

Data Structures

Introduction

CS 225

Brad Solomon

August 26, 2024



UNIVERSITY OF
ILLINOIS
URBANA - CHAMPAIGN

Department of Computer Science



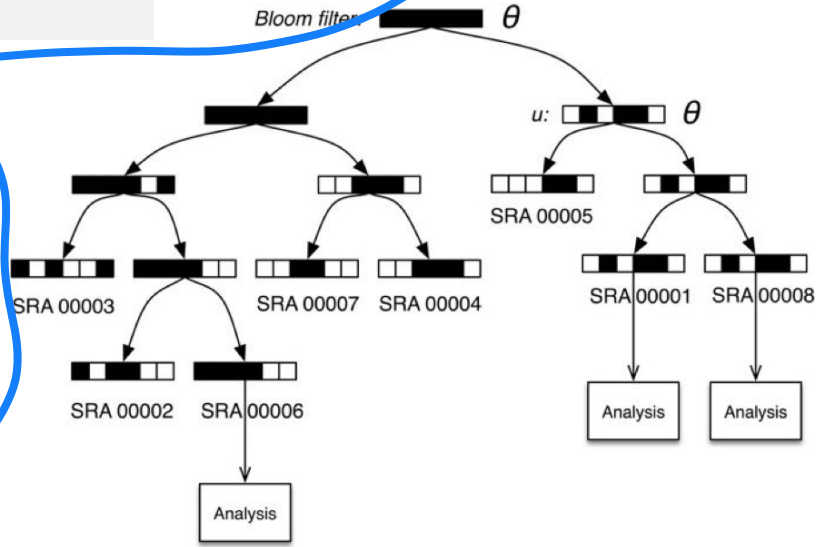
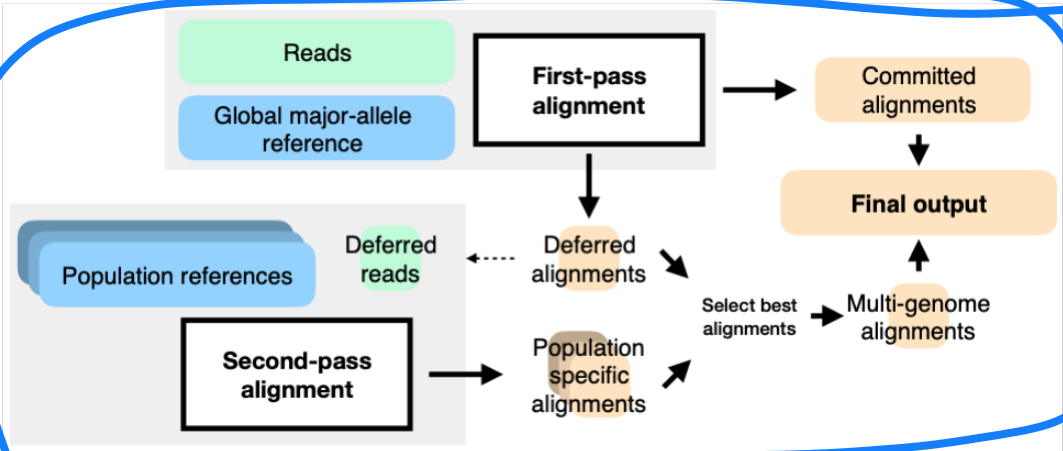
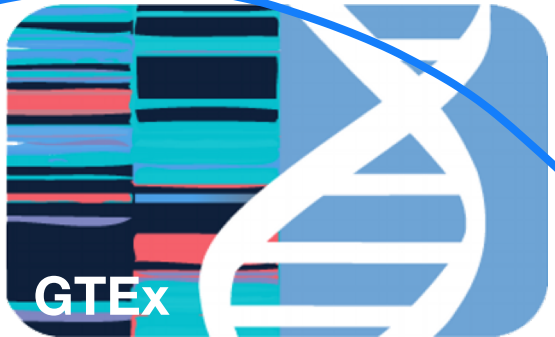
Learning Objectives

Get to know course staff

Introduce this year's Honors Course (String Algorithms)

