

CSE 250

Data Structures

Dr. Eric Mikida
epmikida@buffalo.edu

Dr. Oliver Kennedy
okennedy@buffalo.edu

212 Capen Hall

Day 1
Course Overview

Who are we?

- Eric Mikida [epmikida@buffalo.edu]
 - Tentative office hours:
 - Capen 212: M/T 1:00-3:00, W 3:00-5:00
- Oliver Kennedy [okennedy@buffalo.edu]
 - Tentative office hours (database, data, or HEMA questions also welcome):
 - Capen 212: Weds 11:00-1:00

**Please keep class discussions on Piazza (private/anonymous posts exist)
Always include [CSE-250] in the subject line when emailing**



212 Capen: Take these elevators, then turn right.

Who are the TAs?

Undergraduate TAs

- Hope Kara
- Vrund Patel
- Sean Grzenda
- Andrew Schick
- Heba Mahran
- Nawar Khouri
- Jacky Lin
- Amelia Graca

- Anton Kalinin
- Joey Poblete
- Thinh Ho
- Riad Mukhtarov
- David Lam
- Kartike Chaurasia
- Tirth Shah
- Dikshit Khandelwal

Graders

- Sphoorthi Keshannagari
- Vindhya Nuthalpati
- Rohit Joseph

Logistics

- **Course Forums + Live Q&A:** Sign up for Piazza
 - <https://piazza.com/buffalo/fall2022/cse250>
- **Course Website / Syllabus:**
 - <https://odin.cse.buffalo.edu/teaching/cse-250/2022fa/>
- **Assignment Submission:** Autolab
 - <https://autograder.cse.buffalo.edu>
- **Assignment Distribution:** Github Classroom

Development Environment

- Supported Operating Systems
 - MacOS
 - Ubuntu Linux
 - Windows + WSL/Ubuntu
- Supported Dev Environments
 - Emacs + Scala-SBT
 - IntelliJ (Community Edition is Free) + Scala Plugin
 - <https://www.jetbrains.com/community/education/>

Other setups are ok, but the more your setup differs the lower the chance we'll be able to help you

Course Syllabus

Grading

Grade Breakdown:

- Assignments: 30%
- Participation: 10%
- Midterm: 20%
- Final Exam: 40%

Score (x)	Letter Grade	Quality Points
$90\% \leq x \leq 100\%$	A	4
$85\% \leq x < 90\%$	A-	3.67
$80\% \leq x < 85\%$	B+	3.33
$75\% \leq x < 80\%$	B	3
$70\% \leq x < 75\%$	B-	2.67
$65\% \leq x < 70\%$	C+	2.33
$60\% \leq x < 65\%$	C	2
$55\% \leq x < 60\%$	C-	1.67
$50\% \leq x < 55\%$	D	1
$0\% \leq x < 50\%$	F	0

Written Assignments

Bi-Weekly Written Assignemnts

- Expect to spend about a week per assignment
- Submit up to 24hrs after deadline with a 50% penalty

You are responsible for submission formatting

- Submit only PDFs
- Submissions that do not load will receive a 0

We recommend writing solutions by hand

- Better retention of what you have written
- Easier to write out math by hand than on a computer

Programming Assignments

Grading for most projects will be as follows:

- Write test cases (~15/100 points)
 - Submit as many times as you like
- Test submissions (0/100 points, but predicts ~50/85 points)
 - Submit as many times as you like
- Final submission (~85/100 points)
 - Tests will run once **AFTER** the deadline has passed

Grades will always be based on the **LAST** submission you make

Programming Assignments

You have 2-3 weeks per assignment

- Plan to start early and work throughout
- 25% penalty per day late, up to 48 hours
- Bonus for early submissions (up to 5/100 points)

3 'grace days' for the semester

- Applied automatically, even if your score does not increase

Exams

One In-Class Midterm (Wednesday October 19)

- Content covered is roughly Weeks 1-7 in the syllabus
- More details as exam approaches

One Final Exam (Monday December 12, 7:15-10:15)

- Comprehensive, covering any topics from throughout the semester
- Check for conflicts ASAP
- If HUB changes the date/location...trust the HUB

If you need accommodations, contact [Accessibility Resources](#) ASAP

Class Participation

Lecture

- No recorded attendance
- Easy access to ask questions live (use it)

Recitation

- Recitations start Tue, Sept 6 (Next week)
- Attendance is mandatory

Collaboration, AI, Extra Resources

Do...

