CS4710 – CS for Bioinformatics
Fall 2020

Instructor

Concettina Guerra
Email: guerra@gatech.edu (use a subject including "cs4710")
Office: College of Computing, TRSB #234A (it is really unlikely that we will ever meet in my offce)

Course Objective

Provide a broad and thorough background in formal modeling in CS. Provide algorithmic tools useful for computational biology applications. Specifically, the students will be presented

1. an introductory study of two central areas of the theory of computation: computability and complexity. This provides a formal background to understand what can and cannot be computed and how efficiently.
2. fundamental data structures, such as arrays, strings and trees, and algorithms useful in analyzing genomic and proteomic sequences
3. basic computational geometry algorithms useful in analyzing the three dimensional structure of proteins
4. graphs algorithms useful for analyzing biological networks, i.e. protein-protein interaction networks and transcriptional networks.

Requirements for admission

MS, PhD Bioinformatics

Information-Discussions:

We will be using Piazza (accessible via canvas) for class-related discussions. The system allows fast interactions with classmates and the instructor. Rather than asking question by writing emails to the instructor or TA, I encourage you to post your questions on Piazza.
If you prefer that your question addresses to only our TA and the instructor, you can use the private post feature (i.e., check the "Individual Students(s) / Instructors(s)" radio box).

If you want to talk to me, you have to request a virtual meeting by sending me an email.

Recommended Textbooks:

There are no required textbooks. I will recommend a few textbooks for each topic presenting their features and suitability in class.

Assignments

We will have 3 home works on automata theory, 1 on complexity analysis and 1 on graph algorithms. The first homework will be assigned after the first two lectures. The next three home works will be assigned approximately one every week. The last homework on graphs will be in the first decade of November. There will be 4-5 programming projects at least two of which on graphs algorithms. The first project will be assigned in the first half of September.

Canvas will be used for submission of assignments and projects.

Please note that the number and timing of the assignments is subject to changes as the course evolves.

Grade Breakdown

Your grade will be based on the following.

a. 40% homework

b. 50% projects

c. 10% participation

The grade of the last two (or three projects) will be assigned based on the correctness and quality of the code (25% of the project worth) and 25% on the project presentation, either a poster presentation or an oral presentation. More details later.

Participation will be based on the student's absences. Excused absences involve those to attend conferences, job fairs etc. Furthermore, a student's involvement in classes through actively engaging in discussions and question/answer activity will be considered.

Late assignment policy.

There will be a penalty for late assignments of 5% for each late day. No penalties for medical reasons or emergencies. And should they arise, you must contact the Dean of Students office. Doctor's notes, medical documentation, explanation of emergencies, etc. should be submitted to the Dean's office. After their office receives the information, they will notify me on your behalf.

Grading Scale

When assigning course grades, I will use the standard grade thresholds (90 for A, 80 for B, etc.). However, I may lower (and never raise) the thresholds (i.e., to your benefits). For example, I may use 88 instead of 90 for an A.

List of topics

(minor changes are possible)

1. Theory, such as formal models, data structures, and algorithms
    - a. Basic proof techniques such as induction
    - b. Counting
        - Sum and series, permutations and combinations
    - c. Formal language (computing theory) topics will include the following:
        - Deterministic and non-deterministic finite automata
        - Regular and Non-regular languages
        - Context-free languages and grammars
        - Computability Theory: Turing machines and Undecidability
        - Complexity theory
    - d. Algorithm and data structures topics will include the following:
        - Algorithm techniques: divide & conquer, greedy method.

      Arrays, lists, strings, trees:
        - Searching and sorting, traversals

      Graphs:
        - Global and local properties of graphs
        - Algorithms:
        - Tree traversals such as depth-first search, breadth-first search

Finding shortest paths
Graph isomorphism

2. Programming

- The students will learn and use the language Python during the lab sessions. The sessions will be mostly remote.

The students will be involved in programming projects on relevant applications in bioinformatics. The projects will be assigned by the instructor. A student might choose to work on a project of choice that I approve. The projects will be done individually, although students can work together on parts of some larger project.

Academic Integrity

Georgia Tech aims to cultivate a community based on trust, academic integrity, and honor. Students are expected to act according to the highest ethical standards. For information on Georgia Tech's Academic Honor Code, please visit http://www.catalog.gatech.edu/policies/honor-code/ or http://www.catalog.gatech.edu/rules/18/.

Any student suspected of cheating or plagiarizing on a quiz, exam, or assignment will be reported to the Office of Student Integrity, who will investigate the incident and identify the appropriate penalty for violations.

Accommodations for Students with Disabilities

If you are a student with learning needs that require special accommodation, contact the Office of Disability Services at (404)894-2563 or http://disabilityservices.gatech.edu/, as soon as possible, to make an appointment to discuss your special needs and to obtain an accommodations letter. Please also e-mail me as soon as possible in order to set up a time to discuss your learning needs.