Experiments in computer science

2022/2023

Luís Paquete
University of Coimbra

Experiments in computer science

Contents

- Computer science as a discipline
- Goals in computer science experiments
- Engineering cycle
- Measurements, simulation and analytical modelling

Experiments in computer science

What is computer science as a discipline?

"Computer science is in part a **scientific** discipline concerned with the empirical study of a class of phenomena, in part a **mathematical** discipline concerned with the formal properties of certain classes of abstract structures, and in part a **technological** discipline concerned with the cost-effective design and construction of commercially and socially valuable products."

P. Wegner, *Research paradigms in computer science*. In proc. of the 2nd International Conference on Software Engineering, 322-330, 1976.

Experiments in computer science

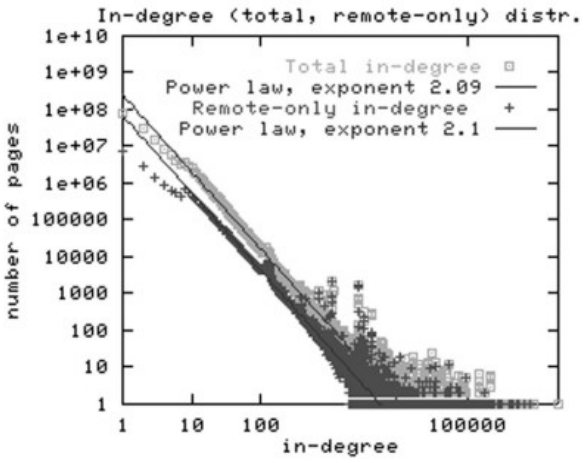
Computer science as an empirical science:

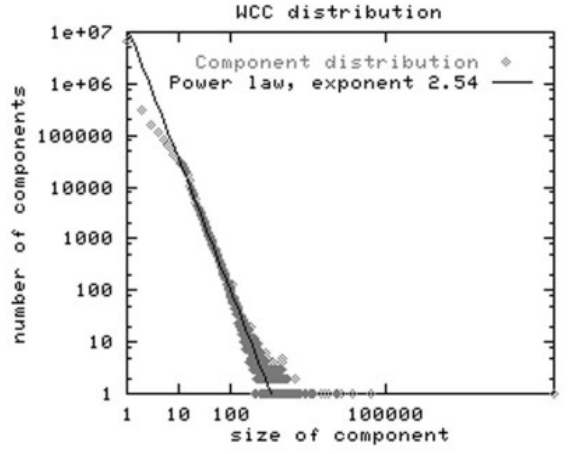
- Data-collection phase: Collect large amount of data concerning phenomena
- Observed uniformities and differences between phenomena lead to classification schemes and to (succint) models and theories on the observed data
- Phenomena in computer science: man-made entities and concepts developed in the computer

Experiments in computer science

Graph structure in the Web
(Broder et al. 2000)

- Analysis on 200 millions pages and 1.5 billion links
- The number of links to individual pages is distributed according to a power law
- The size of weakly connected components also follows a power law
- If a path exists between two nodes, its average length is close to 7





Experiments in computer science

Computer science as mathematics:

- Development of abstractions from a class of phenomena
- Study of properties of the abstraction independently of the phenomena from which they were derived
- Study of algorithms, computational theory, formal methods, information theory, criptography

Experiments in computer science

Quicksort

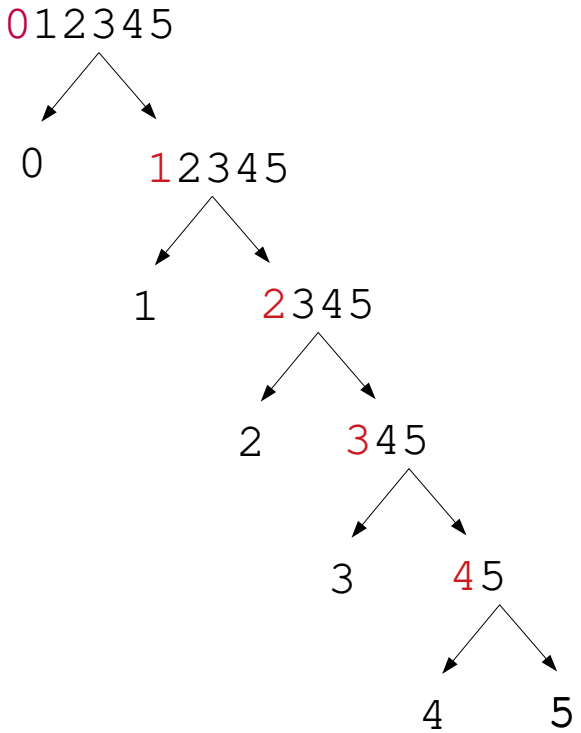
(Hoare 1959)

