**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 .

}

**Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente**

**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:**

Interfaz de usuario gráfica, Texto, Aplicación, Sitio web

Descripción generada automáticamente

**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 .

}

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

**Answer:**

Un florero con flores

Descripción generada automáticamente

**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 .

}

**Interfaz de usuario gráfica, Texto, Aplicación, Sitio web

Descripción generada automáticamenteAnswer:**

**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) .

}

**Tabla

Descripción generada automáticamenteAnswer:**

**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:**

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

**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 .

}

Interfaz de usuario gráfica

Descripción generada automáticamente con confianza baja**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") .

}

Tabla

Descripción generada automáticamente**Answer:**

**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:**

Texto

Descripción generada automáticamente

**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 .

}

}

**Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamenteAnswer:**