An overview of course expectations, goals, and structure

Brad Solomon

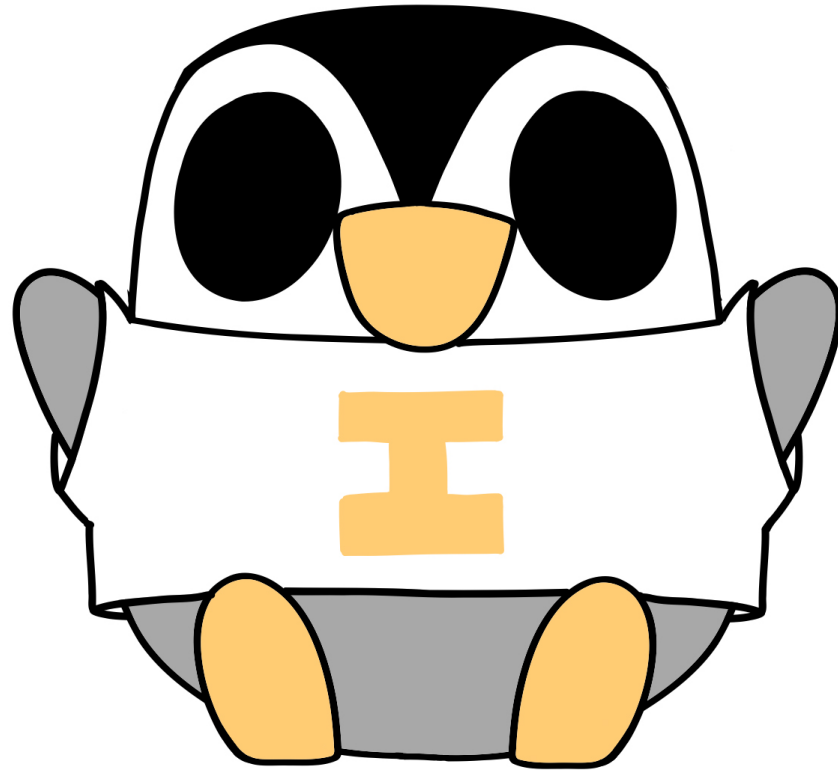


Fast search of thousands of short read sequencing experiments. Brad Solomon and Carl Kingsford. *Nature Biotech* 2016.

Reducing reference bias using multiple population reference genomes. Chen et al. *Genome Biology* 2021

Thierry
Ramais





CS225 STAFF



How to contact us?

Admin Email: cs225admin@lists.cs.illinois.edu

For best results, give a descriptive subject header!

It may take a day or two to get a response.

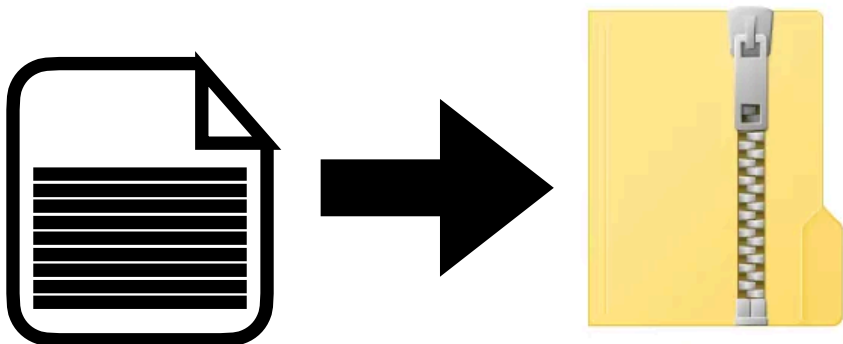
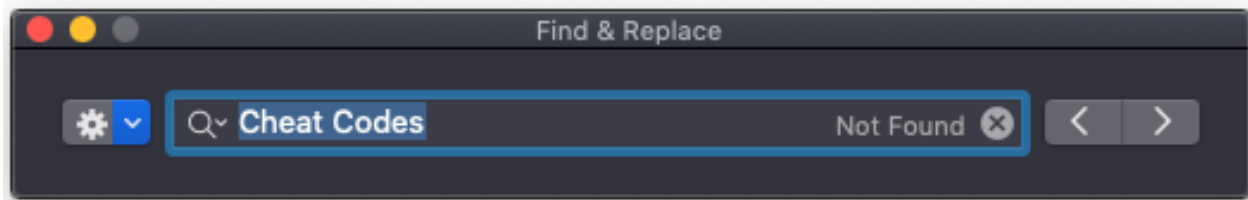
Discord: <https://discord.gg/YuEwhnR>

Don't DM course staff on Discord

Be respectful to one another online

Use Discord in class to ask questions!

