C S 3C: ADVANCED DATA STRUCTURES & ALGORITHMS IN PYTHON

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 3B.
Advisory:	Demonstrated proficiency in English by placement via multiple measures OR through an equivalent placement process OR completion of ESLL 125 & ESLL 249.
Degree & Credit Status:	Degree-Applicable Credit Course
Foothill GE:	Non-GE
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

A systematic treatment of advanced data structures, algorithm analysis, and abstract data types in the Python programming language, intended for computer science majors as well as non-majors and professionals seeking advanced Python experience. Coding topics include large program software engineering design, multi-dimensional arrays, string processing, primitives, compound types, and allocation of instance and static data. Data structure concept topics include dynamic memory, inheritance, polymorphism, hierarchies, recursion, linked-lists, stacks, queues, trees, hash tables, and graphs. Algorithm 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:

a. Demonstrate a working knowledge of data abstraction through various data types, data structures, and built-in Python classes, and choose the appropriate data structure to model a given problem
b. Design, implement, test, and debug intermediate-level Python programs that use each of the following fundamental programming constructs: user interaction and communication, string processing, numeric computation, simple I/O, arrays, and the Python built-in classes
c. Analyze the basic algorithms of a general tree ADT
d. Use object-oriented programming (OOP) to create alternative implementations of binary search trees in Python, and verify or compare the logN behavior of each
e. Compute the big-O, little-o, omega, and theta time complexity of selected algorithms
f. Define asymptotic behavior and perform empirical benchmarks to compare brute-force techniques with divide-and-conquer strategies
g. Analyze the basic algorithms of a general tree ADT
h. Use object-oriented programming (OOP) to implement a binary search tree in Python and verify its logN behavior
i. Describe the advantages of balanced trees and analyze the performance of AVL trees
j. Write code that realizes a Splay Tree using either top-down or bottom-up splaying
k. Define linear probing, quadratic probing, and open addressing as used in the hash tables ADT
l. Design a priority queue using heaps in Python
m. Implement a Quicksort and at least one other NlogN sort and compare the results as the number of data items approaches infinity
n. Define indirect sorting and explain when it is needed
o. Create a graph data and implement shortest path, minimum spanning tree, and maximum flow problems

Course Content

a. The notion of run-time complexity
    i. Exhaustive search vs. Heuristic search
    ii. Search space pruning
    iii. Analysis of recursive solutions which are not appropriate due to exponential time complexity vs. iterative polynomial complexity
b. Formal treatment of time complexity
    i. O(n) order of magnitude
    ii. o(n) order of magnitude
    iii. theta(n) order of magnitude
    iv. omega(n) order of magnitude
    v. Constant, polynomial, logarithmic, and exponential time complexity
    vi. NlogN time complexity
    vii. Improper use of recursion leading to exponential time complexity
c. Measuring asymptotic behavior
    i. Empirical methods of measurement (benchmarking)
    ii. Brute-force vs. strategies that use branching or divide-and-conquer algorithms
    iii. Timing greedy and depth-first algorithms in graphs
d. General trees
    i. Tree nodes, roots, leaves, children, and siblings
    ii. Binary node implementation of a general tree
    iii. Insertion and deletion in general trees
    iv. Traversal with recursion
e. Searching and Binary Search Trees (BSTs)
    i. Ordering condition and structure condition
    ii. Object-oriented programming (OOP) implementation

