         mtnx:pept ?pept .
   ?gpr mtnx:cata ?cata ;
        mtnx:reac ?reac .
   ?reac a mtnx:REAC ;
         mtnx:mnxr ?mnxr .
   ?mnxr rdfs:label ?id .
}
```

Answer:

id
"MNXR145046"
"MNXR165934"

FEDERATED QUERY - UniProt and MetaNetX
Q10: 5 POINTS: What is the official Gene ID (UniProt calls this a "mnemonic") and the MetaNetX Reaction identifier (mnxr…..) for the protein that has "Starch synthase" catalytic activity in Clostridium difficile (taxon 272563).

```
PREFIX mnx: <https://rdf.metanetx.org/schema/>
PREFIX uniprotkb: <http://purl.uniprot.org/uniprot/>
PREFIX up: <http://purl.uniprot.org/core/>
PREFIX taxon: <http://purl.uniprot.org/taxonomy/>

SELECT DISTINCT ?ID ?MNXID ?activity
WHERE
{
 service <http://sparql.uniprot.org/sparql> {
   ?protein a up:Protein ;
            up:organism taxon:272563 ;
            up:mnemonic ?ID ;
            up:classifiedWith ?GO .
   ?GO rdfs:label ?activity .
   filter contains(?activity, "starch synthase")
   bind (substr(str(?protein),33) as ?prot_ac)
   bind (IRI(CONCAT(uniprotkb:,?prot_ac)) as ?uniprotRef)
 }
 service <https://rdf.metanetx.org/sparql> {
   ?pept mnx:peptXref ?uniprotRef .
   ?cata mnx:pept ?pept .
   ?gpr mnx:cata ?cata ;
        mnx:reac ?reac .
   ?reac rdfs:label ?MNXID .
 }
}
```

Answer:

Results

Sparql XML | Sparql JSON | CSV | Show query | Share

ID	MNXID	activity
"GLGA_CLOD6"xsd:string	"mnxr165934"xsd:string	"starch synthase activity"xsd:string
"GLGA_CLOD6"xsd:string	"mnxr145046c3"xsd:string	"starch synthase activity"xsd:string