Georgetown University
Medical Center
Dept. of Biochemistry and Molecular & Cellular Biology
BINF-6970-01: Databased for Bioinformatics
Spring 2024
Project III

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

1. How many proteins are listed without genes?
2. Number of proteins with "proteinExistence" having "Predicted" values.
3. List of all proteins that 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.
5. List the genes of each protein.
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.
   6.2. List only proteins that have genes listed.
   6.3. List the proteins and their associated genes of all proteins with "RecommendedName":
   6.4. List of proteins that are associated with a gene name using "regex" operator, e.g., starting with "H"
   6.5. Give a query example to implement an "AND" condition.
   6.6. Give a query example to implement an "OR" condition.
   6.7. Give a query example to implement an operator match (anything other than equality).