- Quicksort is a divide-and-conquer algorithm to sort *n* numbers
- It takes O(n^2) time in the worst case if the pivot is always the smallest element in the list

$$T(n) = T(n-1) + O(n)$$

- It takes O(n log n) time on the average case

$$T(n) = 2T(n/2) + O(n)$$



Experiments in computer science

Computer science as an engineering discipline:

- Efficient accomplishment of specific tasks and towards the development of tools that will enable classes of tasks to be accomplished more efficiently
- *Practicing engineer* - takes a sequence of systematic selection or design decisions which narrow down alternative options for accomplishing the task
- *Research engineer* – may use mathematics and/or empirical methodologies for the development of tools for the practicing engineer (but more practically-oriented than a empirical scientist or a mathematician)

Experiments in computer science

Which sorting algorithm to use?

- A server needs to sort (in-place) 1GB of integers for each PageRank query
- The input data arrives *almost* sorted
- Server: 2 Intel Xeon Silver 4210R 2.4G, 10 Cores / 20 Threads, 9.6GT/s, 13.75M Cache, 2 32GB RDIMM, 2933MT/s, 2 480 GB SSD SATA, OS Debian GNU/Linux 10
- Quicksort or merge sort?
- `void qsort(void *base, size_t nitems, size_t size, int (*comp) (const void *, const void*))` or `std::sort()` ?
- `Java.utils.Arrays.sort()` or `Java.utils.Collection.sort()`?

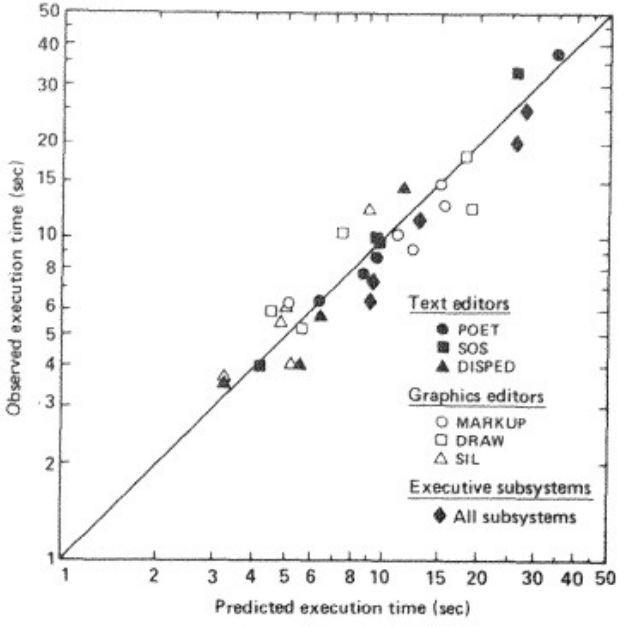
Experiments in computer science

A Keystroke-Level Model for user performance
(Card et al. 1980)

- A model that predicts the time taken by an expert user to perform a given task on a computer
- It takes into account the time to keystroking (*K*), pointing (*P*), homing (*H*), drawing (*D*), as well as the user's mental state (*M*) and system response (*R*)

