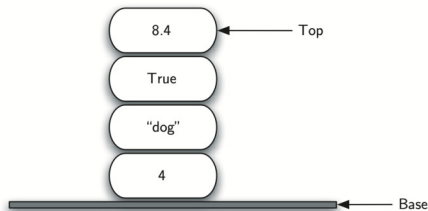


Historical Origins of Data Structures and Algorithms

For Introduction to Algorithms and Data Structure in Python,
CSC 162 Summer 2021

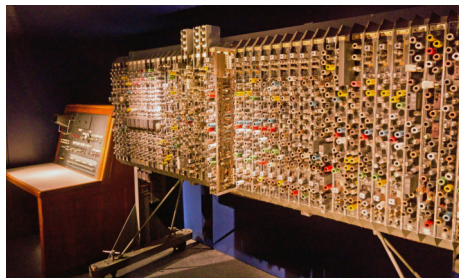
Instructor Alan Beadle, hbeadle@cs.rochester.edu

Stacks



- ▶ First proposed as a computer data structure in 1955 by Klaus Samelson and Friedrich L. Bauer of Technical University Munich
- ▶ Original use was for sequential formula translation (compilers, essentially)

⁰Image from Problem Solving with Algorithms and Data Structures using Python By Brad Miller and David Ranum, Luther College



However, Alan Turing had described¹ the use of a stack for subroutine returns by 1946, using the terms 'bury' and 'un-bury'

¹Report by Dr. A.M. Turing on proposals for the development of an Automatic Computing Engine (ACE): Submitted to the Executive Committee of the NPL in February 1946

²"Pilot Ace" Completed 1950. CPU frequency about 1 MHz. One hundred twenty eight 32-bit words; later expanded to 352 words

⁰Image from https://commons.wikimedia.org/wiki/File:Pilot_ACE3.jpg (CC 3.0 license)

- ▶ Sources also mention that Konrad Zuse's Z4¹ supported subroutines in 1945, but only one at a time (not a stack)

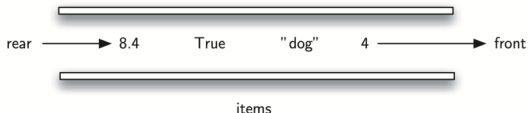


¹“Zuse Z4” Completed 1945. Partially electrical, partially mechanical. CPU frequency about 40 Hz. Sixty-four 32-bit memory locations.

¹Image from https://commons.wikimedia.org/wiki/File:Zuse-Z4-Totale_deutsches-museum.jpg (CC 2.5 License)

- ▶ Polish notation math inspired the later use of a stack as a data structure
- ▶ Polish notation: $(+ 2 2)$
- ▶ Reverse Polish notation: $(2 2 +)$
- ▶ This parenthesis-free notation was invented in 1920 by (Polish) mathematician Jan Łukasiewicz

Queues

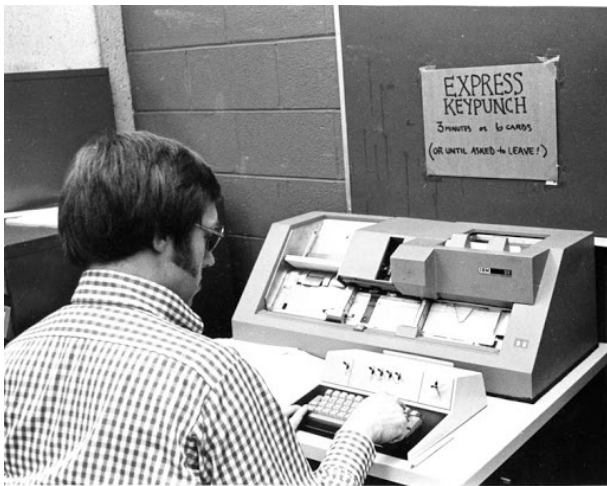


- ▶ Very hard to pinpoint an origin in computing.¹
- ▶ In the early days of batch processing, jobs would wait in line to be run on the machine by the operators
- ▶ This queue was a physical one though, and not really a data structure

