KYBERNETIKA — VOLUME *40* (2004), NUMBER 3, PAGES 381-396

Multidimensional Term Indexing for Efficient Processing of Complex Queries

Michal Krátký; Tomáš Skopal; Václav Snášel

Abstract: The area of *Information Retrieval* deals with problems of storage and retrieval within a huge collection of text documents. In IR models, the semantics of a document is usually characterized using a set of terms. A common need to various IR models is an efficient term retrieval provided via a term index. Existing approaches of term indexing, e. g. the inverted list, support efficiently only simple queries asking for a term occurrence. In practice, we would like to exploit some more sophisticated querying mechanisms, in particular queries based on regular expressions. In this article we propose a multidimensional approach of term indexing providing efficient term retrieval and supporting regular expression queries. Since the term lengths are usually different, we also introduce an improvement based on a new data structure, called *BUB-forest*, providing even more efficient term retrieval.

Keywords: term indexing; complex queries; multidimensional data structures; BUB-forest;

AMS Subject Classification: 68P05; 68P10; 68P20; 14Q15;