CS2010: ALGORITHMS AND DATA STRUCTURES

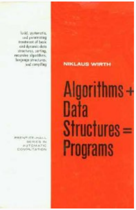
Lecture 1: Module Overview & Introduction

Vasileios Koutavas

School of Computer Science and Statistics

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

→ **Objective:** learn to solve computational problems efficiently



"Algorithms + Data Structures = Programs"

— Niklaus Wirth

→ **Algorithm:** The steps to correctly perform a task that answers a general[1] computational problem

1

- **Algorithm:** The steps to correctly perform a task that answers a general[1] computational problem
  - What is the median age of all people in Ireland?

1

- **Algorithm:** The steps to correctly perform a task that answers a general[1] computational problem
  - What is the median age of all people in Ireland?
  - What is the quickest route from here to the airport?

[1]

- **Algorithm:** The steps to correctly perform a task that answers a general[1] computational problem
  - What is the median age of all people in Ireland?
  - What is the quickest route from here to the airport?

[1] Algorithms answering the above questions should work when the population of Ireland changes & for other destinations!

- **Algorithm:** The steps to correctly perform a task that answers a general[1] computational problem
  - What is the median age of all people in Ireland?
  - What is the quickest route from here to the airport?
- **Data Structures:** The methods to store the information needed for the algorithm.

[1] Algorithms answering the above questions should work when the population of Ireland changes & for other destinations!

WHAT ABOUT EFFICIENCY?

Are these good measures of a program's efficiency?

- → How long it takes the program to run on my laptop
- → How long it takes the program to run on the fastest computer
- → How long it takes the program to run on the slowest computer
- → How long it takes the program to run with the largest input
- → How long it takes the program to run with the smallest input

WHAT ABOUT EFFICIENCY?

Established measure: how well the program **scales** to larger inputs

- → When I double the input size my program takes the same time to run on the same computer (*constant running time*).
- → When I double the input size my program takes twice the time to run on the same computer (*linear running time*).
- → ...

WHAT ABOUT EFFICIENCY?

Established measure: how well the program **scales** to larger inputs

- When I double the input size my program takes the same time to run on the same computer (*constant running time*).
- When I double the input size my program takes twice the time to run on the same computer (*linear running time*).
- ...

We also care about memory needed:

- When I double the input size my program needs the same amount of memory to run (*constant memory space*).
- When I double the input size my program takes twice the amount of memory to run (*linear memory space*).
- ...

In CS2010 (1/2)

The Software Engineer's toolbox

- Learn the most common A&DS that every CS graduate must know.
  - Example algorithms: Merge Sort, Union-Find, Dijstra's Shortest Path Tree, ...

The Software Engineer's toolbox

- Learn the most common A&DS that every CS graduate must know.
  - Example algorithms: Merge Sort, Union-Find, Dijstra's Shortest Path Tree, ...
- Identify which known algorithms/data structures best fit specific problems
  - Example: What is the best algorithm for finding the median? – It depends:[a] QuickSelect, MedianOfMedians, IntroSelect, using SoftHeaps

IN CS2010 (1/2)

The Software Engineer's toolbox

→ Learn the most common A&DS that every CS graduate must know.

  → Example algorithms: Merge Sort, Union-Find, Dijstra's Shortest Path Tree, ...

→ Identify which known algorithms/data structures best fit specific problems

  → Example: What is the best algorithm for finding the median? – It depends:[a] QuickSelect, MedianOfMedians, IntroSelect, using SoftHeaps

→ Learn Abstract Data Types: interfaces of A&DS

IN CS2010 (1/2)

The Software Engineer's toolbox

- → Learn the most common A&DS that every CS graduate must know.
  - → Example algorithms: Merge Sort, Union-Find, Dijstra's Shortest Path Tree, ...
- → Identify which known algorithms/data structures best fit specific problems
  - → Example: What is the best algorithm for finding the median? – It depends:[a] QuickSelect, MedianOfMedians, IntroSelect, using SoftHeaps
- → Learn Abstract Data Types: interfaces of A&DS
- → Practice implementing A&DS

[a] https://www.quora.com/

IN CS2010 (2/2)

The Computer Scientist's toolbox

- Learn to evaluate new algorithms
  - Efficiency: **calculate** the running time and memory usage
    - **how well they scale**
    - **Measuring aspects:** Worst-case, average-case, amortised, experimental performance
    - **Measuring systems:** big-O notation, tilde notation, cost models
  - Correctness: **rigorous testing** and some informal correctness arguments (see Unit Testing, test coverage)

CS2010 Logistics

CS2010 LOGISTICS

`https://www.scss.tcd.ie/Vasileios.Koutavas/teaching/cs2010/`

Example Problem

A software engineer was asked to design an algorithm which will input two **unsorted** arrays of integers, `A` (of size N) and `B` (also of size N), and will output `true` when all integers in `A` are present in `B`.

