

C S 2C: ADVANCED DATA STRUCTURES & ALGORITHMS IN C++

Foothill College Course Outline of Record

Heading	Value
Effective Term:	Summer 2023
Units:	4.5
Hours:	4 lecture, 2 laboratory per week (72 total per quarter)
Prerequisite:	C S 2B.
Advisory:	One of the following: ENGL 1A or 1AH or ESLL 26.
Degree & Credit Status:	Degree-Applicable Credit Course
Foothill GE:	Area V: Communication & Analytical Thinking
Transferable:	CSU/UC
Grade Type:	Letter Grade (Request for Pass/No Pass)
Repeatability:	Not Repeatable

Student Learning Outcomes

- The successful student will be able to write and incorporate balanced trees, hash tables, directed graphs and priority queues in his or her software.
- The successful student will be able to analyze the time complexity of a variety of algorithms and data structure access techniques and choose the best algorithm and/or data structure for the project at hand.

Description

Systematic treatment of advanced data structures, algorithm analysis and abstract data types in the C++ programming language. Coding topics include the development of ADTs from scratch, building ADTs on top of the STL templates, vectors, lists, trees, maps, hashing functions and graphs. Concept topics include searching, big-O time complexity, analysis of all major sorting techniques, top down splaying, AVL tree balancing, shortest path algorithms, minimum spanning trees and maximum flow graphs.

Course Objectives

The student will be able to:

- Implement a user-defined vector abstract data type (ADT) and its associated iterator from scratch, and compare the performance to the built-in C++ standard template library (STL) vector.
- Implement a user-defined linked-list ADT and its associated iterator from scratch, and compare the user-defined performance to the built-in C++ STL list.
- Build stack, queue and sparse matrix ADTs using C++ vectors and lists.
- Compute the big-O, little-o, omega and theta time complexity of search and sort algorithms.
- Define asymptotic behavior and perform empirical benchmarks to compare brute-force techniques with divide-and-conquer strategies.

- Analyze the basic algorithms of a general tree ADT.
- Use object-oriented programming (OOP) to create alternative implementations of binary search trees in C++ and verify or compare the logN behavior of each.
- Describe the advantages of balanced trees and analyze the performance of AVL trees.
- Write code using the STL that realizes a splay tree using either top-down or bottom-up splaying.
- Define linear probing, quadratic probing and open addressing as used in the hash tables ADT.
- Design a priority queue using heaps in C++.
- Analyze, classify and measure the main non-NlogN sorts and write a clear report of the results.
- Implement a C++ Quicksort and at least one other NlogN sort and compare the results as the number of data items approaches infinity.
- Define indirect sorting and explain why it is particularly important in C++.
- Create a graph data structure using OOP techniques and the STL container classes, and implement shortest path, minimum spanning tree and maximum flow problems.
- Describe common applications for each data structure studied in the course.
- Arrive at a strategy for selecting the right data structure for the job.

Course Content

- The Vector ADT
 - Implementing a vector from scratch
 - Using STL
- Linked List ADT
 - Implementing a linked list from scratch
 - Using STL list and associated iterator
- Applications of Vectors and Lists
 - Sparse matrices
 - Queues
 - Stacks
 - The "subset sum problem" and its solution using vectors and lists
- Time Complexity
 - $O(f)$ order of magnitude
 - $o(f)$ order of magnitude
 - $\theta(f)$ order of magnitude
 - $\omega(f)$ order of magnitude
 - Constant, polynomial, logarithmic, and exponential time complexity
 - NlogN time complexity
 - Improper use of recursion leading to exponential time complexity
- Measuring Asymptotic Behavior
 - Empirical methods of measurement (benchmarking)
 - Brute force vs. strategies that use branching or divide-and-conquer algorithms
 - Timing greedy and depth-first algorithms in graphs
- General Trees
 - Tree nodes, roots, leaves, children and siblings
 - Binary node implementation of a general tree
 - Insertion and deletion in general trees
 - Traversal with recursion
 - Cloning trees
- Searching and Binary Search Trees (BSTs)
 - Ordering condition and structure condition
 - OOP (object-oriented-programming) implementation
 - Time complexity consequence of the divide-and-conquer algorithm in of BSTs