¹Perhaps some digital archaeology could find early occurrences in preserved code

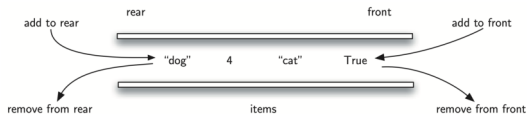
¹Image from Problem Solving with Algorithms and Data Structures using Python By Brad Miller and David Ranum, Luther College

Queues



"A student at the express keypunch in the Vogelback Computing Center, February 28, 1977, 11 p.m." (Northwestern University)

Dequeues



- ▶ Don't know, couldn't find much ¹
- ▶ Probably existed in practice before it was identified as a unique concept

¹Perhaps some digital archaeology could find early occurrences in preserved code

¹Image from Problem Solving with Algorithms and Data Structures using Python By Brad Miller and David Ranum, Luther College

Runtime Analysis

- ▶ Big O notation dates to 1894, introduced by German mathematician Paul Bachman
- ▶ Little-o was introduced in 1909 by German mathematician Edmund Landau
- ▶ Other symbols were introduced over time
- ▶ Were mostly used in mathematics for many years
- ▶ Was only popularized in computer science for the study of algorithms by Donald Knuth in the 1970s!

Donald Knuth

- ▶ Computer scientist, mathematician, Professor Emeritus at Stanford
- ▶ Turing Award winner
- ▶ "father of the analysis of algorithms"
- ▶ author of "The Art of Computer Programming"
- ▶ Creator of the TeX computer typesetting system



¹Image: taken October 25, 2005 by Jacob Appelbaum
<https://commons.wikimedia.org/wiki/File:KnuthAtOpenContentAlliance.jpg>

Recursion

- ▶ The LISP programming language dates to 1958 ¹
- ▶ LISP depends heavily on recursion, so the designers must have already had a strong understanding of it

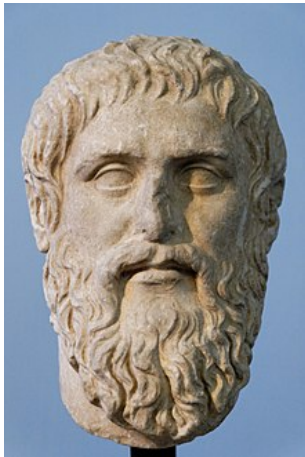


¹LISP was invented by John McCarthy while a research fellow at MIT. He is also considered by many to be the father of AI.

¹Image from <https://engineering.stanford.edu/news/stanford-s-john-mccarthy-seminal-figure-artificial-intelligence-dead-84>

- ▶ If you think about it, recursion is really based on mathematical induction (like inductive proofs)
- ▶ You've got a base case, and a bunch of cases that successively depend on previous ones
- ▶ So how old is mathematical induction?

Recursion



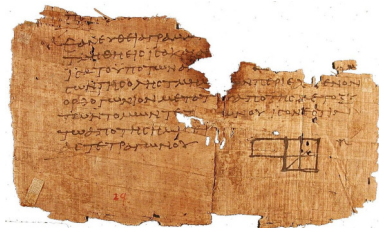
Ancient! Plato's dialogue 'Parmenides' in 370 BCE contained an inductive proof, but the concept was not formalized until much later

¹Image from public domain

https://commons.wikimedia.org/wiki/File:Plato_Silanion_Musei_Capitolini_MC1377.jpg

- ▶ The 4th century BCE philosopher Eubulides enjoyed paradoxes and contradictions
- ▶ Here is an example of a recursive paradox attributed to him:
- ▶ 1,000,000 grains of sand is a heap of sand. A heap of sand minus one grain is still a heap. So 1 grain of sand can be a heap.

