CS 15-121 Data Structures

Lecture 1 - Introduction

Prof. Ananda Gunawardena ("Guna")

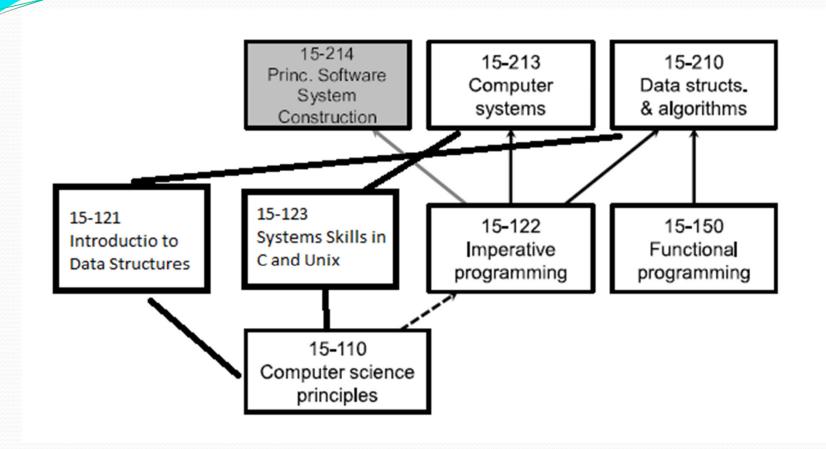
Summer

2011

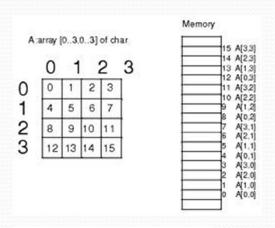
Course Description 15-121

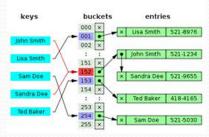
15-121 Introduction to Data Structures Fall and Spring: 10 units

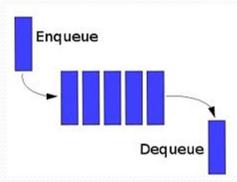
An introduction to the process of program design and analysis using the Java programming language for students with some prior programming experience (functions, loops, and arrays) in a language other than Java. Topics to be covered include an overview of fundamental programming concepts using Java as well as object-oriented programming techniques, data aggregates, data structures (e.g., linked lists, stacks, queues, trees, and graphs), and an introduction to the analysis of algorithms that operate on those data structures. This course, along with 21-127, serves as a prerequisite for 15-211. NOTE: students who receive a grade of C or less in 15-111 should discuss whether they are adequately prepared for 15-211 with their academic advisor.

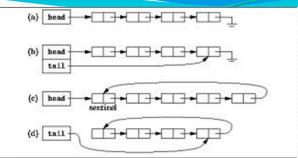


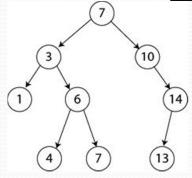
This course is about Data Structures

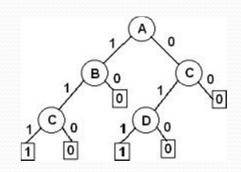


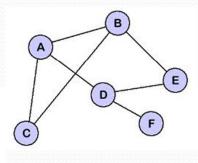


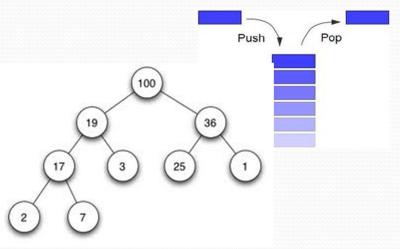












And algorithms that operates on these Data Structures

- Searching
- Sorting
- Updating
- Rearranging
- Rendering
- Swapping
- Inverting
- Restructuring
- Combining
- Selecting and deselecting

This course is also about

- Understanding Object Oriented Programming Concepts
- Understanding computers and storage
- Understanding Data Organization
- Understanding applications that require complex data structures and algorithms
- Understanding algorithms
- Understanding complexity of large programs
- Understanding debugging techniques
- Understanding large software development environments

What questions can be answered using what we learn in this course?