   iii. Time complexity consequence of the divide-and-conquer algorithm in of BSTs
f. Searching and BSTs
   i. Ordering condition and structure condition
   ii. OOP implementation
   iii. Time complexity of BSTs
   iv. Lazy deletion of tree nodes
   v. Threaded trees
   vi. Bin heaps
g. Balanced BSTs 1: AVL trees
   i. Tree height and rebalancing
   ii. Single and double rotations as the fundamental rebalancing tools
   iii. Implementing AVL trees by inheriting from a generic BST
h. Balanced BSTs 2: Splay trees
   i. Splaying
   ii. Top down vs. bottom-up splaying
   iii. Implementing splay trees using Python
i. Hashing
   i. Hashing functions
   ii. Separate chaining
   iii. Linear and quadratic probing
j. Priority queues
   i. The percolate down operation
   ii. Bin heap implementation of priority queues
   iii. Heap sort
k. Non-NlogN sorts
   i. Insertion sort
   ii. Shellsort
   iii. Benchmarking non-NlogN sorts compared with standard library sort
l. Indirect sorting
   i. What it is
   ii. When it is needed
m. Graph theory
   i. Structures of a graph
   ii. Nodes, edges, and adjacency tables
   iii. Shortest path algorithms and Dijkstra
   iv. Minimal spanning tree algorithms and Kruskal
   v. Maximum flow graphs and their algorithms

Lab Content

a. Implementation of time-intensive algorithms on various data types
   i. Design and implement an algorithm whose execution time and/or memory requirements grow significantly when data size increases
   ii. Use generics (a.k.a. templates), which are a universal tool in advanced data structures, in some aspect of the algorithm
   iii. Demonstrate that the algorithm adapts correctly when the generic that you use is applied to at least two distinct underlying data types
   iv. 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
   i. 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
   ii. Incorporate generics so as to allow the algorithm to work on various underlying data types
   iii. Try different sized data for the linked-list and demonstrate that it handles growth properly
   iv. Summarize the results along with sample program runs
c. Analyzing time complexity in the lab
   i. Implement an assigned algorithm after first predicting its time complexity (linear, quadratic, NlogN, etc.)
   ii. Run the algorithm on various sized data sets, recording times
   iii. Describe the largest size data set that the computer can handle without running out of memory or taking an unreasonable amount of time
   iv. Compare the expected growth rate with the observed growth rate
d. Demonstrating competence with binary search trees
   i. Implement a binary search tree (BST) from scratch, or make significant assigned adjustments to an existing BST data structure supplied by your instructor
   ii. Use recursion as appropriate for some of the BST methods
   iii. Demonstrate that the class works on various underlying base type by use of generic specialization
   iv. Supply runs and report on expected vs. observed time complexity
e. Demonstrating competence with balanced trees
   i. Implement an 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
   ii. Use recursion as appropriate for some of the balanced tree methods
   iii. Demonstrate that the class works on various underlying base type by use of generic specialization
   iv. Supply runs and report difference between balanced tree times and simple BST times
f. Incorporating hash tables into programs
   i. Produce a lab that creates or modifies a hash table and hashing function
   ii. Write a client that tests out the hash table on various data
   iii. Using a large data set, demonstrate that near-constant time access is produced by the hashing function and hash table
   iv. Supply runs and report results with varying sized data sets
g. Analysis of a single sort algorithm
   i. Implement a single sort algorithm as directed by the instructor
   ii. Experiment with coding adjustments to try to improve the performance
   iii. Compare the known time complexity of that algorithm with what you observe using increasingly larger data sets
   iv. Attempt to explain any discrepancies in the expected vs. observed growth rate of the sort algorithm
h. Analysis of multiple sort algorithms
   i. Implement multiple sort algorithms, at least two of which involve Shell sort and quicksort
   ii. 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
   iii. Time the algorithms on very small to very large data sets

iv. Report on which algorithms work better on small sets, and which on large sets

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

Special Facilities and/or Equipment

1. The college will provide access to a computer laboratory with Python interpreters.
2. The college will provide a 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.
3. 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.
4. 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 Python 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

Cormen, Thomas H.. Introduction to Algorithms, 4th ed.. 2022.

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

a. Reading
   i. Textbook assigned reading averaging 30 pages per week
   ii. Reading the supplied handouts and modules averaging 10 pages per week
   iii. Reading online resources as directed by instructor though links pertinent to programming
   iv. Reading library and reference material directed by instructor through course handouts
b. Writing
   i. Writing technical prose documentation that supports and describes the programs that are submitted for grades
   ii. Writing specifications using prose to connect natural English language to the formulaic programming languages

Discipline(s)

Computer Science