Recursion



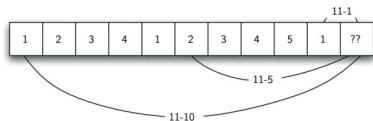
- ▶ Euclid's Elements (300 BCE) also contains an inductive proof to show that every integer is a product of primes
- ▶ However, induction was not a formally defined technique until centuries later ¹

¹(Gersonides (1288–1344), Blaise Pascal (1665))

¹Image from public domain

https://en.wikipedia.org/wiki/File:P._Oxy._I_29.jpg

Dynamic Programming



- ▶ The technique was invented by American mathematician Richard Bellman¹ while he was working for RAND Corporation
- ▶ Bellman first used this term in the 1940's and refined it to match the modern definition by 1953

¹Bellman Equations, Bellman-Ford shortest path algorithm

¹Image from Problem Solving with Algorithms and Data Structures using Python By Brad Miller and David Ranum, Luther College

Searching sorted items

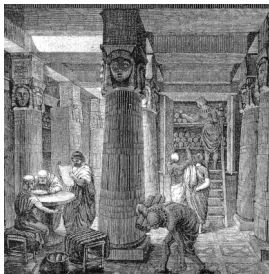
- ▶ An ancient Babylonian¹ tablet from 200 BCE called the Inakibit-Anu tablet contains a sorted table of numbers and their reciprocals

1	𐍪	11	𐍪𐍪	21	𐍪𐍪𐍪	31	𐍪𐍪𐍪𐍪	41	𐍪𐍪𐍪𐍪𐍪	51	𐍪𐍪𐍪𐍪𐍪𐍪
2	𐍪𐍪	12	𐍪𐍪𐍪	22	𐍪𐍪𐍪𐍪	32	𐍪𐍪𐍪𐍪𐍪	42	𐍪𐍪𐍪𐍪𐍪𐍪	52	𐍪𐍪𐍪𐍪𐍪𐍪𐍪
3	𐍪𐍪𐍪	13	𐍪𐍪𐍪𐍪	23	𐍪𐍪𐍪𐍪𐍪	33	𐍪𐍪𐍪𐍪𐍪𐍪	43	𐍪𐍪𐍪𐍪𐍪𐍪𐍪	53	𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪
4	𐍪𐍪𐍪𐍪	14	𐍪𐍪𐍪𐍪𐍪	24	𐍪𐍪𐍪𐍪𐍪𐍪	34	𐍪𐍪𐍪𐍪𐍪𐍪𐍪	44	𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪	54	𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪
5	𐍪𐍪𐍪𐍪𐍪	15	𐍪𐍪𐍪𐍪𐍪𐍪	25	𐍪𐍪𐍪𐍪𐍪𐍪𐍪	35	𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪	45	𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪	55	𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪
6	𐍪𐍪𐍪𐍪𐍪𐍪	16	𐍪𐍪𐍪𐍪𐍪𐍪𐍪	26	𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪	36	𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪	46	𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪	56	𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪
7	𐍪𐍪𐍪𐍪𐍪𐍪𐍪	17	𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪	27	𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪	37	𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪	47	𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪	57	𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪
8	𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪	18	𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪	28	𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪	38	𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪	48	𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪	58	𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪
9	𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪	19	𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪	29	𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪	39	𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪	49	𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪	59	𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪𐍪
10	𐍪	20	𐍪𐍪	30	𐍪𐍪𐍪	40	𐍪𐍪𐍪𐍪	50	𐍪𐍪𐍪𐍪𐍪		

¹Ancient Babylon used a base-60 number system

¹Image from public domain

Searching sorted items



The ancient library at Alexandria (about 300 BCE) is said to have had the world's first library catalog based on alphabetical sorting of the author's names.

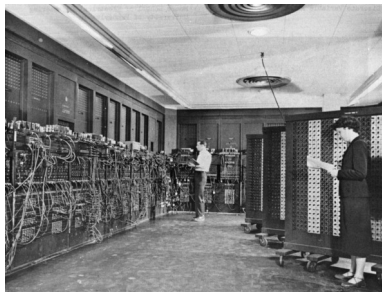
- ▶ Called the Pinakes, compiled by the ancient scholar Callimachus.
- ▶ Nearly 500,000 papyrus scrolls were located in groups based on subject, not author, so finding a work by a particular author required using a catalog.