CS 199-225: String Algorithms and Data Structures



```
GATCACAGGTCTATCACCTATTAACCACTCACGGGAGCTCTCCATGCATTTGGTATTTT
CGTCTGGGGGGTATGCACGCGATAGCATTGCGAGACGCTGGAGCCGGAGCACCTATGTC
GCAGTATCTGTCTTTGATTCTGCCTCATCCTATTATTTATCGCACCTACGTTCAATATT
ACAGGCGAACATACTTACTAAAGTGTGTTAATTAATTAATGCTTGTAGGACATAATAATA
ACAATTGAATGTCTGCACAGCCACTTTCCACACAGACATCATAACAAAAAATTTCCACCA
AACCCCCCTCCCCGCTTCTGGCCACAGCACTTAAACACATCTCTGCCAAACCCAAAA
ACAAAGAACCCTAACACCAGCCTAACAGATTTCAAATTTTATCTTTTGGCGGTATGCAC
TTTTAACAGTCACCCCCCACTAACACATTATTTTCCCCTCCCCTCCATACTACTAAT
CTCATCAATAACAACCCCGCCATCCTACCCAGCACACACACCCGCTGCTAACCCATA
CCCCGAACCAACCAACCCCAAGACACCCCCACAGTTTATGTAGCTTACCTCCTCAA
GCAATACACTGACCCGCTCAAACCTCCTGGATTTTGGATCCACCAGCGCCTTGGCCTAAA
CTAGCCTTTCTATTAGCTCTTAGTAAGATTACACATGCAAGCATCCCCGTTCCAGTGAGT
TCACCCTCTAATCACCACGATCAAAGGAACAAGCATCAAGCACGCAGCAATGCAGCTC
AAAACGCTTAGCCTAGCCACACCCCCACGGGAAACCTAGTGAATTAACCTTTAGCAATAA
ACGAAAGTTTAACTAAGCTATACTAACCCAGGCTTGGTCAATTCTTGGCCAGCCACCGC
GGTCACACGATTAACCAAGTCAATAGAAGCTGGCGTAAAGAGTGTCTAGATCACCCCC
TCCCAATAAAGCTAAAACCTCACCTGAGTTTAAAAAACTCCAGTTGACAAAATAGAC
TACGAAAGTGGCTTTAACATATCTGAACACCAATAGCTAAGACCCAAATGGGATTAGA
TACCCCACTATGCTTAGCCCTAACCTCAAAGTTAAATCAACAAAACCTCTCGCCAGAA
CACTACGAGCCACAGCTTAAAACCTCAAAGGACTTGGCGGTGCTTCATAACCTCTAGAGG
AGCCTGTTCTGTAATCGATAAACCCCGATCAACTCACCACCTCTGCTCAGCCTATATA
CCGCCATCTTCAGCAAACCCTGATGAAGGCTACAAAGTAAGCGCACTAACCCAGTAAAG
ACGTTAGGTCAAGGTGTAGCCCATGAGGTGGCAAGAAATGGGCTACAACTAACCCAG
AAAACCTACGATAGCCCTTATGAAACTTAAAGGGTGAAGGTGGATTTAGCACTAAG
AGTAGAGTGCTTAGTTGAACAGGGCCCTGAAGCGCGTACACACCCGCCGTTTCCCTC
AAGTATACTTCAAAGGACATTTAACTAAAACCCCTACGCATTTATATAGAGGAAAGT
CGTAACCTCAAACCTCCTGCCTTTGGTGATCCACCCGCTTGGCCTACCTGCATAAAG
AAGCACCCAACTTACACTTAGGAGATTTCAACTTAACTTGACCGCTCTGAGCTAAACCTA
GCCCCAAACCCACTCCACCTTACTACCAGACAACCTTAGCCAAACCATTTACCCAAATAA
AGTATAGGCGATAGAAATTGAAACCTGGCGCAATAGATATAGTACCGCAAGGGAAAGATG
AAAAATTATAACCAAGCATAATATAGCAAGGACTAACCCCTATACCTTCTGCATAATGAA
TTAACTAGAAATAACTTTGCAAGGAGAGCCAAAGCTAAGACCCCCGAAACCAGACGAGCT
ACCTAAGAACAGCTAAAAGAGGACACCCGCTCTATGTAGCAAAATAGTGGGAAGATTTATA
```


