

CS 199-225: Honors Data Structures

Fall 2022 Syllabus

Updated: August 24, 2022

Instructor Information

Name: Brad Solomon
Office: Siebel Room 2233
Email: bradsol@illinois.edu
Office Hours: Thursday 10:30 AM – 12 PM (CST)

Course Description

Each semester, CS 225 offers a zero credit hour honors section that covers an advanced topic in CS related to data structures. This is the honors component to receive James Scholar or HCLA credit for CS 225. As an honors course, CS 199-225 will be much less structured than CS 225 and requires significant independent work and learning.

This semester, the course will focus on algorithms and data structures used to process sequences of characters (string data) and their use in large-scale data collections. Topics will include string search, string data structures, inexact pattern matching, and topics chosen by you. While the curriculum is heavily motivated by Big Data problems in genomics and CS, no knowledge of biology is assumed. Assignments and projects will be done primarily in C++.

Course Goals

Target outcomes for this course include:

- Understand fundamental string algorithms for search, storage and indexing
- Experience applying data structures, algorithms, and algorithm design principles to real-world data science problems
- Justify implementation choices based on theoretical or practical considerations
- Develop a foundation for future data science projects in CS

Textbooks

There is no required textbook for this course. Listed below are some optional textbooks which may supplement this course in some way. No readings will be required.

Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology
by Dan Gusfield

Introduction to Algorithms (*Third Edition*) by Thomas Cormen, Charles Leiserson, Ronald Rivest, and Clifford Stein

Compact Data Structures: A Practical Approach by Gonzalo Navarro

Course Topics

While the exact schedule is subject to change, a rough idea of the core concepts presented in this course include:

- Exact Pattern Matching
 - Naïve exact matching
 - Z algorithm
 - Boyer Moore
- Compressed Full-text Indexes
 - Suffix Trie
 - Suffix Tree
 - Suffix Array
 - BWT and FM Index
- Inexact Pattern Matching
 - Edit distance
 - Pigeonhole principle
 - Bitmasks
 - Global and Local Alignment

In addition to core content, the course has historically included bonus topics such as:

- String Graph Assembly
 - String coverage
 - Overlaps and Shortest Common Superstring
 - Overlap Layout Consensus assembly
 - De Bruijn graph assembly
 - Eulerian walks
- Sketching and Similarity Estimation
 - Sets and cardinalities
 - Cardinality estimation
 - Union, intersection, Jaccard, containment
 - Bloom Filter
 - MinHash sketch
 - HyperLogLog Sketch

Online Resources

The Piazza site is the home for discussions, questions, announcements, homeworks, solutions, lecture notes, etc. You can sign up for the class using this link: [Piazza Link](#)

Course Expectations & Grading

While no attendance will be taken and recordings will be made available, it is strongly encouraged that you show up to each lecture to take advantage of the synchronous lecture experience.

Most lectures will be accompanied by either a short one-week assignment involving a practical implementation of a recently covered topic or part of a multi-week short project broken into one-week chunks. Each assignment [each week] is worth 100 points and an assignment is considered a 'pass' if you achieve at least 80 points. **To receive honors credit for the course, you must 'pass' at least nine of the assignments for the semester.** This, in essence, gives each student two 'free passes' to miss or otherwise fail to complete an assignment. You are strongly encouraged not to intentionally use these passes, as it is meant to help students deal with emergency situations.

Late Policy

This semester we will be testing out a deadline-free variant of the class. Assignments will be released once the corresponding content is covered but can be submitted freely until the end of the semester. As a consequence of this, no solutions will be posted.

Commitment to Diversity, Equity, and Inclusion

As the state's premier public university, the University of Illinois at Urbana-Champaign's core mission is to serve the interests of the diverse people of the state of Illinois and beyond. The institution thus values inclusion and a pluralistic learning and research environment, one which we respect the varied perspectives and lived experiences of a diverse community and global workforce.

We support diversity of worldviews, histories, and cultural knowledge across a range of social groups including, but not limited to, race, ethnicity, gender identity, sexual orientation, abilities, economic class, age, religion, and their intersections. The Grainger College of Engineering recognizes that we are learning together in the midst of the Black Lives Matter movement, that Black, Hispanic, and Indigenous voices and contributions have largely either been excluded from, or not recognized in, science and engineering, and that both overt racism and micro-aggressions threaten the well-being of our students and our university community.

This course and its social resources like Discord and Piazza should be safe and encouraging learning environment that allows for the open exchange of ideas while also ensuring equitable opportunities and respect for all. Everyone is expected to help establish and maintain an environment where students, staff, and faculty can contribute without fear of personal ridicule, or intolerant or offensive language. If you witness or experience racism, discrimination, micro-aggressions, or other offensive behavior, you are encouraged to bring this to the attention of the course director or the course staff either directly or anonymously. You can also report these behaviors to the Bias Assessment and Response Team ([BART](#)). Based on your report, BART members will follow up and reach out to students to make sure they have the support they need to be healthy and safe. If the

reported behavior also violates university policy, staff in the Office for Student Conflict Resolution may respond as well and will take appropriate action.

Students with Disabilities

Disability Resources and Educational Services (DRES), a unit of the College of Applied Health Sciences, serves as the designated office of the University that coordinates campus-wide services for students with disabilities. DRES obtains and files disability documentation, certifies eligibility for services, and coordinates the planning and provision of auxiliary aids as well as the implementation of modifications and adjustments in academic rules, policies and procedures. Professional staff are available at DRES to provide University students with disability-related consultation and to assist them in obtaining both academic and nonacademic supports and services.

Any student with a disability who may need accommodations in this class should obtain an accommodation letter from the Disability Resources and Education Services (DRES) Office. Accommodations should be made known early in the semester and cannot be retroactively applied to assignments due before the time the accommodation was made known. For more details please refer to the university-wide academic accommodations [policy](#).

Mental Health

Diminished mental health, including significant stress, mood changes, excessive worry, substance/alcohol abuse, or problems with eating and/or sleeping can interfere with optimal academic performance, social development, and emotional wellbeing. The University of Illinois offers a variety of confidential services including individual and group counseling, crisis intervention, psychiatric services, and specialized screenings at no additional cost. If you or someone you know experiences any of the above mental health concerns, it is strongly encouraged to contact or visit any of the University's resources provided below. Getting help is a smart and courageous thing to do – for yourself and for those who care about you.

Counseling Center: 217-333-3704, 610 East John Street Champaign, IL 61820

McKinley Health Center: 217-333-2700, 1109 South Lincoln Avenue, Urbana, Illinois 61801

Academic Integrity

This course inherits the Academic Integrity Policy of CS 225. For a more detailed description of this Academic Integrity Policy, please see [Academic Integrity](#). You are responsible for the content there, but as an overview:

You may NOT reference any code outside of that provided in lecture and the textbook, or receive help from ANYONE outside of course staff or other active students in this course. Be sure to credit any students who you discussed the homeworks with and be sure that your final work is 100% your own. To be clear: you may discuss how to approach a problem, but you must code the assignment by yourself. You should never give your code to other students nor allow them to read from your exact solution. Any form of code sharing or code copying from online sources is an academic integrity violation.