$$T_e = T_K + T_P + T_H + T_D + T_M + T_R$$

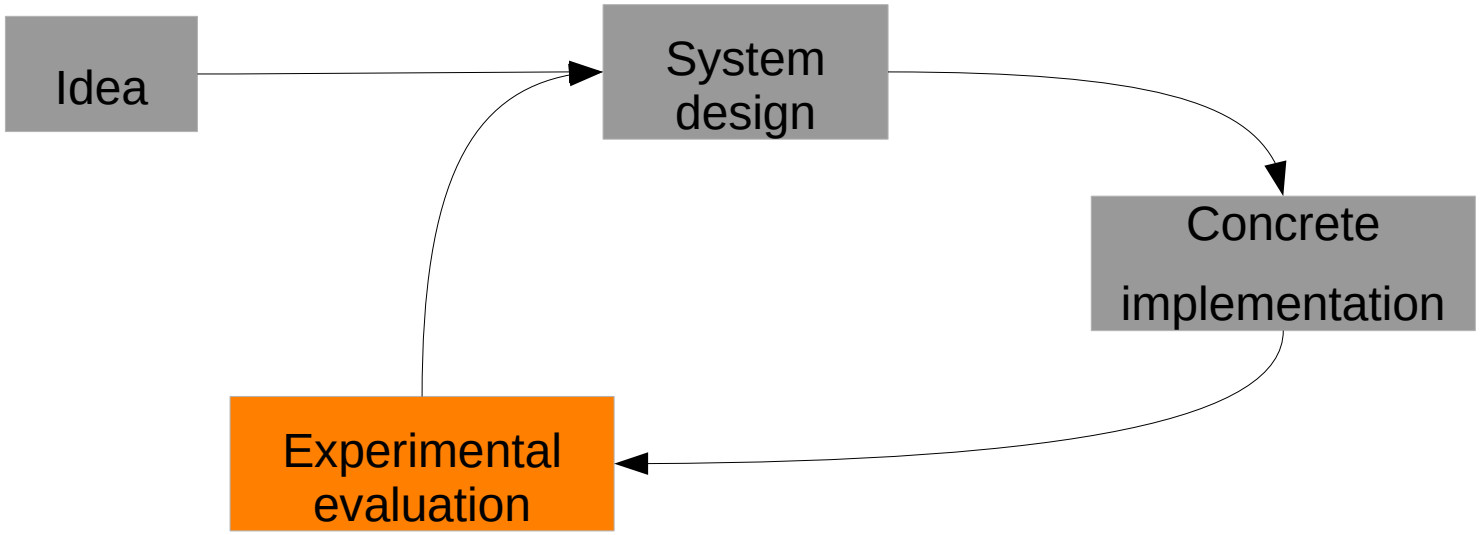
- The model gives a lower bound on the effectiveness of new user interface proposals



Experiments in computer science

Engineering cycle

- **Recall**: Efficient accomplishment of specific tasks and towards the development of tools that will enable classes of tasks to be accomplished more efficiently



Experiments in computer science

Goals in computer science as an engineering discipline:

- **Compare alternatives**: provide quantitative information about which configurations are best under specific conditions
- **Determine the impact of a feature**: determine the impact of adding or removing a specific feature of the system
- **System tuning**: find the set of parameter values that produces the best overall performance
- **Identify relative performance**: quantify the change in performance relative to history or client expectations
- **Performance debugging**: apply tools and analysis techniques to determine why the system is not meeting performance expectations
- **Set expectations**: set expectactions for what a system is actually capable of doing

Experiments in computer science

Solution techniques:

- **Measurements**: it provides the most reliable results since, given the necessary measurement tools, no simplifying assumptions are required. However, direct measurement may not be feasible for some systems.
- **Simulation**: a program that models important features of the system being analysed. It can be easily modified to test different scenarios. It requires simplifying assumptions since it may be impossible to model every detail of the system being studied, which results in less accuracy.
- **Analytical modeling**: mathematical description of the system. It provides quick insight into the overall behaviour of a system or its components. It may require even more simplifying assumptions in order to be tractable.

Experiments in computer science

Characteristic	Solution technique		
	Analytical modeling	Simulation	Measurement
Flexibility	High	High	Low
Cost	Low	Medium	High
Believability	Low	Medium	High
Accuracy	Low	Medium	High

Source: Lilja (2002)

Experiments in computer science

Recap:

- Computer science can be seen as a discipline that uses empirical, mathematical and engineering paradigms
- From engineering perspective, the goals in computer science are: to compare alternatives, to determine the impact of a feature, to perform system tuning, to identify relative performance, to debug performance, to set expectations
- Performance can be assessed with measurements, simulation or analytical modelling. Each one has its own pros and cons.
- **Next**: Measurements

Experiments in computer science

References:

- D.J.Lilja, *Measuring computer performance*, Cambridge University Press, 2002 (see chapter 1)
- P. Wegner, *Research paradigms in computer science*. In proc. of the 2nd International Conference on Software Engineering, 322-330, 1976
- A. Z. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, J.L. Wiener, *Graph structure in the Web*. Computer Networks, 33(1-6): 309-320, 2000
- S. K. Card, T. P. Moran; A. Newell, *The keystroke-level model for user performance time with interactive systems*. Communications of the ACM, 23 (7): 396–410, 1980
- C. A. R. Hoare, *Algorithm 64: Quicksort*, Communications of the ACM, 4 (7): 321, 1961