

Bioinformatics
Professor Hassan Bajwa

OFFICE: Tech 154,
OFFICE HOURS: _TBA-
OFFICE PHONE: 576-4571
E-MAIL ADDRESS: hbajwa@bridgeport.edu
CLASS HOURS: Thursday 06:00PM - 08:30PM, Mandeville Hall, Room 3

Course Description:

Biology has become target of more algorithms than any other fundamental science. This course is about designing and developing algorithms for biological problems. Student will work with popular bioinformatics algorithms not only to understand algorithms design mythologies but also to identify the potential weaknesses in traditional bioinformatics algorithms.

Labs are workshops are integral part of the course. Practical exercises using both informatics software and writing simple programs in Java will give understanding to students how these bioinformatics algorithms really works. The course is aimed both at biomedical engineering and computer science students. Though no prior experience in Java is required, students are exacted to have mathematics and some programming background. During the course students are expected to attend workshops and labs in addition to that they are expected to spend few hours weakly in computer lab to solve problems and lean required programming skills.

Textbook: (REQUIRED)

An Introduction to Bioinformatics Algorithms by Neil C. Jones and Pavel A. Pevzner

ISBN-13: 978-0262101066

Syllabus

1. Introduction
2. Algorithms Design Techniques
 - Exhaustive Search
 - Greedy Algorithms
 - Dynamic Programming
 - Divide and Conquer Algorithms
 - Machine Learning
 - Randomized Algorithms
3. Molecular Biology Primer
 - Genetic Material
 - DNA structure
 - DNA & Protein
 - How to Analyze DNA

4. Exhaustive Search
 - Restriction Mapping
 - Restriction Mapping Algorithms
 - Finding Motifs
 - Finding String

5. Greedy Algorithms
 - Genome Rearrangements
 - Sorting by reversal
 - Approximation Algorithms
 - Approximation for Motif finding

6. Dynamic Programming
 - DNA sequence comparison
 - Least Common Sequences
 - Longest Common Sequences
 - Sequence Alignment

7. Graph Algorithms
 - Graphs and Genetics
 - Sorting Superstring problem
 - Sequencing by Hybridization
 - Protein Sequencing and Hybridization

8. Pattern Matching
 - Repeat finding
 - Hash tables
 - Keyword Tree
 - Approximate Pattern matching
 - BLAST

9. Using Bioinformatics Tools

Assessment

Tentative schedule for the assignments and exams:

Assessment procedure	Date	Proportion
Midterm	-----	20%
Final	Final Examination Week	25%
Project	-----	40%
Homework	Announced in class	15%

Midterm and final: Both midterm and final are in-class closed book exams (only handwritten class notes are allowed). The final is a cumulative test exam

Homework: I encourage every one to take the homework very seriously as they will help you prepare for the exams Homework is due at the beginning of class , I will not accept any homework after I start the lecture.

Project:

Hadoop MapReduce is a programming model and software framework for writing applications that rapidly process vast amounts of data in parallel on large clusters of compute nodes. In this project we will utilize distributed computing model to parallel process biological datasets using Apache Map-Reduce frameworks.

Class project will be the implementation of proven genetic String algorithms like LCS largest common subsequence, edit distance, dot matrix and distributed queries in map reduce framework.

Cheating

Cheating includes and is not limited to looking at a neighbor's answer sheet during an in-class exam, copying the solution of assignment. You are encouraged to discuss homework problems with you class mates but every students is expected to solve problems by themselves.

It is the student's responsibility to familiarize himself or herself with and adhere to the standards set forth in the policies on cheating and plagiarism as defined in Chapters 2 and 5 of the Key to UB <http://www.bridgeport.edu/pages/2623.asp> or the appropriate graduate program handbook.