CS 240 – Data Structures and Data Management

Module 12: In conclusion

Mark Petrick

Based on lecture notes by many previous cs240 instructors

David R. Cheriton School of Computer Science, University of Waterloo

Fall 2020

References: None

version 2020-11-20 23:24

Outline

- What was the course about
- Final
- Comments?

Outline

- What was the course about
- Final
- Comments?

Course summary: what was this about?

- How to re-organize data
  - (Mergesort), Heapsort, Quicksort, count sort, radix sort
  - ADT Priority Queue, finding maximum, Selection
  - Lower Bounds for a problem, decision trees
- How to manipulate structured data (key-value-pairs)
  - Balanced trees, hashing, tries
  - Special keys: words, integers, points
  - Special situations: biased search-requests, range-queries
- How to manipulate unstructured data (text)
  - Searching
  - Compression
- Some general-purpose techniques
  - Randomization: Shift average-requirements from instances to luck
  - Pre-processing: Initial work pays off in faster queries later
  - External-memory: Huge data warrants different thinking

What to remember after the final

> *"This wasn't training for your first job (programmer). It was training for your second job (code designer/manager of programmers)."*
>
> (Based on a quote by J. Malazita)

- Most problems have many possible solutions. Don't implement the first one you can think of. Can you be faster and/or more space-efficient?
- To save on implementation/experimentation-cost, analyze algorithms on paper first to eliminate obviously bad solutions.
- There isn't always one best solution—the answer to "which algo is best?" is almost always "it depends". Know your input.

Future courses that are related

Required: CS341, Algorithms

- Focus on problem solving, especially for graphs
- Lots more lower bounds, especially NP-hardness.

Optional: (CS341 prereq)

- CS466, Algorithm Design & Analysis
  - Amortized analysis, randomized algorithms, online algorithms, ...
- CS348/448: Databases
  - More complicated queries than search and range
  - Big data $\rightarrow$ external memory becomes important
- CS482: Biological Sequence Analysis
  - Lots more on pattern matching, especially with suffix trees
- CO487: Applied Cryptography
  - Lots more on encoding with a focus on making it secure.

Outline

- What was the course about
- Final
- Comments?

Final exam information

- Date and time schedule by the Registrar's Office $\rightarrow$ see web page
- Final help session? $\rightarrow$ see piazza
- Material covered: modules 01-11
  (exceptions, if any, will be posted on web page and piazza)
- Reference-sheet will be provided and published beforehand.
- Strong emphasis on second half (Hashing to Compression)
- Types of question: various

Outline

- What was the course about
- Final
- Comments?

Any comments?

- We are in unprecidented times and pivoting to online learning has been a change and challenge for both students and instructors. Recent news suggests there is a light at the end of the tunnel. I hope everyone has been safe and healthy throughout the term and I wish everyone all the best as we move into the exam period.