Lecture 2: Overview

BT 3051 – Data Structures and Algorithms for Biology

Karthik Raman
Department of Biotechnology
Bhupat and Jyoti Mehta School of Biosciences
Indian Institute of Technology Madras

Overview

What is an algorithm?

Courtesy: Chris Lacher, Florida State University (CIS 4930)

- Well-defined computational procedure that operates on an input set of values (perhaps empty)
- An algorithm is characterised by the following:
  - Assumptions: Things that must be true before the algorithm is executed

What is an algorithm?

Courtesy: Chris Lacher, Florida State University (CIS 4930)

- Well-defined computational procedure that operates on an input set of values (perhaps empty)
- An algorithm is characterised by the following:
  - Assumptions: Things that must be true before the algorithm is executed
  - Outcomes: Things asserted to be true after the algorithm is executed
  - Proof: "If the assumptions are true and the algorithm is executed, the outcomes are true"
  - Runtime: Time required to execute the algorithm — expressed as asymptotic estimate as a function of input size
  - Run space: Space required to execute the algorithm

What is an algorithm?
Courtesy: Chris Lacher, Florida State University (CIS 4930)

- Well-defined computational procedure that operates on an input set of values (perhaps empty)
- An algorithm is characterised by the following:
  - Assumptions: Things that must be true before the algorithm is executed
  - Outcomes: Things asserted to be true after the algorithm is executed
  - Proof: "If the assumptions are true and the algorithm is executed, the outcomes are true"
  - Runtime: Time required to execute the algorithm — expressed as asymptotic estimate as a function of input size
  - Run space: Space required to execute the algorithm

What is an algorithm?

Courtesy: Chris Lacher, Florida State University (CIS 4930)

- Well-defined computational procedure that operates on an input set of values (perhaps empty)
- An algorithm is characterised by the following:
  - Assumptions: Things that must be true before the algorithm is executed
  - Outcomes: Things asserted to be true after the algorithm is executed
  - Proof: “If the assumptions are true and the algorithm is executed, the outcomes are true”
  - Runtime: Time required to execute the algorithm — expressed as asymptotic estimate as a function of input size
  - Run space: Space required to execute the algorithm

What is an algorithm?

Courtesy: Chris Lacher, Florida State University (CIS 4930)

- Well-defined computational procedure that operates on an input set of values (perhaps empty)
- An algorithm is characterised by the following:
  - Assumptions: Things that must be true before the algorithm is executed
  - Outcomes: Things asserted to be true after the algorithm is executed
  - Proof: "If the assumptions are true and the algorithm is executed, the outcomes are true"
  - Runtime: Time required to execute the algorithm — expressed as asymptotic estimate as a function of input size
  - Run space: Space required to execute the algorithm

What is an algorithm?

Courtesy: Chris Lacher, Florida State University (CIS 4930)

- Well-defined computational procedure that operates on an input set of values (perhaps empty)
- An algorithm is characterised by the following:
  - Assumptions: Things that must be true before the algorithm is executed
  - Outcomes: Things asserted to be true after the algorithm is executed
  - Proof: "If the assumptions are true and the algorithm is executed, the outcomes are true"
  - Runtime: Time required to execute the algorithm — expressed as asymptotic estimate as a function of input size
  - Run space: Space required to execute the algorithm

What is an algorithm?

Courtesy: Chris Lacher, Florida State University (CIS 4930)

- Well-defined computational procedure that operates on an input set of values (perhaps empty)
- An algorithm is characterised by the following:
  - Assumptions: Things that must be true before the algorithm is executed
  - Outcomes: Things asserted to be true after the algorithm is executed
  - Proof: "If the assumptions are true and the algorithm is executed, the outcomes are true"
  - Runtime: Time required to execute the algorithm — expressed as asymptotic estimate as a function of input size
  - Run space: Space required to execute the algorithm

What is a data structure?