- Work together to brainstorm ideas
- Explain concepts to each other
- Include a list of your collaborators on all submitted work

Do Not...

- Write solutions when working together
- Describe the details of solutions to problems or code
- Leave your code in a place where it is accessible to another student

When in doubt, ask a member of the course staff!

Resource Policy

Do...

- Use materials provided by course staff (Piazza, Class, OH)
- Use materials from the course textbook or readings
- Cite all materials you reference for written work
- **Cite sources** for all code you reference / copy

Resource Policy

Do NOT...

- Reference random videos on YouTube that “helped you solve the problem”
- Hire “private tutors”
 - Save the money from Chegg
 - If you’re not doing the work yourself, you’re not learning
 - If you have an actual tutor, contact course staff
- Reference exact solutions found online

If you are caught using unauthorized resources, you get an F

Other Ways to Get an F

- Work in a group by assigning each person to a problem
- Copying your friend's homework because you forgot
 - Each homework is not worth a lot on its own
- Sharing your homework with your friend
 - I have no way to know who did the work and who shared
- Submitting work without citations
 - Citing outside work will help you avoid AI repercussions
 - (we grade you on the work you did, but you won't get an AI violation)

Other Ways to Get an F

You are liable/punishable if someone else submits your work as their own.

Ways to Avoid an F

Don't Cheat...but we understand mistakes are made.

We will grant amnesty for any AI violation **IF** you tell us about it **BEFORE**
we discover it

Asking Questions

First...check if the answer exists (syllabus, Piazza, course website)

Then...

Ask in lecture, recitation, Piazza, or office hours

Come prepared, form the question carefully, many times you will answer your own question in the process!

Thinking through your question is a great first step.

**Now...What even is
“Data Structures!?”**

What is a Data Structure?

Data



Container

What is a Data Structure?

Same Data



Different Container

more defensible

What is a Data Structure?

Same Data



more efficient access to
skritches()

Different
Container

What is a Data Structure?

- Store a list of things in some order (“List”)
 - Array
 - LinkedList
 - ArrayBuffer
- Store things organized by an attribute (“Map”, “Dictionary”)
 - Hash Table
 - Binary Search Tree
 - Red-Black Tree

Why should you care?

- **Tactical:** Optimize your Code (“reducing the constants”)
 - Understand the memory hierarchy
 - Understand the CPU / OS
- **Strategic:** Optimize your Design (“reducing the complexity”)
 - Understand how your algorithm scales
 - Understand repetition in your code

Tactical Programming

Go from point A to point B

1. Move up 100 feet
2. Turn right, move forward 200 feet
3. Move north 10 feet then turn left
4. Move forward 20 feet
5. Move south 50 feet
6. Move west 150 feet, then turn left
7. Move forward 60 feet

We can optimize each individual step

- For example, taking a bike will speed up step 2 compared to walking

Strategic Programming

Look at the big picture

**Design (not just implement) and
algorithm**

Focus on complexity

What is “Complexity”?

`mapped_type` must be [emplace_constructible](#) from `std::piecewise_construct`, `std::forward_as_tuple(std::move(key))`, `std::tuple<>()`.
When the default allocator is used, this means that `key_type` must be [MoveConstructible](#) and `mapped_type` must be [DefaultConstructible](#).

No iterators or references are invalidated.

Parameters

key - the key of the element to find

Return value

Reference to the mapped value of the new element if no element with `key` existed. Otherwise a reference to the mapped value of the existing element whose key is equivalent to `key`.

Exceptions

If an exception is thrown by any operation, the insertion has no effect

Complexity

Logarithmic in the size of the container.

Notes

In the published C++11 and C++14 standards, this function was specified to require `mapped_type` to be [DefaultInsertable](#) and `key_type` to be [CopyInsertable](#) or [MoveInsertable](#) into `*this`. This specification was defective

(screenshot: cppreference.com)

What is “Complexity”?



[scala.collection.immutable](https://scala-lang.org/api/scala/collection/immutable)

Vector

Companion object **Vector**

```
sealed abstract class Vector[+A] extends AbstractSeq[A] with IndexedSeq[A] with IndexedSeqOps[A, Vector, Vector[A]] with StrictOptimizedSeqOps[A, Vector, Vector[A]] with IterableFactoryDefaults[A, Vector] with DefaultSerializable
```

Vector is a general-purpose, immutable data structure. It provides random access and updates in $O(\log n)$ time, as well as very fast append/prepend/tail/init (amortized $O(1)$, worst case $O(\log n)$). Because vectors strike a good balance between fast random selections and fast random functional updates, they are currently the default implementation of immutable indexed sequences.