- If we write a program, will it end in a reasonable time?
- Are there problems that are not computable?
- 10000 records need to be organized so median, mode, max and min can be found efficiently
- 100000 records need to be organized only max and min can be found efficiently
- 100000000 records need to be organized so records can be stored and found efficiently
- 10000000 records need to be organized so ranges of data can be found efficiently
- An image with thousands of pixels need to be manipulated
- How would you develop an application for a mobile phone or tablet?

Course Staff

- Ananda Gunawardena (instructor)
 - guna@cs.cmu.edu
 - 412-268-1559
 - Office Gates 6005
 - Office hours M,W 2-3 or by appointment
- Rafee Memon (TA)
 - rmemon@andrew.cmuedu
 - Office hours TBA
 - All questions regarding assignments to Rafee

Course Components

- 6 Assignments 40%
- 5 Quizzes 10%
- Collaborative Activities 15%
- 1 Written Midterm 15%
- 1 Written Final Exam 20%
- Extra Credit opportunities 5%-10%

About Classwork

- All lectures available online (no printouts given)
- Annotated lectures saved online
- Collaborative assignments are 15% of grade
 - Create an account on classroom salon
 - http://www.classroomsalon.org
 - Code: cmu
- There will be many opportunities for class participation
 - All are expected to participate

Programming Assignments

- Requires clear understanding of the problem
 - Use salon discussions to understand the specs
- Ability to read and understand starter code
- Ability to modify code to accomplish the tasks
- Ability to debug and test code
- Typically requires 2-3 days of work a total of 5-10 hours of coding, debugging and testing

Submitting Assignments

- We use AFS for assignment handin
 - /afs/andrew.cmu.edu/course/15/111Mo8/handin
- Authenticate and submit into the correct folder we will show you how to do this in class
- You will receive feedback on your assignments on the handback folder
- We will also record your grades at
 - /afs/andrew.cmu.edu/course/15/111Mo8/grades/YourID
- Start early late programs will lose 50% per day
- No program accepted after 48 hours
- Start early and seek help

Collaborative and Social Learning

- Collaborative Learning
 - Learning and completing assignments as a group
- Social Learning
 - Learning from others (known and unknown)
 - Powerful as knowledge if often distributed and computer science allows us to gather, cluster and analyze user interactions
 - Classroom Salon CMU project funded by Gates Foundation and National Science Foundation and Heinz Endowment
 - Classroom Salon uses collective intelligence, knowledge discovery and expert following

Collaborative and Social Learning is 15% in this course

- Collaborative Learning
 - Submitting some portion of the assignments as a group activity
 - Your contribution will depend on Salon activities you perform (more to come)
- Social Learning
 - When you get stuck with code, you can post *parts* of the code to Salon to seek feedback (up to 1 time per assignment)
 - You will also become part of a group and you can post *parts* of the code to Salon to seek feedback from your group (unlimited postings)
 - You will also receive points for responding/clarifying others posts
 - You will also receive points when others find your responses helpful.
- 15% point distribution is based on
 - Active participation
 - Helpfulness of your comments

Important things to Remember

- All/most assignments are due Sunday at 11:59PM
- Any assignment that is more than 24 hours late automatically loses 50% of the grade
- No assignment is graded if it is more than 2 days late.
- All assignments are individual. You are encouraged to discuss things with others, but final product is "yours"
- Academic honesty is very important
- Start early, so you don't get into unpleasant situations
- Seek help from instructor and TA
 - Guna and Rafee
- All course material from
 - http://www.cs.cmu.edu/~ab/15-121N11
 - http://blackboard.andrew.cmu.edu

Review

Programming Languages

- Language types
 - Natural
 - Artificial
- Language components
 - Syntax Form of language constructs
 - Semantics their meaning
- Examples of High-Level Languages?