- Data structures organise data in the computer, for efficient use by algorithms, e.g. Array
- Abstract data types (ADTs) are theoretical models of data structures defining both the type of data and the operations that can be performed on the data, e.g. Set
- Data structures are implementations of ADTs on a computer

What is a data structure?

- Data structures organise data in the computer, for efficient use by algorithms, e.g. Array
- Abstract data types (ADTs) are theoretical models of data structures defining both the type of data and the operations that can be performed on the data, e.g. Set
- Data structures are implementations of ADTs on a computer

What is a data structure?

- Data structures organise data in the computer, for efficient use by algorithms, e.g. Array
- Abstract data types (ADTs) are theoretical models of data structures defining both the type of data and the operations that can be performed on the data, e.g. Set
- Data structures are implementations of ADTs on a computer

Why study algorithms?

- Algorithms are everywhere!
  - Web search
  - Internet security
  - GPS
  - Image and Video compression/encoding
  - Facebook news feed
  - Recommendation engines
  - NSA Face Recognition
  - Math, Physics, ... Biology!

Why study algorithms?

- Algorithms are everywhere!
  - Web search
  - Internet security
  - GPS
  - Image and Video compression/encoding
  - Facebook news feed
  - Recommendation engines
  - NSA Face Recognition
  - Math, Physics, ... Biology!

Algorithms Are Everywhere …

What algorithm have you used today?

Algorithms are Everywhere ...


BBC
News Sport Weather Capital Future Sho

NEWS TECHNOLOGY

Home World Asia India China UK Business Health Science/Environment Technology Entertainm

16 May 2014 Last updated at 10:29

Share

Algorithm appointed board director

Share Image

Would board meetings be improved by sitting next to an algorithm?

A venture capital firm has appointed a computer algorithm to its board of directors.

The program - called Vital - will vote on whether to invest in a specific company or not.

The firm it will be working for - Deep Knowledge Ventures - focuses on drugs for age-related diseases.

It said that Vital would make its recommendations by sifting through large amounts of data.

The Joy of Algorithms

FROM THE EDITORS

THE JOY OF ALGORITHMS

Francis Sullivan, Associate Editor-in-Chief

THE THEME OF THIS FIRST-OF-THE-CENTURY ISSUE OF *COMPUTING IN SCIENCE & ENGINEERING* IS ALGORITHMS. IN FACT, WE WERE BOLD ENOUGH—AND PERHAPS FOOLISH ENOUGH—TO CALL THE 10 EXAMPLES WE'VE SELECTED "THE TOP 10 ALGORITHMS OF THE CENTURY."

Computational algorithms are probably as old as civilization. Sumerian cuneiform, one of the most ancient written records, consists partly of algorithm descriptions for reckoning in base 60. And I suppose we could claim that the Druid algorithm for estimating the start of summer is embodied in Stonehenge. (That's really hard hardware!)

Like so many other things that technology affects, algorithms have advanced in startling and unexpected ways in the 20th century—at least it looks that way to us now. The algorithms we chose for this issue have been essential for progress in communications, health care, manufacturing, economics, weather prediction, defense, and fundamental science. Conversely, progress in these areas has stimulated the search for ever-better algorithms. I recall one late-night bull session on the Maryland Shore when someone asked, "Who first ate a crab? After all, they don't look very appetizing." After the usual speculations about the observed behavior of sea gulls, someone gave what must be the right answer—namely, "A very hungry person first ate a crab."

The flip side to "necessity is the mother of invention" is "invention creates its own necessity." Our need for powerful machines always exceeds their availability. Each significant computation brings insights that suggest the next, usually much larger, computation to be done. New algorithms are an attempt to bridge the gap between the demand for cycles and the available supply of them. We've become accustomed to gaining the Moore's Law factor of two every 18 months. In effect, Moore's Law changes the constant in front of the estimate of running time as a function of problem size. Important new algorithms do not come along every 1.5 years, but when they do, they can change the exponent of the complexity!