Vectors are implemented by radix-balanced finger trees of width 32. There is a separate subclass for each level (0 to 6, with 0 being the empty vector and 6 a tree with a maximum width of 64 at the top level).

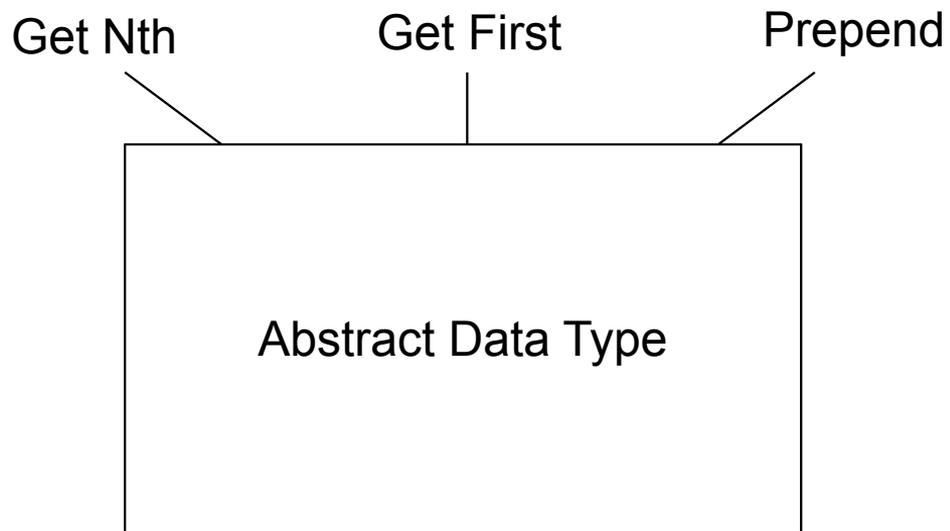
(screenshot: <https://www.scala-lang.org/api>)

What is “Complexity”?

Every (good) standard library’s provides guarantees on the complexity of its data structures’ operations

Understanding complexity bounds can be the difference between code that runs in 6 hours vs code that runs in 8 seconds.

Containers



Containers

Option 1

- Very fast Prepend, Get First
- Very slow Get Nth

Option 2

- Very fast Get Nth, Get First
- Very slow Prepend

Option 3

- Very fast Get Nth, Get First
- Occasionally slow Prepend

Which is better?

Containers

Option 1 (Linked List)

- Very fast Prepend, Get First
- Very slow Get Nth

Option 2 (Array)

- Very fast Get Nth, Get First
- Very slow Prepend

Option 3 (ArrayBuffer...in reverse)

- Very fast Get Nth, Get First
- Occasionally slow Prepend

Which is better?

IT DEPENDS!

Data Structures in a Nutshell

More work now

vs

More work later

Topics Covered - Tools

- **Specific Data Structures/ADTs** (organizational strategies)
 - Collection Types (Lists, Arrays, Vectors, Sets, Heaps)
 - Maps (Hash tables, Search trees)
 - Graphs
- **Algorithms** (recipes for standard tasks)
 - Collection queries / updates
 - Sort
 - Graph/Tree Traversal

Topics Covered - Techniques

- Pseudocode
 - Top-down algorithm design
- Algorithm Analysis / Asymptotic Notation
 - Understand algorithm scaling
- Recursion
 - Expressing tasks in terms of themselves

Topic Order

- Scala
- Asymptotic Notation
- Sequence Collections
- Recursion
- Graphs
- Specialized ADTs
- Hash-based Data Structures
- Advanced Topics

First Assignments

Academic Integrity Quiz

Will be posted before Aug 31 @ 1:00AM

- Posted on AutoLab
- Notifications via Piazza

Should take < 10 minutes

Due Wed September 7 @ 5:00 PM

Scala Project

Will be posted before Wednesday, Aug 31 @ 11 PM

- Posted on Course Website
- Notifications via Piazza

Submit your GitHub username due by Tuesday, September 6

Scala Hello World due Wednesday, September 14

- Load an Open Dataset
- Transform & Analyze the Data

Questions?