03-713: Bioinformatics Data Integration Practicum
Spring 2024

Instructor: Joel McManus
Office: MI 255
Phone: 268 9407
Email: mcmanus@andrew.cmu.edu

Class time: Mondays and Fridays, 12:30 – 1:50 pm, POS 146. Office hours by appointment

Course objectives: The objective of this course is to provide a hands-on, self-directed experience dealing with biological data and integrating it to produce software and analyses that are of use to biologists. Students will be given a project and asked to design a solution using a combination of existing tools and their own developed software. Students will also gain experience working at the command line using the Bridges supercomputer at PSC.

Course work: The primary course work will be the implementation of a large project in teams of 3 to 5 students. The instructor will assign the teams. After about 4–5 weeks, teams will swap their solutions with another team and the other team will use it and improve it. The deliverables for the project include:

1. A runnable, working piece of software that solves the posed problem.
2. A per-group **write-up** of 5–9 polished pages (due on the last day of class). The write-up should include (1) a technical description of the team's solution, including how the team has ensured the solution is correct, (2) a brief comparison of the solution with other similar software for related problems (if any).
3. An approximately 12+3-minute **presentation** given to the class at the end of the course describing the team's solution to the project, including the team's successes, and difficulties. Everyone on the team should speak for some portion of the presentation. The presentations will be graded based on clarity, intelligence of approach, and progress toward solving the problem.
4. A per-group **user manual** that describes how a user would run the software and how they should interpret the results. The length should be sufficient to be useful to real-world users who may not be familiar with the software. This manual is due about 2 weeks before the end of the semester, and should be given to the team to which you deliver your solution.

60% of your grade will be based on the above project deliverables.

The other course work that will be required:

1. **Peer evaluations** of two kinds: (1) at the end of the semester, each group member will anonymously evaluate their teammates' contribution to the project; (2) each team will evaluate the software they used that was created by another team. (10% of your grade.)
2. **Individual Exams**: Each student will individually answer questions about their team's solution and describe their contributions to the project (10% of your grade.).

3. **Progress reports, design documents, class and meeting participation**. During the semester, teams will have to produce several progress reports and documents describing their plans and their progress toward a solution. There will also be meetings with the instructor to discuss progress. (20% of your grade.)

Policies

Excused absences: Students claiming an excused absence for an assessment must supply documentation (such as a doctor's note) justifying the absence. Absences for religious observances must be submitted by email to the instructor during the first two weeks of the semester.

Academic integrity: All class work should be done independently unless explicitly indicated on the assignment handout. The university's policy on cheating and plagiarism can be found here: http://www.cmu.edu/policies/student-and-student-life/academic-integrity.html. In part it reads "…students at Carnegie Mellon are expected to produce their own original academic work. Collaboration or assistance on academic work to be graded is not permitted unless explicitly authorized by the course instructor(s). The citation of all sources is required. When collaboration or assistance is permitted by the course instructor(s), the acknowledgement of any collaboration or source of assistance is likewise required. Failure to do so is dishonest and is the basis for a charge of cheating, plagiarism, or unauthorized assistance. Such charges are subject to disciplinary action." Disciplinary actions are detailed here http://www.cmu.edu/academic-integrity/headernav/policies.html. You should be familiar with these policies in their entirety.

Open Source Software: Students are encouraged to publish their data analysis pipelines so that prospective employers and graduate program admissions committees can evaluate their work. Publications detailing the work, or simply posting your software to GitHub under a BSD open source license are two ways to accomplish this. This is not a requirement of the course, but is strongly encouraged.

Schedule

March 11: Lecture on course organization, introduction to the problem, getting ACCESS user accounts / access to Bridges and completing the course survey.

March 15: Lecture on project goals, software development, teamwork, and using Bridges.

March 18: Lecture on using Bridges and some common software tools

March 20: ** First progress report due (canvas). This report should consist of two sections:
(1) Questions about the problem
(2) Assignments of roles to each member of the team.

March 22: Working on projects and getting help from TAs and instructor.

March 25: Working on projects and getting help from TAs and instructor.

March 29: Discussion of project progress, help from instructor and TAs

April 3: **Second progress report due: Design document.

Submit (canvas) a description of your planned design for solving the problem. This should be more than an uninformed plan. By now, you should have tested software, sketched out algorithms, and have a solid start towards creating a solution. Your progress report should also include a description of which parts of the problem you've decided are too hard to tackle during, and a timeline for completing the work you've chosen to do. You only have **3 weeks** from this point to produce a working solution.

April 5: Project discussion with TAs and instructor / working session with your group.

April 8: Project discussion with TAs and instructor / working session with your group.

April 12: No class, Spring Carnival

April 14: Upload slides for Third progress report. All teams upload slides by 11 PM.

April 15: (Teams 1, 3, 5 and 7 present Monday).

Present to the class your work so far and your plan going forward. You should have at least some piece of the solution working at this point. Get feedback and suggestions.

April 19: (teams 2, 4, 6 and 8 present Friday).

April 22: Project discussion with TAs and instructor / working session with your group.

April 26: Last class meeting - Teams will deliver their software and user manuals (on canvas). Teams will evaluate the other teams' solutions over the following week.

Apr 29-May 4 (TBD) – Presentations, Final Report, and peer evaluations due on Final Exam day.

Support and Resources for Dealing with Stress:

Take care of yourself. Do your best to maintain a healthy lifestyle this semester by eating well, exercising, avoiding drugs and alcohol, getting enough sleep and taking some time to relax. This will help you achieve your goals and cope with stress.

All of us benefit from support during times of struggle. You are not alone. There are many helpful resources available on campus and an important part of the college experience is learning how to ask for help. Asking for support sooner rather than later is often helpful.

If you or anyone you know experiences any academic stress, difficult life events, or feelings like anxiety or depression, we strongly encourage you to seek support. Counseling and Psychological Services (CaPS) is here to help: call 412-268-2922 and visit their website at http://www.cmu.edu/counseling/. Consider reaching out to a friend, faculty or family member you trust for help getting connected to the support that can help.

If you or someone you know is feeling suicidal or in danger of self-harm, call someone immediately, day or night:

CaPS: 412-268-2922

Re:solve Crisis Network: 888-796-8226

If the situation is life threatening, call the police:

On campus: CMU Police: 412-268-2323

Off campus: 911

If you have questions about this or your coursework, please let us know.

Diversity and Inclusion

Carnegie Mellon University is a diverse community of students, faculty, staff, and researchers. Our diversity comes from our own personal identities, and we must not discriminate against others based on these personal identities. This course is intended to serve students from all diverse backgrounds and perspectives.

At CMU, we work to advance and promote diversity and inclusion because it is just and fair to do so. In this class, we are all expected to be respectful to each other and listen with an

open mind to each other's opinions. This is especially important for teamwork, as everyone has something valuable to contribute, and all must be respected.

Unfortunately, bias and discrimination occur, both intentionally and unintentionally. If you feel someone has been subjected to unfair bias based on their identity, you can share your concerns with me or with the Center for Student Diversity and Inclusion at csdi@andrew.cmu.edu, (412) 268-2150. Incidents can also be anonymously reported at reportit.net username**:** *tartans* password**:** *plaid*.