



Term Project

General rules for all projects

Teams:

You are to work in teams of two or three students. Each team will work on a project. Each team is to prepare a final report. All members of the team are expected to participate equally in all aspects of the project. All members of the team will receive the same grade on the project.

Project Requirements:

The team chooses one programming project. Alternatively, the team can propose their own project based on their own research interests, with the approval from the instructor.

Important Date:

Thursday, May 8: Printed copies of the final project are due at the beginning of the lecture. Do not forget to include all pertinent documents.

Submission Requirements:

Please include references for any material you have used, including source code.

A) Submit a hard copy of the project containing all of the following:

- 1) Title page.
- 2) Table of contents (with page numbers).
- 3) An essay of not more than 6, 1 1/2 spaced pages (font size: 11 or 12 points) describing the problem, the overall organization, design of your program. The essay should include a detailed analysis of your results and comparison with the findings presented in the original article. Use MS Excel or a similar package to generate tables and graphs. The essay should give the user an overall roadmap of your code, and would be read by a maintenance programmer before he/she began reading your code. Please do not include definitions, explanations of topics we have covered in the course and do not simply copy the original article. Include a flowchart, or UML or a structure chart to show the design of your program. Do not forget to number pages!
- 4) A one-page description of the test data including accession numbers. You must test your program with the data sets described in the original articles. You may use additional data sets if you wish.
- 5) Include one sample output of your program. Make sure that your output is readable and well formatted.

6) Instructions for running your program, in other words, explain how to compile, and execute your program.

7) A list of references.

B) Submit a CD, labeled with your names, course number and semester, and containing:

- a) The source code (fully documented)
- b) The input files
- c) The document specified in part A

Please make sure that all the files on the CD are readable.

Programming Projects:

A) Building a profile HMM

The profile HMM should

- Be built from at least 20 sequences
- Contain at least 10 match states.

B) Gene prediction algorithm

In this course, we cover different methods for predicting genes. The textbook has additional methods.

Choose an algorithm for gene prediction, and familiarize yourself with all its details before starting your own implementation. Also, explore publicly available packages for gene prediction and decide upon one that will be used in your project (for comparison with your own program). Implement and test your program with different sequences from public databases. You will then compare your results with those obtained from using a publicly available package.

C) Your own project

Submit a proposal by March 18, 2008.