CS 199-225: String Algorithms and Data Structures

GATCACAGGCTATCACCTATTAACCACTCACGGGAGCTCTCCATGCATTTGGTATTTT
 CGTCTGGGGGGTATGCACGCGATAGCATTGCGAGACGCTGGAGCCGGAGCACCCATATGTC
 GCAGTATCTGTCTTTGATTCTCTGCCTCATCTATTATTTATCGCACCTACGTTCAATATT
 ACAGGGCAACATACTACTAAAGTGTGTTAATTAATTAATGCTTGTAGGACATAATAATA
 ACAAATTGAATGCTGCACAGCCACTTCCACACAGACATATAACAAAAAATTTCCACCA
 AACCCCCCTCCCCCGCTTCTGGCCACAGCACTTAAACACATCTCTGCCAAACCCCAAAA
 ACAAAGAACCCTAACCCAGCCTAACAGATTTCAAATTTTATCTTTTGGCGGTATGCAC
 TTTTAAACAGTACCCCCCAACTAACACATTTTCCCTCCCACTCCATACTACTAAT
 CTCATCAATAACAACCCCGCCATCTACCCAGCACACACACCCGCTGCTAACCCATA
 CCCCAGAACCAACCCCAAGACACCCCCACAGTTTATGTAGCTTACCTCCTCAA
 GCAATACACTGACCCGCTCAAACCTCTGGATTTGGATCCACCCAGCGCTTGGCCTAAA
 CTAGCCTTTCTATTAGCTCTTAGTAAGATTACACATGCAAGCATCCCCGTTCCAGTGAGT
 TCACCCTCTAAATCACCAGCATCAAAGGAACAAGCATCAAGCAGCAGCAATGCAGCTC
 AAAACGCTTAGCCTAGCCACACCCCCACGGGAAAGCAGTCTTACCTTAGCAATAA
 ACGAAAGTTAACTAAGCTATACTAACCCCAAGTGGTCAATTGGCCAGCCACCCG
 GGTACACGATTAAACCAAGTCAATAGAAGCGGGTAAAGCTGTATGATCACCCTC
 TCCCAATAAAGCTAAAACCTCACCTGAGTTTAAAAAATCCAGTTGATCAAAAATAGAC
 TACGAAAGTGGCTTAAACATATCTGAACACCAATAGCTAAGACCCAACTGGGATTAGA
 TACCCCACTATGCTTAGCCCTAAACCTCAAAGTTAAATCAACAAAACCTCGCCAGAA
 CACTACGAGCCACAGCTTAAAACCTCAAAGGACTGGCGGTGCTTCACTCCCTCTAGAGG
 AGCCTGTTCTGTAATCGATAAACCCCGATCAAGCTCACCACCTCTTACAGCCTATATA
 CCGCCATCTTCAGCAAACCTGTAGAAAGGCTACAAAAGTGAAGTCAACGAGTAAAG
 ACGTTAGGTCAAGGTGAGCCATGAGGTGGCAAGAAATGGGTACATCAACCCAG
 AAAACTACGATAGCCCTTATGAAACTTAAAGGTGCAAGGTGGATTAGCACTAAG
 AGTAGAGTGCTTAGTTGAACAGGGCCCTGAAGCGCGTACACACCCGCGCTCACTC
 AAGTACTTCAAAGGACATTAACATAAACCCCTACGCATTTATATAGAGGAGTCT
 CGTAACCTCAAACCTGCTGCTTTGGTGATCCACCCGCTTGGCTACCTGCATAATGAAG
 AAGCACCCAAACCTACACTTAGGAGATTTCAACTTAACTTGAACCTGAGCTTAAACCTA
 GCCCAAACCCACTCCACCTTACTACCAGACAACCTTAGCCAAACCATTTACCCAAATAA
 AGTATAGGGGATAGAAATGAAACCTGGCGCAATAGATATAGTACCGAAGGGAAAGATG
 AAAAAATATAACCAAGCATAATATAGCAAGGACTAACCCCTATACCTTCTGCATAATGAA
 TTAAC TAGAAAATAACTTTGCAAGGAGAGCCAAAGCTAAGACCCCGAAACAGAGCAGCT
 ACCTAAGAACAGCTAAAAGAGCACCCCGTCTATGTAGCAAAAATAGTGGGAAGATTTATA
 GGTAGAGGCGCAAAACCTACCGAGCCTGGTGATAGCTGGTTGTCCAAGATAGAATCTTAG
 TTCAACTTTAAATTTGCCACAGAACCTCTAAATCCCTTGTAAATTTAACTGTTAGTC
 CAAAGAGGAACAGCTCTTTGGACACTAGGAAAAACCTTGTAGAGAGAGTAAAAATTTA
 ACACCCATAGTAGGCCCTAAAAGCAGCCACCAATTAAGAAAGCGTTCAAGCTCAACACCCA
 CTACCTAAAAATCCAAACATATAACTGAACTCCTCACACCCAATTTGGACCAATCTATC
 ACCCTATAGAAAGAACTAATGTTAGTATAAGTAACTAAGTAACTTCTCCTCCGCAATAGC
 CTGCGTCAGATTAACACTGAACTGACAATTAACAGCCCAATATCTACAATCAACCAAC
 AAGTCATTATTACCCTCACTGTCAACCAACACAGGCATGCTCATAAGGAAAGGTTAAAA
 AAAGTAAAAGGAACTCGGCAAACTTACCCCGCCTGTTTACCAAAAACATCACCTCTAGC
 ATCACCAGTATTAGAGGCCACCCCTGCCAGTGACACATGTTTAAACGGCCGCGTACCT
 AACCGTGCAAGGTAGCATAATCACCTGTTCTTAAATAGGGACCTGTATGAATGGCTCC
 ACGAGGTTAGCTGTCTTCTTAACTTTAAACAGTGAATTTGACCTGCCCTGAAGAGGCG
 GGCATAACACAGCAAGACGAGAAGACCCCTATGGAGCTTAAATTTAATGCAAAACAGTA
 CCTAACAAACCCACAGGCTCTAAACTACCAAACTGCATTAATAAATTTCCGTTGGGGCGA