For me, great algorithms are the poetry of computation. Just like verse, they can be terse, allusive, dense, and even mysterious. But once unlocked, they cast a brilliant new light on some aspect of computing. A colleague recently claimed that he'd done only 15 minutes of productive work in his whole life. He wasn't joking, because he was referring to the 15 minutes during which he'd sketched out a fundamental optimization algorithm. He regarded the previous years of thought and investigation as a sunk cost that might or might not have paid off.

Researchers have cracked many hard problems since 1 January 1900, but we are passing some even harder ones on to the next century. In spite of a lot of good work, the question of how to extract information from extremely large masses of data is still almost untouched. There are still very big challenges coming from more "traditional" tasks, too. For example, we need efficient methods to tell when the result of a large floating-point calculation is likely to be correct. Think of the way that check sums function. The added computational cost is very small, but the added confidence in the answer is large. Is there an analog for things such as huge, multidisciplinary optimizations? At an even deeper level is the issue of reasonable methods for solving specific cases of "impossible" problems. Instances of NP-complete problems crop up in attempting to answer many practical questions. Are there efficient ways to attack them?

I suspect that in the 21st century, things will be ripe for another revolution in our understanding of the foundations of computational theory. Questions already arising from quantum computing and problems associated with the generation of random numbers seem to require that we somehow tie together theories of computing, logic, and the nature of the physical world.

The new century is not going to be very restful for us, but it is not going to be dull either!

2 COMPUTING IN SCIENCE & ENGINEERING

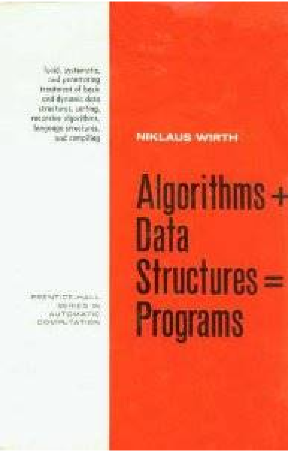
"For me, great algorithms are the poetry of computation. Just like verse, they can be terse, allusive, dense, and even mysterious. But once unlocked, they cast a brilliant new light on some aspect of computing."

— Francis Sullivan

Why Study Data Structures?

- Algorithms + Data Structures = Programs!
- Data structures underlie ∼every algorithm
- The choice of data structure can greatly impact the performance of an algorithm
- *"Bad programmers worry about the code. Good programmers worry about data structures and their relationships."*

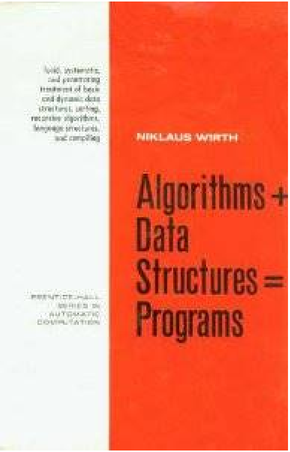
  — Linus Torvalds



Why Study Data Structures?

- Algorithms + Data Structures = Programs!
- Data structures underlie $\sim$every algorithm
- The choice of data structure can greatly impact the performance of an algorithm
- *"Bad programmers worry about the code. Good programmers worry about data structures and their relationships."*

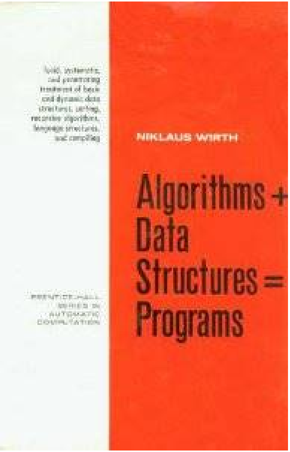
  — Linus Torvalds



Why Study Data Structures?

