CS4710 – CS for Bioinformatics
Fall 2015

Instructor: Concettina Guerra
Email: guerra@cc.gatech.edu   (use a subject including "4710")
Office: College of Computing, TRSB #234 A

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 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.

CS4710 is a 4 credit hour course consisting of a lecture on Tuesdays and Thursdays

Requirements for admission: MS, PhD Bioinformatics

Students are required to frequent the course web site for handouts, assignments, updates, and useful URLs.

Required Textbooks:
Introduction to the Theory of Computation by Michael Sipser.

Beginning Perl for Bioinformatics by James Tisdall 2nd edition.

Recommended Textbooks:
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

1. **Theory, such as formal models, data structures, and algorithms**
   a. **Formal language (computing theory) topics will include the following:**
      Basic proof techniques such as induction
      Basic techniques to solve recurrences
      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
   b. **Algorithm and data structures topics will include the following:**
      Algorithm techniques: divide & conquer, greedy method, and dynamic programming
      Arrays, lists, strings, trees

      Graphs:

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

2. **Programming, basic software engineering**

   The students will be involved in a programming project on a relevant application in bioinformatics. They will learn and use the language **Perl,** considered by many to be the de facto standard programming language for Bioinformatics applications.