Faculty of Computer Science 2005-2006 Seminar Series

Succinct Indexes and their Applications

By

J. Ian Munro

University of Waterloo

Monday, December 5th, 2005
3:30pm
ITC317

We combine two different techniques, one from data structures the other algorithmic, to solve various problems in databases and information retrieval. The result is an improvement to known solutions in both space and time.

The concept of a succinct data structure is that of encoding information in an amount of space very close to the information theoretic lower bound while still permitting basic operations to be performed in (nearly) constant time. Much of the work in this area has focused on trees and text searches.
The notion of an adaptive algorithm is that of a method that solves easy problems quickly and inherently harder ones more slowly. A natural metric of difficulty could be output (as well as input) size or the length of the shortest proof that our answer is indeed correct. This approach has been applied to convex-hull covers, sorting, set union and intersection. Such algorithms perform "optimally" not only in the worst case among instances of same size, but also among instances of same difficulty, where the definition of the difficulty of an instance depends of the problem and of the analysis.
We combine those techniques to encode efficiently binary relations such as the one associating web-pages references to keywords in the index of a search engine, and to solve conjunctive queries (i.e. Google-like queries) much faster than previous solutions.
Extending these techniques, we propose an index for file systems, and an algorithm to search in it in almost non-deterministic time.

****

STUDENTS ARE ENCOURAGED TO ATTEND

****