¹Image from public domain

<https://commons.wikimedia.org/wiki/File:Ancientlibraryalex.jpg>

Searching sorted items

- ▶ In 1946, American Physicist John Mauchly¹ made the first recorded mention of 'binary search', as part of the Moore School lectures²



¹One of ENIAC's designers

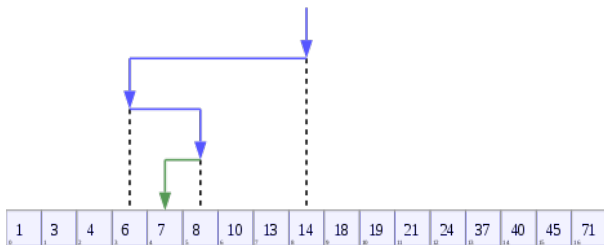
²July 8, 1946, thru August 30, 1946 at the University of Pennsylvania, the first time any computer topics had ever been taught to an assemblage of people

¹Image from public domain

<https://commons.wikimedia.org/wiki/File:Eniac.jpg>

Searching sorted items

- ▶ Until 1960 when Derrick Henry Lehmer¹ published an improved version, all binary searches only worked on arrays having a length of a power of 2! (8, 16, 32, etc)

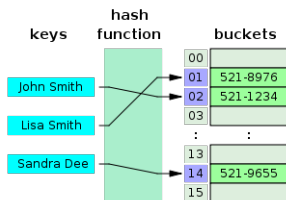


¹One of the first people to develop programs for ENIAC. Also presented at the Moore School Lectures.

¹Image from https://commons.wikimedia.org/wiki/File:Binary_Search_Depiction.svg (CC 4.0 License)

Hashing

- ▶ Donald Knuth notes that a 1953 IBM memo includes the use of a hash function, but the term did not appear in published literature until the late 1960's
- ▶ There are claims that Gene Amdahl, Elaine M. McGraw, Nathaniel Rochester, and Arthur Samuel authored a program that used hashing around this time as well
- ▶ US computer architect Gene Amdahl and USSR computer scientist Andrey Ershov independently invented linear probing



¹Image:

https://commons.wikimedia.org/wiki/File:Hash_table_3_1_1_0_1_0_0_SP.svg
(CC 3.0 License)

Bubble sort

- ▶ No idea if there are older mentions, but a 1956 thesis describes a variant of bubble sort¹
- ▶ The thesis called it a “filing scheme for sorting” (does not use the word “algorithm”)

¹Demuth, Howard B. (1956). Electronic Data Sorting (PhD thesis). Stanford University <https://www.proquest.com/docview/301940891>

- ▶ Donald Knuth writes that insertion sort “was mentioned by John Mauchly as early as 1946, in the first published discussion of computer sorting.” (The Art of Computer Programming, Volume 3, Page 82)

Selection sort

- ▶ The aforementioned 1956 thesis also describes a "selective scheme" for sorting, which we now call selection sort.
- ▶ This is probably not the origin of this algorithm, but at least we know it is old

Mergesort

- ▶ Mergesort was invented by John von Neumann¹ in 1945
- ▶ Hand-written on paper to run on the EDVAC computer
- ▶ Original manuscripts show that "TOP SECRET" was written but then erased from the cover page, since all computer work was classified at this time

¹In addition to his work in computing, von Neumann worked on the Manhattan project

- ▶ Quicksort was developed by British mathematician Tony Hoare in 1959 (published 1961)
- ▶ He is also known for his 1969 proposal of Hoare Logic, an important formal technique for reasoning about (and formally proving!) the correctness of programs

- ▶ Python's default sorting algorithm was developed in 2002 by Tim Peters



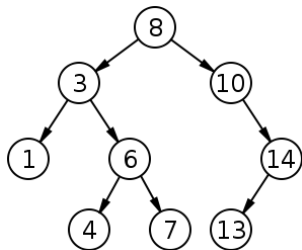
¹Image from

https://commons.wikimedia.org/wiki/File:Python_logo_and_wordmark.svg

- ▶ Trees are a kind of graph, so this goes back to pre-computer graph theory
- ▶ The mathematical term “tree” was coined in 1857 by British mathematician Arthur Cayley
- ▶ Also used when discussing genealogy (“family tree”), so this is an ancient concept

Binary (search) Trees

- ▶ First published in 1960, P.F. Windley, A.D. Booth, A.J.T. Colin, and T.N. Hibbard

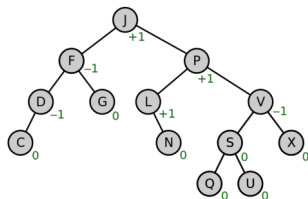


¹Image from public domain

https://commons.wikimedia.org/wiki/File:Binary_search_tree.svg

Balanced Trees

- ▶ The AVL tree was the first balanced tree
- ▶ Published¹ in 1962 by Soviet mathematicians Georgy Adelson-Velsky and Evgenii Landis



¹Adelson-Velsky, Georgy; Landis, Evgenii (1962). "An algorithm for the organization of information". Proceedings of the USSR Academy of Sciences (in Russian).

