Cpt S 471/571: Computational Genomics

Spring 2021

When: MWF 11:10-12:00

Where: Zoom

Instructor weekly office hour for Spring 2021 (walk-ins):
Wednesdays 1:30-2:30pm @ Zoom office hour

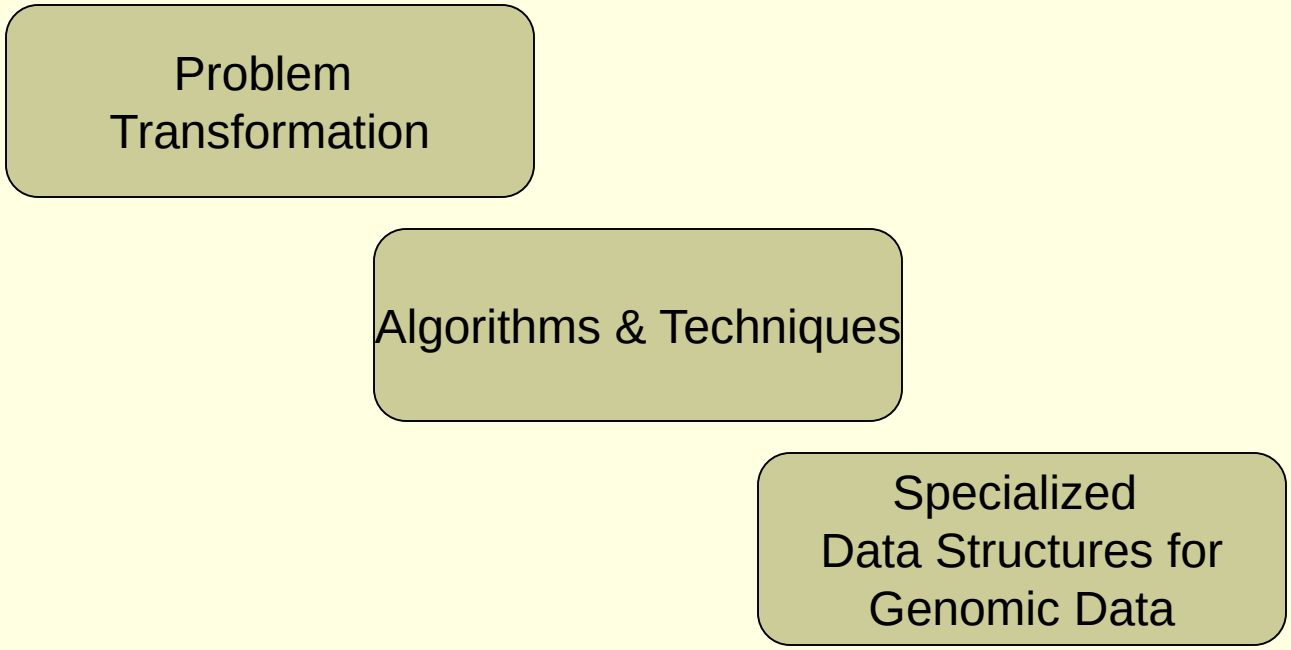
Course Objectives

- To introduce the set of algorithms and data structures that have applications to computational genomics
- To be able to formulate and/or model a biological problem/system as a computer science problem
- To be able to design algorithms using appropriate data structures to solve the underlying biological problem
- To be able to appreciate the role of computer science in modern day biological sciences (interdisciplinary training)
- To see applicability of algorithms & techniques in other domains such as text mining, pattern matching, etc.

Course Organization

- Topics:
  - Approximate string matching
  - Exact string matching
  - Probabilistic modeling for biological sequence analysis
  - Applications
    - Genome sequencing, and annotation, Read mapping, Gene identification, Clustering/transcriptomics, Phylogenetics.

