

CSC2417 Project Description

October 4, 2016

Rather than having a final exam this course will have a term project. You have two options for the project - you can either implement some (fairly complex) algorithm or you can write a report about some area of computational biology (see Guidelines below). In either case the exact topic is up to you, as long as it is broadly related to the content of this course. You are encouraged to mix techniques across fields (this is very much in the spirit of computational biology) - for example if your background is in machine learning you might choose to apply some ML technique to a computational biology problem or if your background is in software engineering you might choose to write an optimized implementation of a classic algorithm.

The project has two deliverables. First you will provide a short project proposal (1-2 pages, due **October 25th**) describing the topic you selected and how you will approach it. The second deliverable is the full project (**due December 14th**) as laid out in the guidelines below.

Guidelines

Here are the requirements for the project depending on whether you choose to implement an algorithm or write a longer report. In either case your report should be fully referenced with citations to the appropriate papers for the ideas, algorithms and datasets that you used for your project.

Algorithm option

In this option you will implement some algorithm from either the literature or of your own design. You will test your implementation on a publicly available

data set (eg. some open access sequencing data) and write a 5 page report describing:

- The problem you will solve
- The algorithm you implemented
- A description of the data set you tested your algorithm on, and how you evaluated the results
- A critical analysis of the strengths and weaknesses of your approach in comparison to published work

Your implementation can be in the language of your choice but please check with me first if you are going to use a less popular language (Python, C/C++, Java are all OK, check with me for others).

Report option

In this option you will write a journal-style article that discusses some area of computational biology in depth. Your report should include:

- A description of the problem your report focuses on
- A discussion of the various approaches people have tried to solve this problem
- A critical analysis of strengths and weaknesses of each approach and suggestions of future directions to explore

This report should be around 10 pages in length.

Grading

The proposal is worth 20% of the project and the final software/report is worth the remaining 80%. For the algorithm option the choice of algorithm, the quality of your implementation, how you tested your software and your analysis of its strengths and weaknesses will be assessed. The report option will be assessed based on how well you analyzed the strengths and weaknesses of the various approaches and in particular the depth of your analysis (eg. did you consider all relevant aspects of the problem).

Teamwork

If you would like you can work with one other person on your project, however I expect the project to have a greater scope/depth when completed in pairs. If you work with someone else I expect the report to include a paragraph describing the contributions of each person (who implemented each part of the algorithm, etc).

Example projects

Here are some example projects that might help you generate ideas:

Algorithm option examples

- Implement a bloom filter-based representation of a de Bruijn graph
- Develop a simulator of Oxford Nanopore sequencing data with a realistic error model
- Write an optimized implementation of the “partial order alignment” method for computing a multiple alignment between long reads

Report option examples

- Compare the indexing strategies of aligners used for short sequencing reads
- Compare the de Bruijn graph and overlap graph methods of genome assembly
- Discuss fast methods of calculating species abundance from metagenomic datasets

If you are unsure of a project or would like to discuss ideas please feel free to email me and I will help.