

Course Overview and General Information

My goal in this class is to give you enough background to understand and apply cutting-edge algorithms for rapidly searching very large bodies of textual data, particularly biological (DNA, RNA, and protein) sequences. Rather than dwell too much on specific biological motivations, we will focus on key algorithmic results that have been successfully applied to problems of biosequence and other textual comparisons. Michael Brent's CSE 587A / BIO 5495 course is a great place to get a broader perspective on important questions in computational biology.

Biosequence comparison is a challenging computational task because the data volumes are enormous! We'll be studying cutting-edge computational tools for building sequence indices in small space, which is extremely relevant to searching a collection of short "next-generation" DNA sequence reads. However, the same techniques, with some tweaks, work for text of *all* kinds – web pages, paper abstracts, Chinese newspapers, you name it.

Your primary source for class notes, assignments, and handouts is the class web site, <https://classes.engineering.wustl.edu/cse584/>. Please check this site regularly for course announcements. You can send me email at jbuhler@wustl.edu. Ongoing discussion and announcements are distributed via the class Piazza board at <https://piazza.com/wustl/spring2018/cse584a/home>. Homeworks will be turned in and grades posted via Blackboard at <https://blackboard.wustl.edu>.

- **Where and When:** Mondays and Wednesdays 2:30-4:00 PM, Rudolph 203
- **Prerequisites:** Some formal algorithms experience and programming ability
- **Your Instructor:** Dr. Jeremy Buhler, jbuhler@wustl.edu
- **Office Hours:** My hours will be posted to the course Piazza board.
- **Required text:** none. I will use a combination of my own class notes (which will be posted to the course website) and selected papers and review articles from the literature.
- **Optional texts:** Gonick and Wheelis's "Cartoon Guide to Genetics"; Gusfield's "Algorithms on Strings, Trees, and Sequences" (not at the bookstore, but easily and cheaply available through Amazon)

1 Assignments

There are two types of assignment in this class: homeworks, and a final course project. Homeworks will give you opportunities to think more about the algorithms and ideas presented in class. There will be three to four of them, containing written and/or programming parts. The final project will entail implementing one or more of the algorithms discussed in class to build a useful bioinformatics application.

I will try my best to reduce the need for physical paper in this class! We will use Blackboard to turn in written assignments, and you will have repositories to which you can commit your coding assignments. Stay tuned for more detail once we get to our first homework. *For those who took my 347 class, I am going to relax the turn-in criteria to allow hand-written and scanned homework.*

Assignments must be turned in on Blackboard or in the repositories by **class time** on the due date. Late assignments will be accepted within reason (1-2 days, except by prior arrangement) but may not receive full credit.

My tentative weightings for the various assignments are as follows:

1. homework: 65%
2. final project: 35%

I will say in advance that some aspects of grading, especially for final projects, are necessarily subjective. I'm looking for

- evidence of substantial effort;
- a polished presentation that makes things easy to grade;
- a demonstration that you understand the material well enough to work creatively with it.

To ensure timely grading, I will ask for student volunteers to help me grade homeworks. Volunteers should post their solutions as *private* Piazza posts well in advance of the due date for me to evaluate. If you are chosen to grade a problem, you will receive extra credit equal to one-half that problem's value the homework you are grading.

2 Policy on Collaborations and Academic Integrity

Please see the separate collaboration policy document on the course website.