

# CSE/BioE 408 Graduate Paper

Deadline: May 8, 12pm

One of the most important parts of graduate research is to learn what is the current state of research in a given field. This is challenging, because it requires a researcher to read and organize a large amount of information in order to understand the history of the field and the importance of the contributions that have been made. Without understanding these two general topics, it is easy to start research projects on topics that have already been addressed, or on topics that would be too tangentially related to the field to gather interest. Avoiding these issues is essential for success as a researcher.

To practice this skill, the 408 graduate paper will assign the task of learning about a subfield of genomics. The task is substantially abridged from the work that would actually need to be done to get up to speed in a new field, but it also keeps the work within the scope of the course.

Instructions.

- 1) Select a topic in from the course:
  - a) Genome Assembly
  - b) Nucleotide Sequence Alignment (including multiple sequence alignment)
  - c) Constructing phylogenetic trees
  - d) Gene Finding
- 2) Beginning with references from the textbook and the lecture slides, and using google scholar, identify exactly 5 additional papers that present a method designed to accomplish the same goal as the methods described in class.
  - a) A method accomplishes the same goal as the method in a starting point paper if the inputs and the outputs of one method are the same as in the other.
  - b) The other methods may improve on the methods described in class, or they may be predecessors of the methods in class.
- 3) Read the papers in detail.

Writing the paper:

- 1) Format:
  - a) 1 cover page + 10 full pages double spaced + 1 reference page = 12 total pages
  - b) Cover page to contain title, name, and email of student
  - c) 12 point font times new roman. Section headings should be no more than 14 point, with the same inter-line spacing.
  - d) 1 inch margins
  - e) No title on the second page.

- f) References for the five papers found on the reference page, in AMA bibliography format.
- 2) Summary section:
  - a) Answer the question: what is the general purpose of algorithms in this subfield?
  - b) What biological question are they trying to answer?
- 3) Five sections about each of the five papers you selected. In each section,
  - a) Describe how the method works in a technically detailed way. A person should be able to code the method from your description alone.
  - b) Describe what makes the method different from the other methods you chose
  - c) Describe why that difference is superior from a biological or computational perspective. Why should that paper have been published?
  - d) Describe the experiments run by the authors of the paper to prove that the method they created actually works and achieves the goal of the paper.
- 4) Conclusion section:
  - a) Describe what the methods all had in common
  - b) Describe one thing about each method that made them different from the rest
  - c) Explain why that differentiating factor was biologically important to measure

#### Grading:

10% Summary section clearly describes the general purpose of algorithms in the subfield.  
(Summary a)

10% Summary section clearly describes the biological question they are trying to answer.  
(Summary b)

50% (5x10%) each section summarizing each paper, covering points a-d.

30% (3x10%) each of the conclusion section components a-c.

#### Submitting the 408 paper

Make a zip file that includes (a) your paper as a PDF, and (b) all five papers that you read as PDFs. Submit this zip file using Lehigh FileSender. The file name of the zip file should include your name and "CSE408Project". Reports submitted without all five PDFs will not be graded.