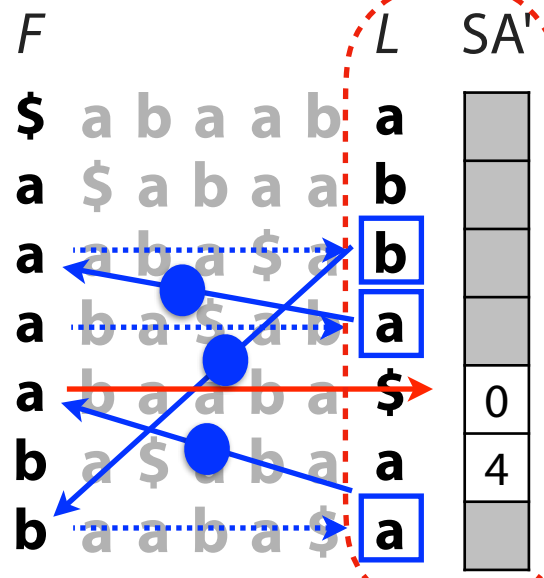
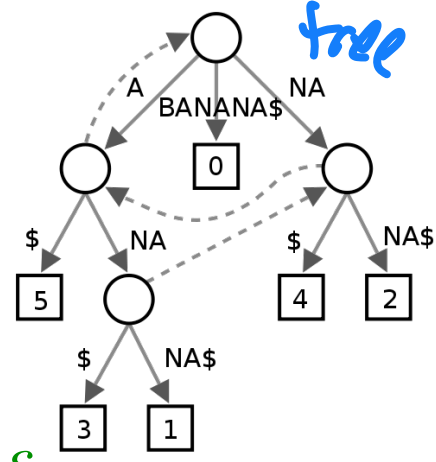
A deep dive into algorithms (and DS)

Suffix tree

T: CCTTCTGCT(A)CCTTTTGGCGGCGCG

P: CCTTTTGC

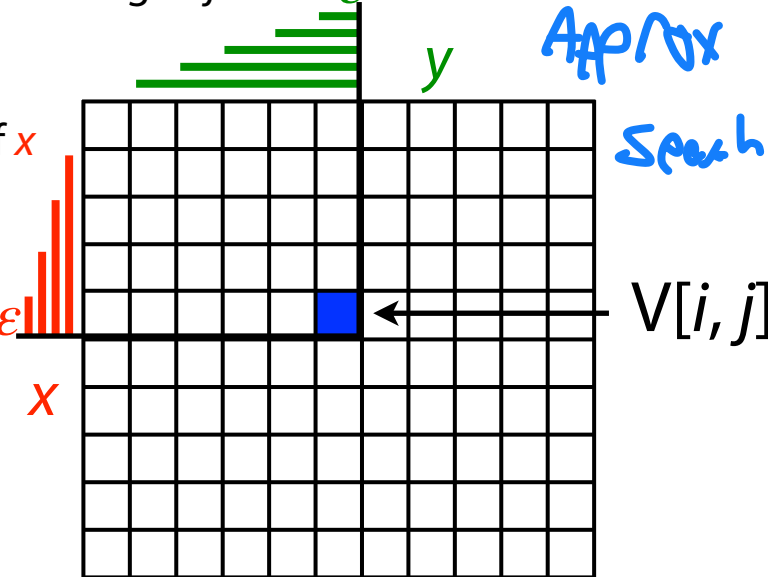
BM



FM Index

Substrings of y ending at j

Substrings of x ending at i



CS 199-225: String Algorithms and Data Structures



Logistics: 0-credit optional honors

Weekly lecture followed by weekly assignment

Monday 5-5:50 PM

Lecture → Assignment

Siebel 0216

First lecture 9/09/24



<https://courses.grainger.illinois.edu/cs225/fa2024/pages/honors.html>

Syllabus has information on enrollment / HCLAs

Everything about CS 225

<https://courses.engr.illinois.edu/cs225/>

Information on:

Staff

Communications

Lab Sections

MPs

Exams

Grading

Academic Integrity



A surprise pop quiz!

I missed the Discord link. Where can I find it?

↳ website → course information
↳ Lectures page

When is exam 0? What is on it?

