University of Colorado at Colorado Springs

CS 509: Syllabus

Bioinformatics

MW 5:50-7:05 p.m.\(^1\)

Summary: Bioinformatics is the science and engineering of organizing and analyzing complex biological data. Bioinformatics primarily deals with computational processing of biological sequence data representing genes and the contents of genomes, the complete genetic materials of organisms, and with predicting the function and structure of organic macro-molecules such as proteins. It also deals with simulation of life processes at various levels of detail. We will learn about the minimal biological basics, string comparison algorithms, computational learning algorithms in the context of bioinformatics, and also read and perform independent research.

Instructor: Jugal Kalita (262-3432)
Office: Engineering 178
Hours: MW 2:00-3:30 p.m., TR 5:45-6:15 p.m., or by appointment

Text Book & Recommended Books

I looked around a lot but didn’t find the single book that discusses all the topics that I want to cover in the class at the level I want. I decided that I will recommend two text books.

I will also direct you to various Web sites, and you should also surf the Web to find relevant sites. If you find a site that is particularly useful, please let me know so I as well as the rest of the class can learn from it. It is important that you go to the Web sites I mention in class or are listed in the books to get a feel for, and learn to use the many databases and programs that are on the Web, and the numerous tutorials and papers that are available.

The first text book is Essential Bioinformatics by Jin Xong. It’s comprehensive and up-to-date in terms of its coverage. It discusses the important topics in a manner that non-computer science students can follow. However, the topics discussed are important for everyone. The chapters are usually short and to-the-point.

The second text book is Biological Sequence Analysis by Durbin et al. It is a mathematical book. We will refer to sections of this book, but it’s quite likely that we will not work out all the math.

Here is the list of other books for you to look at. You need not buy any of these books.

\(^1\)Because the class is really small, I am willing to negotiate another time to meet if everyone agrees.
• *Statistical Methods in Bioinformatics*, Ewens and Grant. It is a fairly dense mathematical book, and is expensive.

• *Fundamental Concepts of Bioinformatics* by Krane and Raymer. It is a small but expensive book. I used it as a text book a couple of times but it’s out of print now.

• *Introduction to Bioinformatics*, by Arthur M. Lesk. The objective of this book is to generate an understanding of the biological background of bioinformatics and to integrate it with an introduction to the use of computational skills. It does not describe the algorithmic or computer science details, but is very stimulating in its description of the computational tools used in bioinformatics. One of the most interesting topics discussed is the analysis and prediction of protein structures with an aim to drug discovery.

In addition to the books mentioned earlier, it would be good to have a book on algorithms that has a chapter on string processing. One such book is *Introduction to Algorithms* by Cormen, Leiserson, Rivest and Stein, MIT Press and McGraw-Hill, 2001. Another great book devoted entirely to the topic is *Algorithms on Strings, Tress, and Sequences: Computer Science and Computational Biology* by Dan Gusfield, 1997.

**Schedule of Topics**

The following schedule is tentative. I will give out handouts when appropriate.

<table>
<thead>
<tr>
<th>Weeks</th>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Introduction to Genomes and Bioinformatics, Resources on the Web</td>
</tr>
<tr>
<td>3</td>
<td>Data Searches and Pairwise Alignment, String algorithms</td>
</tr>
<tr>
<td>3</td>
<td>Substitution Patterns</td>
</tr>
<tr>
<td>3</td>
<td>Phylogenetic Tree Construction</td>
</tr>
<tr>
<td>3</td>
<td>Protein Structure Prediction, Proteomics and Drug Discovery</td>
</tr>
</tbody>
</table>

**Grading Scheme:**

• Home work assignments will count for 30% of the final grade. There will be three or four home-work assignments. Each one will involve theory as well as programming. All assignments may not be equally weighted. You will program in a language of choice, Perl or C/C++ preferred. **It is extremely important that you do not miss any home work assignment.**

• Midterm will count 20% toward the final grade.

• An individual semester project is the basis for most of the rest of the class grade. A project proposal is due by the 4th week of the semester. The proposal must show that you have read some scholarly papers.² You are required to do a mid-term demo. You are also required to

²Visit http://scholar.google.com and http://citeseer.ist.psu.edu to search for scholarly articles in addition to any other sites you find.
do a final demo of your project and hand in a professionally-written report with details of everything you have done. The final project report and the demo are worth 25%. The final demo and report are due on the last day of classes. The final report should be in the style of a conference or journal paper, say 10-12 pages long, single-spaced and properly illustrated and referenced. Use IEEE author style for all your reports.\(^3\)

- The final exam is a 20-30 minute class presentation. The final exam is worth 20% of the class grade. Please hand in your 10-12 page Powerpoint or similar presentation on this day.

- Class attendance and participation throughout the semester are worth 5% of the grade.

**Note:** If you have special reasons for not being able to hand in an assignment on time or take an examination on a scheduled date, please make prior arrangements with the instructor.

**Important Dates:**

<table>
<thead>
<tr>
<th>Event</th>
<th>Due Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semester Project Proposal due</td>
<td>2/7/2007</td>
</tr>
<tr>
<td>Midterm Exam</td>
<td>3/12/2007</td>
</tr>
<tr>
<td>Midterm Project Report and Demo</td>
<td>By 3/20/2007</td>
</tr>
<tr>
<td>Final Project Report and Demo</td>
<td>By 5/3/2007</td>
</tr>
<tr>
<td>Final Exam and Presentation</td>
<td>5/10/2007</td>
</tr>
</tbody>
</table>

---

\(^3\)On Google, search for “IEEE Guide for Transactions Authors” to find the style files.