**Data Structures and Algorithms for Computational Biology**

Fall 2022

Dates and locations: Tuesday-Thursday, 10:00-11:30 AM

**Instructor:**               Dr. Mina C. Moghadam

**TAs:** Chandrima Bhattacharya:

Gabriel Deards

**Office Hours:**         by appointment

**Grading:**

All students will receive a letter grade (A-F).

* Two assignments (25% each)
* Midterm (20%)
* Final presentations (20%)
* Class participation (10%), Class attendance and participation policy: **Mandatory**.

**Course references:**

* An Introduction to Bioinformatics Algorithms (Computational Molecular Biology), Neil Jones and Pavel Pevzner, MIT Press, 2004
* Bioinformatics Algorithms: an Active Learning Approach (Compeau and Pevzner)
* Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology, Dan Gusfield, Cambridge University Press
* RECOMB, ISMB, WABI, ACM-BCB conference papers

**Academic Integrity**

Each student in this course is expected to abide by the Cornell University Code of Academic Integrity.  Any work submitted by a student in this course for academic credit will be the student's own work.

You are encouraged to study together and to discuss information and concepts covered in lecture and the sections with other students. You can give "consulting" help to or receive "consulting" help from such students. However, this permissible cooperation should never involve one student having possession of a copy of all or part of work done by someone else, in the form of an e-mail, an e-mail attachment file, a diskette, or a hard copy.

**Note:** Should copying occur, both the student who copied work from another student and the student who gave material to be copied will both automatically receive a zero for the assignment. Penalty for violation of this Code can also be extended to include failure of the course and University disciplinary action.

During examinations, you must do your own work. Talking or discussion is not permitted during the examinations, nor may you compare papers, copy from others, or collaborate in any way. Any collaborative behavior during the examinations will result in failure of the exam, and may lead to failure of the course and University disciplinary action.

**Tentative Course Schedule:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Week** | **Topics** | **Assignments** | **Notes** |
| **1** | Introduction   * Runtime analysis and big O notation * Introduction to Complexity and NP-Completeness   Data structures and complexity analysis   * Linear data structures * Hash-based structures |  | August 23  August 25 |
| **2** | Data structures and complexity analysis (cont.):   * Trees * Tree traversal algorithms * Binary search trees * Heaps and priority queues |  | August 30  September 1 |
| **3** | Algorithms- fundamentals:   * Exhaustive search or brute force * Approximation and Heuristics * Greedy algorithms (knapsack problem, radix sorting problem, Huffman codes) |  | September 5  September 7 |
| **4** | Algorithms- fundamentals (cont.):   * Divide and conquer * Dynamic programming   Algorithm- problem solving:   * Solving few problems using data structures and algorithm discussed in previous weeks |  | September 13  September 15 (zoom) |
| **5** | Algorithm- Applications in CB:   * Genome alignment and greedy algorithms |  | September 20  September 22 (zoom) |
| **6** | Algorithm- Applications in CB:   * Sequence comparison * The Manhattan Tourist Problem * Introduction to the graph theory * Global sequence alignment   Program assignment 1 description | Programming assignment 1  (To be posted on: Sep 28) | September 27  September 29 |
| **7** | Algorithm- Applications in CB:   * Global sequence alignment * Local sequence alignment * Multiple Alignment |  | October 4  October 6  TA session- programming class |
| **8** | Review sessions:   * Assignment 1 problem solving * Midterm review   (Materials: by the end of week 6) | Programming assignment 1  (Due: October 9) | October 11  TA session- Solving assignment 1  October 13  TA session- Midterm review |
| **9** | No class week | Midterm | October 18: Midterm (TAs)  October 20 (no class) |
| **10** | Algorithm- Applications in CB:  (continue on week 7 topics)   * Dynamic programming examples (Longest Common Subsequence, Needleman-Wunsch, Smith-Waterman, Hirschberg, Nussinov) |  | October 25  October 27 |
| **11** | Algorithm- Applications in CB:   * Exact pattern matching algorithms | Assignment 2  (Posted on November 5) | November 1  November 3  (Guest speaker) |
| **12** | Algorithm- Applications in CB:   * Advanced pattern matching techniques * Suffix trees, Suffix arrays   Programming assignment 2 description |  | November 8  November 10  TA session- Programming class |
| **13** | Algorithm- Applications in CB:   * Burrows-Wheeler Transform & FM Index |  | November 15  November 17 |
| **14** | More advanced topics in CB  (Papers and literature review) | Students’ presentation  (5 students) | November 22  Thanksgiving |
| **15** | More advanced topics in CB  (Papers and literature review) | Students’ presentation  (8 students)  Programming assignment 2  (Due: Dec 3) | November 29  December 1 |
| **16** | More advanced topics in CB  (Papers and literature review) | Students’ presentation  (8 students) | December 6  December 8 |