← 9/4 - 9/6 → Prairie test
→ website → Exams
→ slide deck!

Exam 0 (9/4 — 9/6)

An introduction to CBTF exam environment / expectations

Quiz on foundational knowledge from all pre-reqs

Practice questions can be found on PL

Topics covered can be found on website

Registration started August ~~22~~

26

Grading — Point Distribution

Category	Contribution	Notes
<u>Machine Problems</u>	360	6 x 60 points each
Lab Assignments	120	12 x <u>10 points each</u>
Exams	360	6 x 60 points each
Final Exam	160	

All MPs have a one-day late policy for 93% credit

There are no built-in extensions for labs

Grading — Final Grades

~~920 A-~~

Points	Grade	Points	Grade	Points	Grade
<u>$[930, \infty)^*$</u>	A+	<u>$[930, \infty)^*$</u>	A	<u>$[900, 930)$</u>	A-
$[870, 900)$	B+	$[830, 870)$	B	$[800, 830)$	B-
$[770, 800)$	C+	$[730, 770)$	C	$[700, 730)$	C-
$[670, 700)$	D+	$[630, 670)$	D	$[600, 630)$	D-
		$(600, 0]$	F		

* An A+ requires both a minimum amount of points and the support of one or more course staff members who have found some part of your work exceptional.

Extra Credit Opportunities

MP Extra Credit Submission (40 pts)

Problems of the day (40 pts)

Course Feedback Surveys (20 pts)

Other points
maybe.

Extra credit is capped at 100 points.

Plagiarism Policy

Don't share your code with anyone! Ever!

Don't use or look up code solutions from any source!

Don't use generative AI tools!

Infractions will result in 0s on the assignment, 100 point loss in class, and a loss of all extra credit opportunities*

All infractions will be reported through FAIR and remain on your permanent record.

Other Syllabi Policies



Adhere to the CS Values and Code of Conduct

Your mental and physical health is important

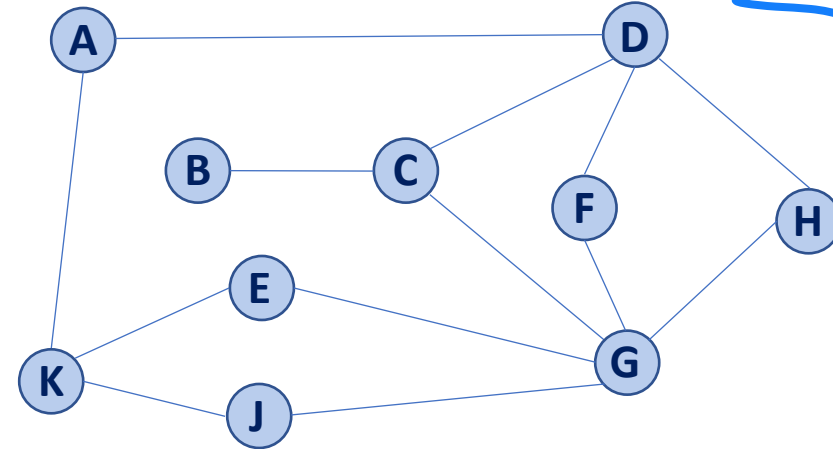
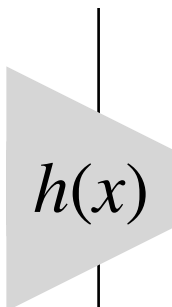
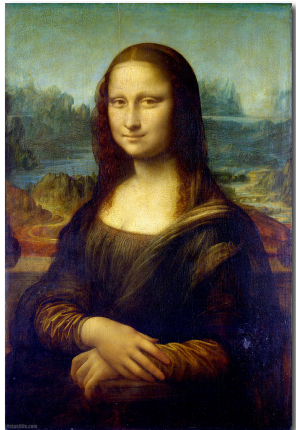
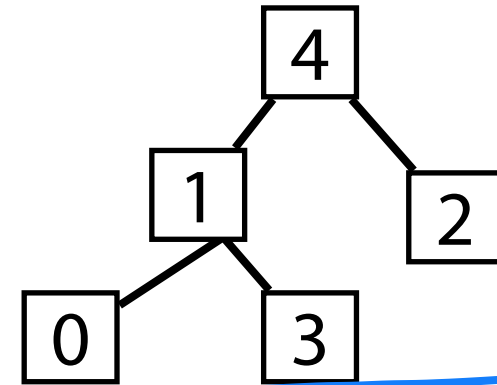
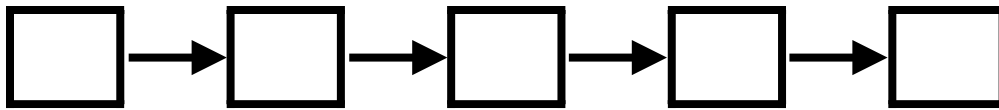
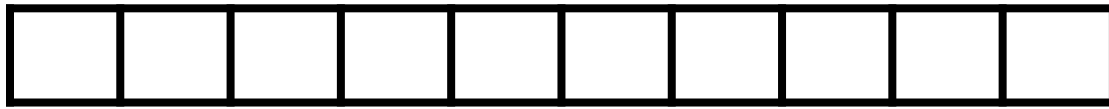


What is this course about?



