

Class 48: Merge Sort

Held: Monday, 28 April 2008

Summary: We continue our exploration of sorting by considering the applicability of divide-and-conquer to the problem of sorting. We look at one particular divide-and-conquer algorithm, *merge sort*. We explore how the running time for that algorithm varies based on the number of values we are sorting.

Related Pages:

- EBoard.
- Lab: Merge Sort.
- Reading: Merge Sort.

Notes:

- Are there questions on Exam 3?
- EC for attending the stats poster session Thursday at 11:00 in the Science elbow.
- EC for attending Thursday's Thursday extra (the last of the semester).

Overview:

- More efficient sorting techniques.
- Divide and conquer, revisited.
- Merge sort.
- Analyzing merge sort.

Key Ideas of Merge Sort

- Divide and conquer is often a useful design strategy.
- For sorting, natural division is "first half" / "second half".
- What do we do after sorting the two halves? Merge 'em.

An Alternate Implementation

- We can do something very much like merge sort while avoiding the splitting step.
- We start with a list of sorted lists, and repeatedly merge neighboring pairs.
- This technique is slightly easier to implement.
- This technique is much easier to analyze.

The Costs of Merge Sort

- What's the running time? We do approximately $N \cdot \log_2 N$ comparisons.
- The analysis:
 - N steps to merge 2 sorted lists of length $N/2$
 - N steps to merge 4 sorted lists of length $N/4$ into those two sorted lists.
 - N steps to merge 8 sorted lists of length $N/8$ into those four sorted lists.
 - And so on and so forth.
- Can we do better? Not if our sorting is based on comparing values to each other.

Lab

- Do the lab.
- Pause to reflect.

Copyright © 2007-8 Janet Davis, Matthew Kluber, and Samuel A. Rebelsky. (Selected materials copyright by John David Stone and Henry Walker and used by permission.) This material is based upon work partially supported by the National Science Foundation under Grant No. CCLI-0633090. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. This work is licensed under a Creative Commons Attribution-NonCommercial 2.5 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/2.5/> or send a letter to Creative Commons, 543 Howard Street, 5th Floor, San Francisco, California, 94105, USA.