- 4. The alternative "lazy deletion" implementation of tree nodes and its comparative performance to "hard deletion"
- 5. Threaded trees
- 6. Bin heaps
- H. Balanced BSTs 1: AVL Trees
 - 1. Tree height and rebalancing
 - 2. Single and double rotations as the fundamental rebalancing tools
 - 3. Implementing AVL trees by inheriting from a generic BST
- I. Balanced BSTs 2: Splay Trees
 - 1. Splaying
 - 2. Top down vs. bottom-up splaying
 - 3. Implementing splay trees using STL classes
- J. Hashing
 - 1. Hashing functions
 - 2. Separate chaining
 - 3. Linear and quadratic probing
- K. Priority Queues
 - 1. The percolate down operation
 - 2. Bin heap implementation of priority queues
 - 3. Heap sort
- L. Non-NlogN Sorts
 - 1. Insertion sort
 - 2. Shellsort
 - 3. Benchmarking non-NlogN sorts compared with the STL sort
- M. NlogN Sorts
 - 1. Merge sort
 - 2. Heapsort
 - 3. Quicksort
 - 4. Benchmarking NlogN sorts compared with the STL sort
- N. Indirect Sorting
 - 1. What it is
 - 2. Why it's needed in C++
- O. Graph Theory
 - 1. Structures of a graph
 - 2. Nodes, edges and adjacency tables
 - 3. Shortest path algorithms and Dijkstra
 - 4. Minimal spanning tree algorithms and Kruskal
 - 5. Maximum flow graphs and their algorithms
- P. The use of data structures in common applications areas
 - 1. Math
 - 2. Physics
 - 3. Chemistry
 - 4. Biology
 - 5. Astronomy
 - 6. Business and finance
 - 7. Internet
- Q. Strategies for selecting the right data structure
 - 1. Storage allocation and use memory for the data structure
 - 2. Deciding on a user-defined or built-in data structure
 - 3. How declaration models, binding and visibility affect different ADT implementations

Lab Content

- A. Implementation of time-intensive algorithms on various data types
 - 1. Design and implement an algorithm whose execution time and/or memory requirements grow significantly when data size increases
 - 2. Use generics (AKA templates), which are a universal tool in advanced data structures, in some aspect of the algorithm
 - 3. Demonstrate that the algorithm adapts correctly when the generic that you use is applied to at least two distinct underlying data types

- 4. Document the results of the algorithm when the program is applied to different sized data and different underlying data types
- B. Using linked-list ADTs to optimize for size-varying or space-sensitive data types
 - 1. Demonstrate the ability to use programming language-supplied linked-list structures in a problem that is not easily solved using fixed-size ADTs such as arrays
 - 2. Incorporate generics so as to allow the algorithm to work on various underlying data types
 - 3. Try different sized data for the linked-list and demonstrate that it handles growth properly
 - 4. Summarize the results along with sample program runs
- C. Analyzing time complexity in the lab
 - 1. Implement an assigned algorithm after first predicting its time complexity (linear, quadratic, NlogN, etc.)
 - 2. Run the algorithm on various sized data sets, recording times
 - 3. Describe the largest size data set that the computer can handle without running out of memory or taking an unreasonable amount of time
 - 4. Compare the expected growth rate with the observed growth rate
- D. Demonstrating competence with binary search trees
 - 1. Implement a binary search tree (BST) from scratch, or make significant assigned adjustments to an existing BST data structure supplied by your instructor
 - 2. Use recursion as appropriate for some of the BST methods
 - 3. Demonstrate that the class works on various underlying base type by use of generic specialization
 - 4. Supply runs and report on expected vs. observed time complexity
- E. Demonstrating competence with balanced trees
 - 1. Implement a some assigned balanced tree algorithm (such as AVL, splay or red-black) from scratch, or make significant adjustments to an existing balanced tree algorithm supplied by your instructor
 - 2. Use recursion as appropriate for some of the balanced tree methods
 - 3. Demonstrate that the class works on various underlying base type by use of generic specialization
 - 4. Supply runs and report difference between balanced tree times and simple BST times
- F. Incorporating hash tables into programs
 - 1. Produce a lab that creates or modifies a hash table and hashing function
 - 2. Write a client that tests out the hash table on various data
 - 3. Using a large data set, demonstrate that near-constant time access is produced by the hashing function and hash table
 - 4. Supply runs and report results with varying sized data sets
- G. Analysis of a single sort algorithm
 - 1. Implement a single sort algorithm as directed by the instructor
 - 2. Experiment with coding adjustments to try to improve the performance
 - 3. Compare the known time complexity of that algorithm with what you observe using increasingly larger data sets
 - 4. Attempt to explain any discrepancies in the expected vs. observed growth rate of the sort algorithm
- H. Analysis of multiple sort algorithms
 - 1. Implement multiple sort algorithms, at least two of which involve Shellsort and Quicksort
 - 2. Experiment with coding adjustments to try to improve the performance on any one of them to see if you can beat the fastest of the algorithms
 - 3. Time the algorithms on very small to very large data sets
 - 4. Report on which algorithms work better on small sets, and which on large sets
- I. Writing projects that use graph theory