CS 225 — Course Goals

Understand foundational data structures and algorithms

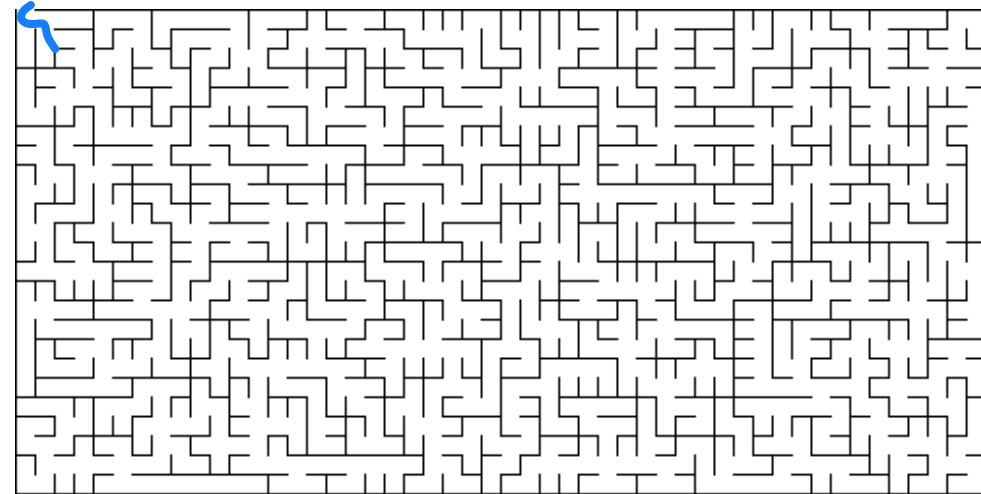
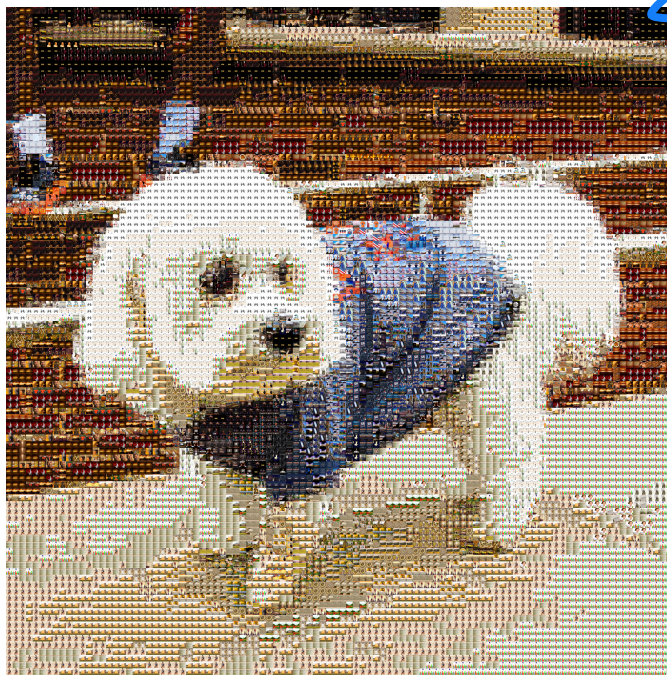
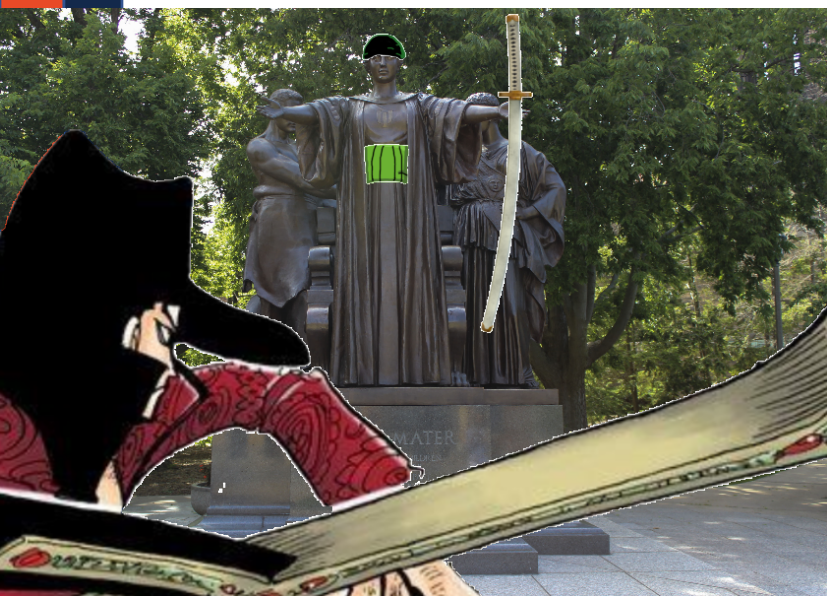