- Algorithms + Data Structures = Programs!
- Data structures underlie $\sim$every algorithm
- The choice of data structure can greatly impact the performance of an algorithm
- *"Bad programmers worry about the code. Good programmers worry about data structures and their relationships."*

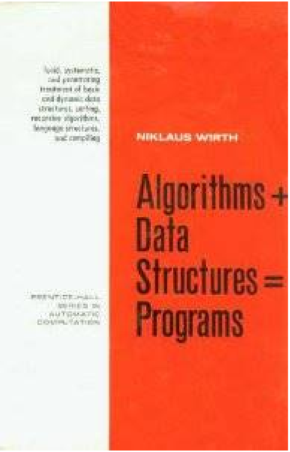
  — Linus Torvalds



Why Study Data Structures?

- Algorithms + Data Structures = Programs!
- Data structures underlie ∼every algorithm
- The choice of data structure can greatly impact the performance of an algorithm
- *"Bad programmers worry about the code. Good programmers worry about data structures and their relationships."*

  — Linus Torvalds



What is the use of algos/data structures in biology?

- Many biological problems cannot be solved without a computer
  - Sequencing large genomes
  - Protein folding (still unsolved)
  - Protein ligand interactions
  - ...
- Modern science is increasingly simulation-driven: "Computational models are replacing math models in scientific inquiry"[a]
- *In silico* **simulations can predict the outcome of real-life experiments!**

[a] http://algs4.cs.princeton.edu/lectures/00Intro.pdf

What is the use of algos/data structures in biology?

- Many biological problems cannot be solved without a computer
  - Sequencing large genomes
  - Protein folding (still unsolved)
  - Protein ligand interactions
  - ...
- Modern science is increasingly simulation-driven: "Computational models are replacing math models in scientific inquiry"[a]
- *In silico* simulations can predict the outcome of real-life experiments!

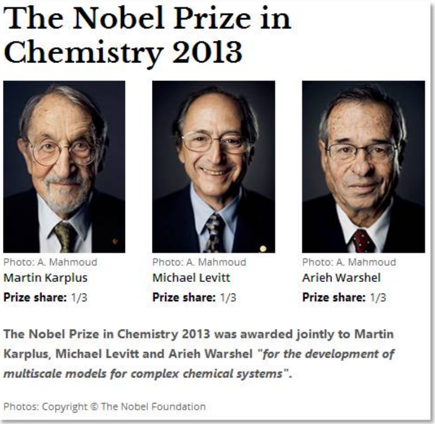
[a] http://algs4.cs.princeton.edu/lectures/00Intro.pdf

What is the use of algos/data structures in biology?

- Many biological problems cannot be solved without a computer
  - Sequencing large genomes
  - Protein folding (still unsolved)
  - Protein ligand interactions
  - ...
- Modern science is increasingly simulation-driven: “Computational models are replacing math models in scientific inquiry”[a]
- ***In silico* simulations can predict the outcome of real-life experiments!**

[a] http://algs4.cs.princeton.edu/lectures/00Intro.pdf

What is the use of algos/data structures in biology?

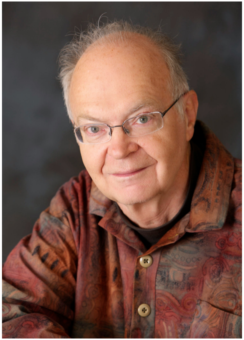


*"In the 1970s, Martin Karplus, Michael Levitt and Arieh Warshel laid the foundation for the powerful programs that are used to understand and predict chemical processes. Computer models mirroring real life have become crucial for most advances made in chemistry today. ... **Today the computer is just as important a tool for chemists as the test tube.** Simulations are so realistic that they predict the outcome of traditional experiments."*

— The Royal Swedish Academy of Sciences (2013)[a]

[a] http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2013/press.html

What is the use of algos/data structures in biology?



"I can't be as confident about computer science as I can about biology. Biology easily has 500 years of exciting problems to work on. It's at that level."

— Donald E Knuth (Turing Award, 1974)

Self-assessment Exercise

- Discuss an algorithm (simplistic overview) underlying something in daily life
  - Bring it up on Piazza

Self-assessment Exercise

- Discuss an algorithm (simplistic overview) underlying something in daily life
  - Bring it up on Piazza

Questions?

Also remember