

Mini-Rosetta Project Milestone 2

Due: Saturday, April 14

There are two programs for milestone 2. The first program uses JDBC to do bulk loading of proteins. The second program uses JDBC to retrieve proteins, compute matches and save the results into the database.

The first program is provided to get your familiar with JDBC programming. You are expected to do the second one by yourselves. You should turn in the screen shots of successfully executing your program.

Program 1

Postgres.java, ProteinLoader.java

Data: bacil-gb.txt, ecoli-gb.txt, ecoli-sp.txt, human-gb.txt, yeast-gb.txt

To import these data, you first need to delete existing records from the database.

Note: You may need to make some changes using provided code to fit with your own postgres settings/table name/column name etc. It is necessary that you look through and understand the code.

Program 2

You will

* Use JDBC to retrieve proteins from the database.

Use Smith-Waterman algorithm to

- compute matches for each pair of proteins.
- store the final alignment score for the pair
- store the computed alignment (equivalent to storing “chained rectangles”
- Note: you should use the Blosum62 matrix to compute similarity scores.