In this project, we need to query UniProt using MapReduce and Mango Query

1. How many proteins are listed without genes?

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

7 proteins are listed without genes. (Not sure why it is not showing the number instead of the list, even though I set the Reduce to “count”)

1. Number of proteins with “proteinExistence” having “Predicted” values.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A close-up of a text

Description automatically generated

There are 586 proteins with “proteinExistence” having “Predicted” values.

1. List of all proteins that have references.

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

A screenshot of a computer code

Description automatically generated

If a protein has references, then it must come with a citation and a title. Therefore, I filtered the database with existing citation and titles. It turns out that 1224 proteins have references.

4. Evidence codes are used for each protein submitted name. They are ontology codes that help explain the provenance of the evidence. List of evidence codes used by UniProt.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

The evidence code used by uniport is ECO:0000313 (all 699 entries have the same evidence code).

1. List the genes of each protein.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

6. Implement some interesting CouchDB \_find selectors on the UniProt database…

6.1. Give a “Selector” statement to list all proteins with “proteinExistence” having “Predicted” values.

A screen shot of a computer code

Description automatically generated

A screenshot of a computer

Description automatically generated

6.2. List only proteins that have genes listed.

A screenshot of a computer

Description automatically generated

6.3. List the proteins and their associated genes of all proteins with “RecommendedName”:

A screenshot of a computer

Description automatically generated

6.4. List of proteins that are associated with a gene name using “regex” operator, e.g., starting with “H”

A screenshot of a computer

Description automatically generated

6.5. Give a query example to implement an “AND” condition.

A screenshot of a computer

Description automatically generated

6.6. Give a query example to implement an “OR” condition.

A screen shot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

6.7. Give a query example to implement an operator match (anything other than equality).

A screenshot of a computer

Description automatically generated