CS4710 – CS for Bioinformatics
Fall 2016

Instructor: Concettina Guerra
          Email: guerra@cc.gatech.edu   (use a subject including "cs4710")
          Office: College of Computing, TRSB #234

Teaching Assistant: Tian Haozheng
          Email: tianhzh@gatech.edu   (use a subject including "cs4710")

Course Objective: Provide a broad and thorough background in formal modeling tools and algorithmic tools useful for computational biology applications. Specifically the student 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. fundamental data structures, such as trees and graphs, and graph algorithms useful for analyzing biological networks, i.e. protein-protein interaction networks and transcriptional networks.

Requirements for admission: MS, PhD Bioinformatics

Students are required to frequent the course web site for handouts, assignments, updates, and useful URLs. We will be using Piazza for class-related discussions. The system allows fast interactions with classmates and the instructor. Rather than asking questions by writing emails to the instructor, I encourage you to post your questions on Piazza.

Recommended Textbooks:
Introduction to the Theory of Computation by Michael Sipser. Of this book I will cover only the first few chapters.
Beginning Perl for Bioinformatics by James Tisdall 2nd edition.
Mastering Perl for Bioinformatics by James Tisdall.

Grade Breakdown: Your grade will be based on the following.

a. 20% homework
b. 30% project
c. 25% midterm
d. 15% final exam
e. 10% participation

What follows is the list of topics that I taught in the Fall 2015. There might be some minor changes.

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:**
        Regular languages:
            Deterministic and non-deterministic finite automata
            Regular expressions
            Non-regular languages
        Context-free Languages:
            Context-free grammars
            Chomsky normal form
            Pushdown automata, parsing
        Computability Theory:
            Turing machines
            Undecidability
        Complexity theory
            NP-completeness
    d. **Algorithm and data structures topics will include the following:**
        Algorithm techniques: divide & conquer, greedy method.
        Arrays, lists, strings, trees:
         Searching and sorting
         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 be involved in few programming projects on relevant applications in bioinformatics. They will be assigned by the instructor. A student might instead work on a project related to other biological applications of his choice that I approve. The project will be done individually, although students can work together on parts of some larger project.

The students will preferably use the language **Perl**. On the other hand, a student might code in a different language (like Python) if she prefers so.