Assignment 5 SPARQL queries

I would like you to create the SPARQL query that will answer each of these questions.  Please submit the queries simply as a text document (NO programming is required!) - submit to GitHub as usual.

*For many of these you will need to look-up how to use the SPARQL functions ‘COUNT’ and ‘DISTINCT’ (we used ‘distinct’ in class), and probably a few others...*

-----------------------------------------

**UniProt SPARQL Endpoint:** [**http://sparql.uniprot.org/sparql**](http://sparql.uniprot.org/sparql)

**Prefixes:**

PREFIX uniprotkb: <http://purl.uniprot.org/uniprot/>

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

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

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

PREFIX label: <http://www.w3.org/2004/02/skos/core#>

PREFIX xsd: <[http://www.w3.org/2001/XMLSchema#](http://www.w3.org/2001/XMLSchema)>

PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

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

SELECT DISTINCT (count(?protein) as ?count)

WHERE {

?protein a up:Protein .

}

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

SELECT DISTINCT (count(?protein) as ?count)

WHERE {

?protein a up:Protein .

?protein up:organism ?taxon .

?taxon a up:Taxon .

?taxon up:scientificName 'Arabidopsis thaliana' .

}

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

SELECT DISTINCT ?image

WHERE {

?taxon <http://xmlns.com/foaf/0.1/depiction> ?image .

?taxon up:scientificName 'Arabidopsis thaliana' .

}

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

SELECT ?description

WHERE {

uniprotkb:Q9SZZ8 up:enzyme ?enzyme .

?enzyme up:activity ?activity .

?activity rdfs:label ?description .

}

**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”)

SELECT DISTINCT ?protein ?date

WHERE

{

?protein a up:Protein .

?protein up:created ?date . FILTER( ?date >= '2020-01-01'^^xsd:date)

}

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

SELECT DISTINCT (count(?taxon) as ?count)

WHERE

{

?taxon a up:Taxon .

}

**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!)

SELECT (count(DISTINCT ?taxon) as ?count)

WHERE {

?protein a up:Protein .

?protein up:organism ?taxon .

?taxon a up:Taxon .

}

**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”

K  
SELECT DISTINCT ?name ?locus

WHERE

{

?taxon up:scientificName 'Arabidopsis thaliana' .

?protein up:organism ?taxon .

?protein a up:Protein .

?protein up:annotation ?annotation .

?annotation a up:Function\_Annotation .

?annotation rdfs:comment ?text .

FILTER CONTAINS (?text, 'pattern formation') .

?protein up:encodedBy ?gene .

?gene skos:prefLabel ?name .

?gene up:locusName ?locus .

}

**From the MetaNetX metabolic networks for metagenomics database SPARQL Endpoint: https://rdf.metanetx.org/sparql**

**Prefixes**:

PREFIX mnx: <https://rdf.metanetx.org/schema/>

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

PREFIX uniprotkb: <http://purl.uniprot.org/uniprot/>

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

SELECT DISTINCT ?reac\_label

WHERE{

?pept mnx:peptXref uniprotkb:Q18A79 .

?cata mnx:pept ?pept .

?gpr mnx:cata ?cata .

?gpr mnx:reac ?reac.

?reac rdfs:label ?reac\_label .

}

**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).

SELECT DISTINCT ?mnemonic ?reac\_label

WHERE {

?protein up:organism taxon:272563 .

?protein a up:Protein .

?protein up:classifiedWith ?goTerm .

?goTerm rdfs:label ?label .

FILTER regex (?label, 'starch synthase', 'i') .

?protein up:mnemonic ?mnemonic .

SERVICE <https://rdf.metanetx.org/sparql> {

?pept mnx:peptXref ?protein .

?cata mnx:pept ?pept .

?gpr mnx:cata ?cata .

?gpr mnx:reac ?reac .

?reac rdfs:label ?reac\_label .

}

}