¹Image from

https://commons.wikimedia.org/wiki/File:AVL-tree-wBalance_K.svg (CC 4.0 License)

Balanced Trees

- ▶ Adelson-Velsky would go on to lead the development an early chess program which would win the first match played between two computer programs in 1966-1967.
- ▶ The competing (and older) program¹ was written by students of John McCarthy at MIT and had been the first chess program to play convincingly.



¹Kotok A. (1962) A Chess Playing Program for the IBM 7090 Computer. (MIT)

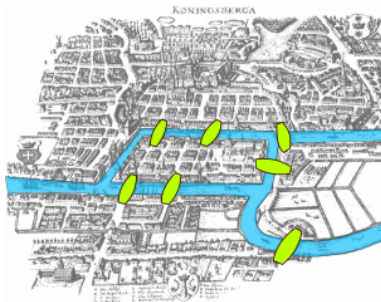
¹USSR M20 Computer. Image from <https://code.google.com/archive/p/m20/> (GPL v2 License)

- ▶ In 1964, the binary heap was introduced by Welsh-Canadian mathematician J. W. J. Williams
- ▶ He also invented heapsort at the same time¹
- ▶ I was not able to find information about the context of his work (Feel free to see what you can find and let me know)

¹Williams, J. W. J. (1964), "Algorithm 232 - Heapsort", Communications of the ACM, 7 (6): 347–348, doi:10.1145/512274.512284

Graphs

- ▶ Graphs in computer science come directly from graph theory in mathematics
- ▶ This field of work dates to Leonhard Euler's 1736 paper on the "Seven Bridges of Königsberg" problem



¹Image from https://en.wikipedia.org/wiki/File:Königsberg_bridges.png (CC 3.0 License)

- ▶ A 19th-century French author named Charles Pierre Trémaux described a form of DFS for solving mazes
- ▶ I cannot find conclusive evidence about when DFS was used in computing, but probably 1950's

Graphs: BFS

- ▶ Described in 1945 by Konrad Zuse for his rejected thesis¹ on the Plankalkül² programming language, finally published in 1972
- ▶ American professor of mathematics and computer science Edward F Moore reinvented it as a maze solving algorithm in 1959

¹Zuse forgot to pay an enrollment fee and apparently did not care that the thesis was rejected

²First high level computer language. It was intended for engineering calculations.

Dijkstra's Algorithm



- ▶ Easy. Edsger Dijkstra 1956 (published 1959)
- ▶ At the National Research Institute for Mathematics and Computer Science in Amsterdam, Netherlands

¹Image from

https://commons.wikimedia.org/wiki/File:Edsger_Wybe_Dijkstra.jpg (CC 3.0 License)

Dijkstra's Algorithm



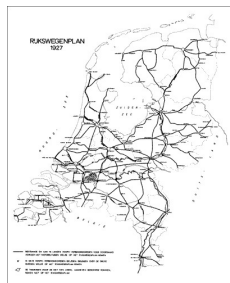
Dijkstra was on a team designing a new computer named ARMAC, and they needed a program to demonstrate the capabilities of the machine.

They wanted to solve a problem that non-computer people would understand.

¹Image from

<http://www-set.win.tue.nl/UnsungHeroes/machines/armac.html>

Dijkstra's Algorithm



Shortest path between 64 cities in the Netherlands (6 bits to identify each city)

Later, rediscovered a similar algorithm (Primm's algorithm) to help with designing the wiring in another new computer

¹Image from public domain

https://commons.wikimedia.org/wiki/File:Rijkswegenplan_1927.jpg