Assignment 5 SPARQL queries

I would like you to create the SPARQL query that will answer each of these questions. Please submit the queries as a Jupyter notebook with the SPARQL kernel activated. NO programming is required! Submit to GitHub as usual, WITH THE ANSWERS STILL VISIBLE IN THE NOTEBOOK. Thanks!

*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**

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

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

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

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

**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”)

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

**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!)

**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”

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

**(this slide deck will make it much easier for you!** [**https://www.metanetx.org/cgi-bin/mnxget/mnxref/MetaNetX\_RDF\_schema.pdf**](https://www.metanetx.org/cgi-bin/mnxget/mnxref/MetaNetX_RDF_schema.pdf)**)**

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

**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).