Java versus C

Compilers

- A compiler translates _____programs into target programs
- Programs are written with static (compile-time) and dynamic (run-time) constructs
- Target programs are machine executable and specifies run-time computations
- The most important function of a compiler is static code checking and analysis
- Compilers also "optimize" the user code for faster execution

Java Compiler

- Java is an _____language
- javac command creates the _____file that contains java byte code
- Java byte code can be interpreted by machine
- We run java programs with > _____
- Where Driver is the class file that contains main() method

A Java Program

```
import java.io.*
public class Driver {
  public static void main() {
    System.out.println("hello world");
  }
}
```

Question: Describe the components of this program

Understanding Memory

Memory

- The smallest measurable unit of memory is 1 bit, holding a zero(o) or one(1)
- A byte is typically defined as smallest storage necessary to hold one ASCII character
- One byte is 8 bits and ______ different characters can be represented using one byte
- ASCII table has _____ characters

Sizes of Basic Data Types in Java

- char _____
- Short _____, int _____, long _____
- float _____, double _____
- Data sizes could vary based on machine

ASCII Characters

Some examples of ASCII characters are

Express them in binary

Memory

- Memory is cheap
 - RAM
 - Flash
 - Hard Drives
 - •
- Measures of memory
 - 1 kilobyte = 1K bytes = bytes
 - 1 megabyte = 1MB = bytes
 - 1 Gigabyte = 1GB = bytes
 - 1 Terabyte = 1TB = bytes
 - 1 PetaByte = 1PB = bytes
 - 1 exabyte = 1 XB = bytes

Storage Facts

- Digital Video
 - 23.6 Mbytes per second
 - How big is a movie that is 2 hrs long?
- Digital Audio
 - 176.4 kbytes/second
 - How big is your favorite MP3 song?

So how big is an exabyte?

- It can hold
 - Your whole life as a video
 - All the words you have spoken
 - All the books and articles you read
 - All the songs you have ever heard
 - All movies you have ever watched
- How much it will cost in 2050
 - Prize of a Pizza?

Question

• I have a machine with 4GB of RAM. The OS gives my program a maximum stack memory of 4MB. Can I store 1,000,000 integers in a simple array?

Understanding Data Structures

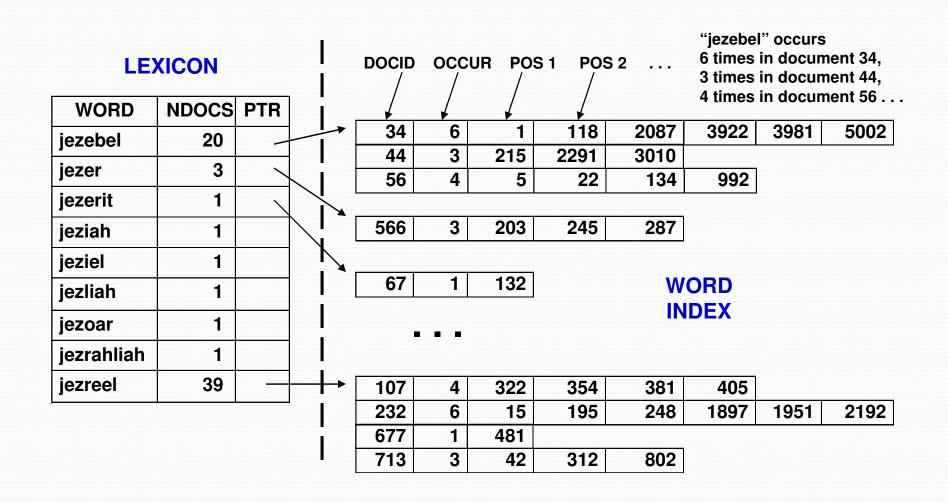
Why Data Structures

- Data must be organized efficiently to speed up the programs
 - Eg: imagine searching linearly through an entire array versus searching data based on some heuristic
- Why is Google so fast?
 - Because it organizes the data for faster retrieval
 - It may sometimes cache data for faster search

Google's server farm



Inverted Files for Multiple Documents



Next

- We will discuss Arrays, basic java IO and Strings
- Assignment 1 will be released soon
 - Start early
 - Submit individual component by no later than Friday July 1, 11:59PM