Syllabus of CSE 566 Spring 2023

Class Times and Classroom:
MWF 9:05AM - 9:55AM, Earth and Eng Sciences 118

Instructor:
Name: Mingfu Shao
Email: mxs2589@psu.edu
Office: W205A Westgate

Canvas:
Class page: https://psu.instructure.com/courses/2231423. Announcements, lecture notes, and assignments will be posted through Canvas. So make sure you have access to and check canvas emails regularly.

Slack:
We will use Slack for discussion. You may communicate with me via Slack. Invitation link:
https://join.slack.com/t/slack-dug6002/shared_invite/zt-1mmm9ozcy-OxbKx7cibBcilxWEPDct1

Pre-requisites:
If you do not meet the following prerequisites, you must obtain explicit permission from the instructor to enroll in the course:
1. Proficiency in at least one language from Python/Perl/C/C++ /Java and the Unix environment.
2. CMPSC465 (Data Structures and Algorithms) or equivalent.

Course Description:
Bioinformatics continues to be a growing field with immediate implications for our understanding of biology and treatment of disease. This course covers elegant algorithmic and data structure techniques and their use in bioinformatics. The emphasis is on recurrent ideas that underpin modern biological data analysis, presented in conjunction with their biological applications. The course is suitable both for students interested in doing bioinformatics research or just interested in applications of algorithms to the natural sciences.

Topics:
Some of the topics we may cover include: string matching, suffix trees, suffix arrays, pairwise alignment, multiple alignment, Burrows-Wheeler transform and FM index, genome assembly, transcriptome assembly, hidden Markov models, phylogeny, and other topics.

Materials and Textbooks:

There are no required textbooks. Recommended readings will be posted via Canvas. I will adopt materials (slides, notes, etc) from Prof. Paul Medvedev and Prof. Carl Kingsford (credits goes to and thanks them). Also, some of the materials are from the following books:

1. <u>Introduction to Computational Biology: An Evolutionary Approach</u>, by B. Haubold and T. Wiehe.
2. <u>Algorithms on Strings, Trees and Sequences: Computer Science and Computational Biology</u>, by D. Gusfield.
3. <u>Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids</u>, by R. Durbin, S. R. Eddy, A. Krogh, and G. Mitchison.
4. <u>Compact Data Structures</u>, by G. Navarro.
5. <u>Genome Scale Algorithm Design</u>, by Mäkinen, Belazzougui, Cunial and Tomescu.

Grading Policy:
The grade for the course will be composed as follows:
- Paper presentation (25%)
- Course project (30%)
- Assignments (30%)
- Class participation (15%)

Link to the guideline for paper presentation:
https://docs.google.com/document/d/1qgOU3A80yv6oZxYrt2qcG2iOFfgWFjHQiA4SfCJtTow/edit?usp=sharing
Link to the guideline for course project:
https://docs.google.com/document/d/1a-RulwVsA-r3jGkJDz1dSbFKyFlnvDm4oI2A0FTSRu8/edit?usp=sharing

Academic Integrity Policy:
Information about what constitutes plagiarism is taught at the secondary school level and will not be covered in class. However, you are required to read the following material in order to review what constitutes plagiarism:
https://usingsources.fas.harvard.edu/what-constitutes-plagiarism. Academic dishonesty includes, but is not limited to, cheating, plagiarizing, fabricating of information or citations, facilitating acts of academic dishonesty by others, having unauthorized possession of examinations, submitting work of another person or work previously used without informing the instructor, or tampering with the academic work of other students. For any material or ideas obtained from other sources, such as the text or things you see on the web, in the library, etc., a source reference must be given. Direct quotes from any source must be identified as such. All exam answers must be your own, and you must not provide any assistance to other students during exams. Additional standards Academic Integrity in CSE courses are described in: http://www.eecs.psu.edu/students/resources/EECS-CSE-Academic-Integrity.aspx.

All Penn State policies regarding ethics and honorable behavior apply to this course. Any instances of academic dishonesty will be pursued under the University and Eberly College of Science regulations concerning academic integrity:
http://science.psu.edu/current-students/Integrity/index.html .

Attendance Policy: Class attendance is reflected in the "Class participation" grade. On exceptional occasions, students may miss a class meeting to participate in a regularly scheduled university-approved curricular or extracurricular activity (such as Martin Luther King's Day of Service, field trips, debate trips, choir trips, and athletic contests), or due to unavoidable or other legitimate circumstances such as illness, injury, military service, family emergency, religious observance or post-graduate, career-related interviews when there is no opportunity for students to reschedule these opportunities (such as employment and graduate school final interviews.) In all cases, students should inform the instructor in advance, and discuss the implications of any absence. Note that if class is missed even for a legitimate purpose, there may still be work that cannot be made up, hurting the student's grade in the class. Likewise, students should be prepared to provide documentation when requested by the instructor. Students who will miss a class in accordance with Senate Policy 42-27, should present a class absence form (http://www.psu.edu/oue/aappm/classabs.pdf).

Students with Disabilities Policy: In order to receive consideration for reasonable accommodations, you must first obtain an accommodation letter from the Office for Disability Services (equity.psu.edu/ods). Please share this letter with your instructor and discuss the accommodations with them no later than the first week of courses.