

The evolution of (searching) data structures

Speaker: Paolo Ferragina *

Abstract

In this talk I will survey over 60 years of Research in algorithms and data structures for searching large collections of data. Starting from the classic work about tries (60), we will make a tour that will show how research and applications mutually inspired each other leading to the recent results on compressed data structures, and even more recently, to the surprising results about learned data structures, in which machine learning and classic data structures interplay together.

References:

- Paolo Ferragina, Roberto Grossi: The String B-tree: A New Data Structure for String Search in External Memory and Its Applications. *J. ACM* 46(2): 236-280 (1999) Paolo Ferragina, Giovanni Manzini: Indexing compressed text. *J. ACM* 52(4): 552-581 (2005)
- Paolo Ferragina: Data Structures: Time, I/Os, Entropy, Joules! *Procs of the European Symposium on Algorithms (ESA)*, 2010
- Paolo Ferragina, Rossano Venturini: Compressed Cache-Oblivious String B-Tree. *ACM Trans. Algorithms* 12(4): 52:1-52:17 (2016)
- Gonzalo Navarro: *Compact Data Structures - A Practical Approach*. Cambridge University Press 2016
- Paolo Ferragina, Ugo Scaiella: Fast and Accurate Annotation of Short Texts with Wikipedia Pages. *IEEE Software* 29(1): 70-75 (2012)
- Paolo Ferragina, Francesco Piccinno, Rossano Venturini: Compressed Indexes for String Searching in Labeled Graphs. *Procs of the WWW Conference*, 2015
- Tim Kraska, Alex Beutel, Ed H. Chi, Jeffrey Dean, Neoklis Polyzotis: The Case for Learned Index Structures. *ACM SIGMOD Conference* 2018

*Speaker Biosketch: Prof. Paolo Ferragina is a full professor in the Department of Computer Science, University of Pisa. His research is mainly devoted to the design, analysis and experimentation of algorithms and data structures for storing, compressing, mining and retrieving information from Big Data. He is currently the President of the 'PhD in Computer Science', hosted by University of Pisa, in collaboration with University of Florence and Siena. He funded and lead the Acube Lab, where they design algorithms for Big Data, with collaborations with companies worldwide: Google, Bloomberg, Tiscali, Yahoo!, ST Microelectronics, ENEL, Basilichi, CERVED, Spazio Dati, etc.