Course Focus



Course Material

- Lecture Notes on the course website

- Textbook References:
  - Edited by S. Aluru. Handbook of Computational Molecular Biology, 2005. ISBN: 1584884061 (available through WSU digital library)
  - Durbin, et al. Biological Sequence Analysis: Probabilistic Models of Protein and Nucleic Acids, 1999. ISBN: 0521629713
  - D. Gusfield. Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology, 1997. ISBN: 0521585198

- Other Useful References:
  - C. Setubal and J. Meidanis. Introduction to Computational Molecular Biology, 1997. ISBN: 0534952623
  - M.S. Waterman. Introduction to Computational Biology, 1995. ISBN: 0412993910

Prerequisites

- **Cpt S 223/233/215** or equivalent (advanced data structures)
- **CptS 350** or equivalent (algorithm design & analysis)
- Familiarity with basic Probability fundamentals
- Biological background NOT required but an interest and inclination to learn is!
- C/C++/Java/Python programming experience
  - Perl **not** O.K.

Grading (for Cpt S 571)

- Homework problems (10%)
- 3 Programming projects (52.5%)
- 1 Final Exam (17.5%)
- Survey project (15%) - 6 weeks
  - Propose papers and source material
  - Oral presentation (during the last 2 weeks of class)
  - 5-page survey paper/report
- Classroom participation (5%)
- Grading policy: *curved*

Grading (for Cpt S 471)

- Homework problems (12.5%)
- 3 Programming projects (62.5%)
- 1 Final Exam (20%)
- Classroom participation (5%)

(No Survey project)

- Grading policy: *curved*

Course Webpage, Resources and Announcements

- *http://www.eecs.wsu.edu/~ananth/CptS571*
  - Contents to watch out for:
    - Homeworks, projects
    - Survey project details
    - Lecture notes
    - Tentative course schedule
    - Links to several reference papers, handouts, and other useful web resources
- For assignment dropbox and Zoom lectures and recordings access, use WSU Blackboard: https://learn.wsu.edu/
- All announcements & Discussions will happen on Piazza: https://piazza.com

Please do NOT email me on my WSU email id unless both Piazza and Blackboard are down

Classroom practices

- Attend the lectures and participate!
  - All lecture scribes will be posted on the Lecture Notes link (course website) after each class
  - All lectures will be automatically recorded (available after class via Blackboard Zoom cloud recordings)
- Ask questions
- Unmute to speak up
- Turn on videos if you can (will be appreciated!)
- Use of chat is okay to respond to other classmates

Homeworks & Programming Projects

- Submit on WSU Blackboard dropbox (learn.wsu.edu)
- Due 11:59 pm PDT on the respective due dates -- times are exact.
- 10% late penalty for a 24-hour grace period
- No submissions allowed after that (no exceptions)

- PDF is preferred.
- Cover page

Late Submission Policy

- Extensions *may* be allowed under extraordinary circumstances
- Contact instructor at least 1 week prior to permission

Collaboration Policy

- All assignments should be done *individually* unless a specific problem states that "collaboration is permitted".

- "Collaboration" is defined as a discussion with other students in the same class (no outsiders allowed) aimed at obtaining a better understanding of the problem question and/or exploring potential approaches at a very high level that can lead to a solution.

What is collaboration?	What is NOT allowed:
A discussion leading to a better understanding of the problem statement	Presenting or showing or sharing in any capacity, your solution or the main part of it to another student (PS: this is allowed in team projects)
A high level discussion aimed at arriving at a plausible approach to solving the problem	Referring to an online/web document showing a solution to the same or highly related problem posed in the question
All writing at the end should be 100% yours.	Referring to other sources including previous batch student solutions

Collaboration Policy

- All collaborative efforts should be explicitly acknowledged/cited in the answer sheet by all the participants using the cover sheet.
- Regardless of whether you collaborate or not, the final writing in the answer sheet should be solely yours.
  - No points will be deducted for collaboration as defined above.
- Any deviation from the above guidelines will be considered "cheating" and will be subject to academic dishonesty code. This includes sharing (or even showing of) your solutions, looking up solutions on the web and using them, etc. Depending on the level of offense, the instructor may decide to fail the student.
- You can learn more about Academic Integrity on your campus at http://conduct.wsu.edu