A software engineer was asked to design an algorithm which will input two **unsorted** arrays of integers, A (of size *N*) and B (also of size *N*), and will output `true` when all integers in A are present in B.

The engineer came up with **two** alternatives. The first is:

```
boolean isContained1(int[] A, int[] B) {
  boolean AInB = true;
  for (int i = 0; i < A.length; i++) {
    boolean iInB = linearSearch(B, A[i]);
    AInB = AInB && iInB;
  }
  return AinB;
}
```

A software engineer was asked to design an algorithm which will input two **unsorted** arrays of integers, A (of size *N*) and B (also of size *N*), and will output `true` when all integers in A are present in B. The engineer came up with **two** alternatives:

The second is:

```
boolean isContained2(int[] A, int[] B) {
  int[] C = new int[B.length];
  for (int i = 0; i < B.length; i++) { C[i] = B[i] }
  sort(C); // heapsort
  boolean AInC = true;
  for (int i = 0; i < A.length; i++) {
    boolean iInC = binarySearch(C, A[i]);
    AInC = AInC && iInC;
  }
  return AinC;
}
```

(a) Calculate the worst-case running time of each of the two implementations.

(b) For each implementation, how much extra memory space is it required to store copies of the elements in `A` and `B`? You should take into account any copies made within the methods `sort`, `linearSearch`, and `binarySearch`.

(c) Find an implementation which is more efficient than both of the engineer's implementation.

Why?

VIDEO

http://www.youtube.com/embed/vSi6YoTPWLw?rel=0&start=8&end=165

WHY?

→ To get a technology job
`http://www.careercup.com`

WHY?

→ To create the "New Google"
`http://en.wikipedia.org/wiki/PageRank`

WHY?

→ To make science

`http://ocw.mit.edu/courses/biological-engineering/`
`20-482j-foundations-of-algorithms-and-computational-techniques-in`

Foundations of Algorithms and Computational Techniques in Systems Biology

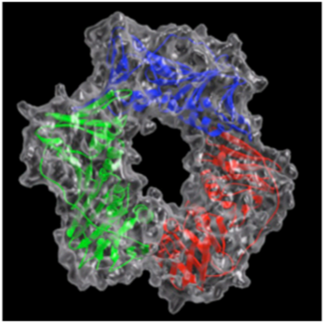
COURSE HOME

SYLLABUS

LECTURE NOTES

ASSIGNMENTS

DOWNLOAD COURSE MATERIALS



The role of the Rad checkpoint complex was inferred from the 3-D structure predicted by comparative modeling at Lawrence Livermore National Laboratory. The Rad complex delays cell division to allow time for DNA repair to take place. (Image courtesy of the U.S. Department of Energy Genomics: GTL Program.)

Instructor(s)
Prof. Bruce Tidor

Prof. Jacob White

MIT Course Number
20.482J / 6.581J

As Taught In
Spring 2006

Level
Graduate

Translated Versions
簡体字

CITE THIS COURSE

WHY?

→ To win big on the stock market
`http://www.theguardian.com/business/2012/oct/21/superstar-traders-lost-magic`

> One theory for the decline of the superstar trader is the rise of the analytical nerd and computerised algorithmic trading. Schmidt says: "The superstars are confronted with a changing market. The punting around is not working. You now need to be either a traditional long-term stock picker, a very short-term person working on algorithms, or a combination of both. There is no future for guys like Coffey.

WHY?

→ To rule the world!
`http://www.theguardian.com/science/2013/jul/01/how-algorithms-rule-world-nsa`


theguardian

News | Sport | Comment | Culture | Business | Money | Life & style

News › Science › Mathematics

How algorithms rule the world

The NSA revelations highlight the role sophisticated algorithms play in sifting through masses of data. But more surprising is their widespread use in our everyday lives. So should we be more wary of their power?

WHY?

→ For fun!
  http://en.wikipedia.org/wiki/Tower_of_Hanoi