

SEMINAR ANNOUNCEMENT

Algorithmic Techniques to search web, xml and text documents

Prof. Paolo Ferragina

Dipartimento di Informatica
Università di Pisa

Abstract

This talk will address two distinct problems. The first, more algorithmic in its nature, concerns with the design of compressed data structures for supporting substring search queries, as the ones that occur in DNA files or text mining applications, and path-based queries, as the ones that occur in XML files. We will introduce the notion of "compressed index" and will survey the most recent advancements on this algorithmic topic that stands at the crossing point of data compression and data structural design.

The second problem addressed in this talk, more IR in its flavor, concerns with a new approach to web search that has been recently introduced by Vivisimo.com, and that has challenged the academic community. Vivisimo is a meta-search engine that organizes on-the-fly the search results into a hierarchy of labeled folders. The hierarchy offers a complementary view to the flat-ranked list of results returned by current search engines. Users can navigate through the hierarchy driven by their search needs, and this turns out to be effective in the presence of informative, polysemous and poor queries.

Unfortunately, the algorithmic ideas behind the functioning of Vivisimo are unknown. Therefore, we will describe in this talk Snaket (<http://snaket.di.unipi.it/>), the first complete and open-source system in the literature that mimics Vivisimo's performance. Then we will propose an approach to personalization of Snaket's results that is fully adaptive, privacy preserving, scalable, and non intrusive for underlying search engines.

Contact Person: Rachele Sprugnoli (tel 0461-405352)

Monday, 5 December 2005, 16:00

**ITC-irst -Sala Edificio Est-
Via Sommarive, 18 (POVO)**