General Information

Instructors:

David Fernández-Baca
Office : 209 Atanasoff Hall
Phone : 294-2168
email : fernande@cs.iastate.edu
Office Hours : by appointment

Srinivas Aluru
Office : 3218 Coover Hall
Phone : 294-3539
email : aluru@iastate.edu
Office Hours : by appointment

Course Overview

Course Outline

This course will cover fundamental algorithmic techniques for solving problems in molecular biology. It will serve as the foundation for students who are interested in doing research in this area, and for students who are interested in further advanced studies in this area. Tentative list of topics covered:

- Introduction to molecular biology – Basic introduction including DNA, proteins, central dogma etc.
- Pairwise sequence alignments – Global, semi-global and local alignments, affine gap penalty functions, Hirshberg’s space-saving algorithm, banded dynamic programming.
- Multiple sequence alignments – sum-of-pairs scoring function, Carillo-Lippman heuristic, approximation algorithms, tree alignments.
- String data structures and algorithms – look-up tables, suffix arrays and suffix trees, construction algorithms, basic applications of suffix trees, lowest common ancestors.
- Applications of string data structures and alignments – database searching, genome assembly, EST clustering.
- Phylogenetics – distance based methods including ultrametric and additive distances, character based methods including parsimony and perfect phylogeny, heuristic methods.
• Microarrays – Introduction to microarray methodology, clustering algorithms.

• Computational structural biology – RNA secondary structure prediction, protein sequence-structure alignments, protein threading, structure comparison.

• Software tools – BLAST, CLUSTALW, PAUP*, Phylip etc.

Text

There is no prescribed text. The following books are useful references:

• Dan Gusfield, Algorithms for Strings, Trees and Sequences: Computer Science and Computational Biology.

• Joao Carlos Setubal and Joao Meidanis, Introduction to Computational Molecular Biology.

• Charles Semple and Mike Steel, Phylogenetics.

Prerequisites

ComS 311 or equivalent.

Course Website

The course has a website at http://www.cs.iastate.edu/~cs548. The website will be used for posting the homeworks. It also contains additional useful information pertaining to the class.

Course Policies

The course grade will be primarily based on homeworks and exams. There will be three exams – two midterm exams and a final exam. The final exam will be comprehensive. The problems given on homeworks and exams may differ between BCB/Gen and CprE/ComS, with more emphasis on algorithm design and analysis on the CprE/ComS side. Homeworks are to be done independently unless otherwise stated. The course grade will be determined based on the following weights:

<table>
<thead>
<tr>
<th>Component</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homeworks</td>
<td>20%</td>
</tr>
<tr>
<td>Midterm I</td>
<td>25%</td>
</tr>
<tr>
<td>Midterm II</td>
<td>25%</td>
</tr>
<tr>
<td>Final Exam</td>
<td>30%</td>
</tr>
</tbody>
</table>

Learning Disabilities

If you have a learning disability that makes it difficult for you to understand the lectures or perform the required work, please inform the instructor(s). Every possible effort will be made to accommodate any special needs required.