CS 599 Special Topics in Algorithms in Bioinformatics

mock syllabus

Description: The application of fast, effective, and pragmatic programming techniques to solve selected computing tasks in bioinformatics and proteomics. By presenting a connection between computer algorithms and their bio-computing applications, the class motivates students to explore unknown domains in computer science and builds a solid foundation for advanced topics. This is a specialized algorithm class, with the focus on biocomputing applications.

Objectives: Upon completion of this course, a student should be able to:

1. Examine and adapt proper algorithms for bioinformatics applications;
2. Implement fundamental algorithms with effective programming;
3. Recognize typical algorithmic problems and be able to apply them to broad issues, including bioinformatics, string manipulation, natural language processing, and more;
4. Use scientific literature

Audience: The course serves computer science graduate students who are working as programmers in Bio-industry. As a graduate level course, students are expected to be able to read extensive literature and to form questions and opinions in the process. To succeed in this class requires extensive reading.

Prerequisites: CS 519

Classes: 1820-2050 Tuesday
HT 319 (Lab), HT 235

Class Website: http://tinyurl.com/cs8er/. The class discussion board is a 'private board', that you need to be explicitly enrolled before you can read or post. First create a general purpose forum account if you do not have one already. Then send a private message to the user *admin*, with your name and course number (CS 599). Students are responsible for monitoring this class website for any new materials posted, such as homework, announcement, etc. Grade of homework assignments and exams will be posted on the college 'Blackboard' system approximately one week after its due date. Students are responsible for monitoring this site for any new material, i.e. homework, announcement, references, etc. as they are posted. You may get 'Blackboard' password information through the college IT department.

Office Hours: Monday, Tuesday: 2050-2120; Monday 1400-1500
Otherwise by appointment. EMail is a reliable way of getting a message to me in a timely fashion. Refer to the "Email Netiquette" which can be found on the class forum for proper Email format.

Textbook:

JONES, N. C. AND PEVZNER, P. A. 2004. *An Introduction to Bioinformatics Algorithms*, The MIT Press. (required)
COHEN, W. W. 2007. *A Computer Scientist's Guide to Cell Biology*, Springer (reference)

Tentative Contents:

1. Algorithms and Complexity
   . correctness . complexity and asymptotic notation . tractable problems
2. Relative Background Information in Molecular Biology
   . genes, DNA and protein . species
3. Exhaustive Search Algorithms;
   . restriction mapping . motif finding . search trees
4. Greedy Algorithms;
   . genome rearrangements . approximation algorithms . motif finding
5. Dynamic Programming Algorithms;
   . dynamic programming . Mahattan tourist problem . longest common sub-sequences . alignments . gene prediction
6. Divide-and-Conquer Algorithms;
   . Graphs . shortest superstring problem . sequencing by hybridization
7. Combinatorial Pattern Matching;
   . repeat finding . hash tables . exact pattern matching . suffix trees . BLAST
8. Hidden Markov Models;
   . Hidden Markov Models . decoding algorithms . HMM parameter estimation
9. Randomized Algorithms
   . randomized sorting algorithms . Gibbs sampling . random projections

Homework: There will be a variety of homework assignments. Programming assignment is by default due before the next class by EMail; or by a date designated when the assignment is given. A print out of the program is <u>also</u> due in class. Non-programming homework is by default due in class one week after it is given, or by a date indicated on the assignment sheet. Be sure to start your homework assignment early. Writing a program is often easy, or at least easier than the process of troubleshooting. You are advised to have the program done by Friday, and use the weekend for any 'just in case' situations. Start a homework assignment on the weekend is often almost too late.

A late submission costs you 10% each 24 hours passing the deadline, and will not be accepted for credit after the 5th day passing the deadline. You have a 'once in a semester chance' of postponing homework submission by up to three days. <u>You must inform the instructor before the deadline, if you want to evoke the '3 days forgivable' rule</u>. The instructor advises you not to request homework extensions except under *extraordinary* situations, which is defined very narrowly and literally.

Students should keep a copy of all completed homework and projects, etc.

Attendance: Attendance is critical to the success in this class, especially in the latter part of the semester when we do extensive discussions on topics not covered in the textbook. Having that said, all students are expected in all classes. If you can not attend a class, please send the instructor an EMail in advance, so that we both have something in writing in case of confusion arises. Discussions within/without class is encouraged and promoted. Absent from a class, approved or not, is not an automatic excuse to miss or delay a homework. Students missing from a class are still responsible in catching up with the class, completing the homework and turning it in on time.

If a student misses any quiz or exam without prior agreement by the instructor, the student receives a zero grade.

Any coursework and derivatives (homework, projects or programming assignments, etc.) submitted during the semester may be later used for evaluation, teaching, or research purposes; they may also be published or made public, without additional notice.

Quiz: There will be two in class quiz, when students are given programming problems to be worked out on computers in limited time.

Bonus: You get bonus points for participation in discussions in and out of class. You also earn bonus for posting on class website discussion board. The bonus points allocation is at the instructor's sole discretion.

Grading Policy:
Homework (60%) + Quiz (10%*2) + Final Exam (20%) + Bonus (10%)
A: 90%; B: 80%, C: 70%, C-: 68%, F: < 68%

Ethics: Exams and homework should reflect individual work. Code '*copy & paste' is strictly prohibited under any circumstances.* If you adopt ideas from a book, talk, or the Internet in your work, please give proper credit to the original sources. *If* a quiz or exam is open book, you can not discuss with *any one, in person or electronically*; and you can not use a computer in any exam unless approved by the instructor in advance.

Be prepared to spend 3-4 hours per week on this course outside of the classroom. Read geek news (slashdot, etc), follow tech trends, make some fun. As to programming skills, *practice makes perfect*.

Tips to succeed in this class:
1. start working on the homework as early as possible, better the same evening of the class;
2. ask questions whenever the moment you feel lost in the class;
3. participate in the forum discussions

Note: This course syllabus/outline is tentative and subject to perpetual revisions as needed. The most up-to-date version will be posted on the class web page. The instructor may change any aspect of it with advance notice.