Bioinformatics: Issues and Algorithms
CSE 397 / 497

Spring Semester 2005
Tuesdays and Thursdays, 2:35 pm – 3:50 pm
Room PL 208

Professor Daniel Lopresti
Department of Computer Science and Engineering
Webpage: http://www.cse.lehigh.edu/~lopresti/
Email: dal9@lehigh.edu

1. Course Description

The field of bioinformatics is becoming one of the most exciting application areas for techniques from computer science, as well as a major driver of new research in computational methods. The past decade has witnessed an explosion in genetic data, with entire genomes for many organisms (including humans) now available online. Computer science plays a central role in the creation, analysis, and management of this data: from the sequencing and assembly of DNA, to searching vast databases of known nucleic acid or protein sequences, to the calculation of evolutionary trees, to attempts to predict the structure, and ultimately the function, of these macromolecules.

CSE 397 / 497 will cover algorithms and computational models applied to bioinformatics. The course will be run in a seminar format. After a set of introductory lectures by the instructor, students will deliver the remaining presentations, prepared with help from the instructor, and take responsibility for leading discussions. Assigned readings will be drawn from a textbook and the current literature.

(Note: the registrar may still be listing this course as "Computational Issues in Molecular Biology." The two titles refer to the same course.)

2. Target Audience / Prerequisites

This course is targeted at two audiences:

* computer science and bioengineering students who wish to learn about the field of bioinformatics and the ways in which computational methods are applied there,

* biology students who would like to understand the algorithmic techniques that lie behind the software tools they use.

Recognizing these distinct backgrounds, early lectures will be designed to bridge the two disciplines and provide a common basis for later discussions. Because of the seminar nature the course, there is substantial flexibility to tailor the material to the students enrolled.

This course is intended primarily for seniors and graduate students. All students should have had some previous programming experience (including data structures) at at least the level of CSE 17 and a college-level understanding of discrete math, probability, and statistics. An advanced course in algorithms is not required.

Computer science students should note that no prior background in biology is assumed. For questions concerning prerequisites, contact the instructor.

3. Course Requirements

All students in the course will be required to present a lecture on a topic of their choice selected from the syllabus. As preparing a high quality lecture takes considerable time and effort, students will choose their topics well in advance and work closely with the instructor on content and delivery. Several one-on-one meetings will be scheduled prior to the student's lecture date.

Beyond this class participation component, there will be a final project, either a research paper or software implementation of one of the algorithms studied in the course. Students taking the course at the graduate level (as CSE 497) will also serve as scribe for one lecture, recording notes that arise as a result of the discussion that ensues. (In terms of the CSE Graduate Core Requirements, CSE 497 falls in the "Computer Applications" and the "Theory" categories.)

4. Textbook

Title: *An Introduction to Bioinformatics Algorithms*
Authors: Neil C. Jones and Pavel A. Pevzner
Publisher: MIT Press
ISBN: 0-262-10106-8

5. Tentative Syllabus (subject to change)

* Introduction to molecular biology for non-biologists
  DNA, RNA, proteins, splicing, sequencing, gene structure

* Introduction to algorithms for non-computer scientists
  Correctness, iteration and recursion, time and space complexity, general methods

* Pairwise sequence alignment
  Homology, global and local alignment, dynamic programming
  Advanced alignment techniques: gap penalties, saving time and space

* Multiple sequence alignment
  Multidimensional dynamic programming
  Approximation techniques

* Physical mapping
  Restriction maps
  Double and partial digests

* Sequencing
  Shotgun sequencing
  Sequencing by hybridization, DNA microarray analysis, clustering
  Genome assembly

* Advanced topics
  Genome rearrangements
  Evolutionary relationships, phylogenetic trees