CS 225 — Course Goals

Justify appropriate algorithms for complex problems

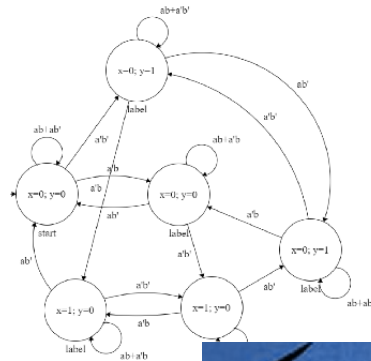
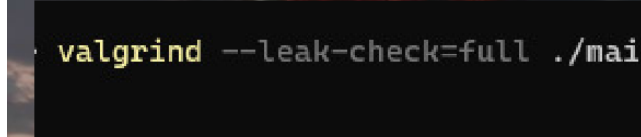
Decompose problem into supporting data structures

Analyze efficiency of implementation choices



CS 225 — Course Goals

Implement intermediate difficulty problems in C++



```
/workspaces/cs225git/mp_stickers/tests/tests-part2.cpp:227: FAILED:
  REQUIRE( sheet.render() == alma )
with expansion:
  PNG(w=961, h=600, hash=103ac580ac38a31d)
  ==
  PNG(w=960, h=600, hash=103ac580ac38a31d)
^C
StickerSheet::getSticker() returns NULL for a removed sticker
/workspaces/cs225git/mp_stickers/tests/tests-part2.cpp:255
/workspaces/cs225git/mp_stickers/tests/tests-part2.cpp:256: FAILED:
  due to a fatal error condition:
```



✓ [10/10] myImage.png exists and contains stickers

CS 225 — Course Goals

Understand foundational data structures and algorithms

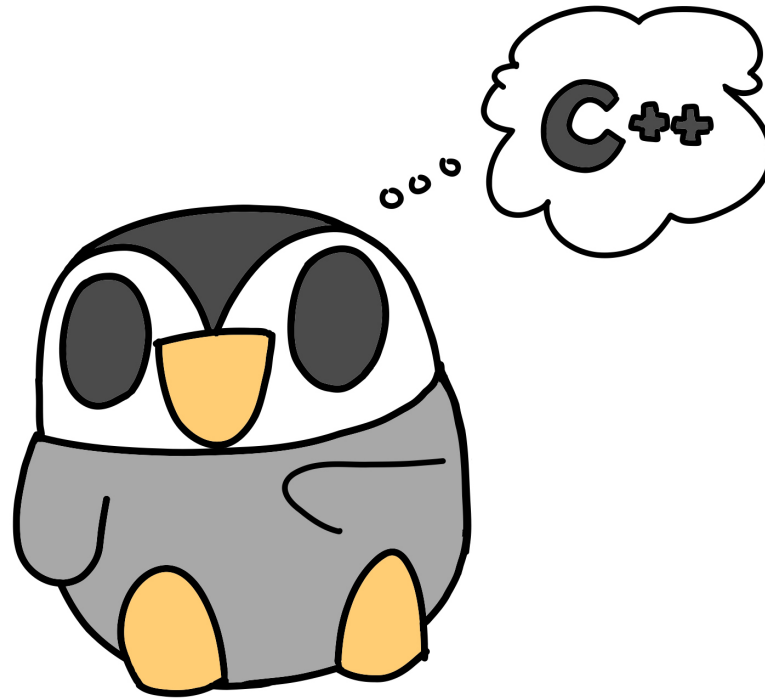
Justify appropriate algorithms for complex problems

Implement intermediate difficulty problems in C++

Improve your foundation of CS theory



What about C++



Lectures from Previous Semesters Covering C++ Available Here

https://mediaspace.illinois.edu/playlist/dedicated/177553201/1_s10ctiib/1_z2cz05fi



Problems of the Day (PoTDs)

(Optional) Open Lab This Week

This week's lab is open office hours

Focus is making sure your machine is setup for semester

Installation information available on website