1. Implement a generic (AKA template) that will realize/specialize a graph of any underlying data type, either from scratch or using code provided by your instructor
2. Write one of the common algorithms for graphs: shortest path, maximum flow or minimum spanning tree
3. Discuss the problems that arise when debugging labs which involve data structures as complex as graph theoretic algorithms
4. Devise a reasonable output for displaying graphs and supply samples with your program runs

Special Facilities and/or Equipment

- A. Access to a computer laboratory with C++ compilers.
- B. Website or course management system with an assignment posting component (through which all lab assignments are to be submitted) and a forum component (where students can discuss course material and receive help from the instructor). This applies to all sections, including on-campus (i.e., face-to-face) offerings.
- C. When taught via Foothill Global Access on the Internet, the college will provide a fully functional and maintained course management system through which the instructor and students can interact.
- D. When taught via Foothill Global Access on the Internet, students must have currently existing email accounts and ongoing access to computers with internet capabilities.

Method(s) of Evaluation

Methods of Evaluation may include but are not limited to the following:

Tests and quizzes

Written laboratory assignments which include source code, sample runs and documentation

Final examination

Method(s) of Instruction

Methods of Instruction may include but are not limited to the following:

Lectures which include motivation for syntax and use of the C++ language and OOP concepts, example programs, and analysis of these programs

Online labs (for all sections, including those meeting face-to-face/on campus), consisting of:

1. A programming assignment webpage located on a college-hosted course management system or other department-approved internet environment. Here, the students will review the specification of each programming assignment and submit their completed lab work
2. A discussion webpage located on a college-hosted course management system or other department-approved internet environment. Here, students can request assistance from the instructor and interact publicly with other class members

Detailed review of programming assignments which includes model solutions and specific comments on the student submissions

In-person or online discussion which engages students and instructor in an ongoing dialog pertaining to all aspects of designing, implementing and analyzing programs

When course is taught fully online:

1. Instructor-authored lecture materials, handouts, syllabus, assignments, tests, and other relevant course material will be delivered through a college-hosted course management system or other department-approved internet environment

2. Additional instructional guidelines for this course are listed in the attached addendum of CS department online practices

Representative Text(s) and Other Materials

Weiss, Mark Allen. [Data Structures and Algorithm Analysis in C++, 4th ed.](#) 2013.

This text is a classic text in the field and is used by many universities in both undergraduate and graduate classes on the subject of data structures.

Types and/or Examples of Required Reading, Writing, and Outside of Class Assignments

A. Reading

1. Textbook assigned reading averaging 30 pages per week.
2. Reading the supplied handouts and modules averaging 10 pages per week.
3. Reading online resources as directed by instructor though links pertinent to programming.
4. Reading library and reference material directed by instructor through course handouts.

B. Writing

1. Writing technical prose documentation that supports and describes the programs that are submitted for grades.
2. Writing specifications using prose to connect natural English language to the formulaic programming languages.

Discipline(s)

Computer Science