UC San Diego
Technical Reports

Title

Flexible and Efficient XML Search with Complex Full-Text Predicates

Permalink

https://escholarship.org/uc/item/3gv5t284

Authors

Amer-Yahia, Sihem
Curtmola, Emiran
Deutsch, Alin

Publication Date

2005-12-12

Peer reviewed

Flexible and Efficient XML Search with Complex Full-Text Predicates

Sihem Amer-Yahia   Emiran Curtmola   Alin Deutsch

Abstract

Recently, there has been extensive research that generated a wealth of new XML full-text query languages, ranging from simple Boolean search to combining sophisticated proximity and order predicates on keywords. While computing least common ancestors of query terms was proposed for efficient evaluation of conjunctive keyword queries by exploiting the document structure, no such solution was developed to evaluate complex full-text queries. We present efficient evaluation algorithms based on a formalization of XML queries in terms of keyword patterns and an algebra which manipulates pattern matches. Our algebra captures most existing languages and their varying semantics and our algorithms combine relational query evaluation techniques with the exploitation of document structure to process queries with complex full-text predicates. We show how scoring can be incorporated into our framework without compromising the algorithms complexity. Our experiments show that considering element nesting dramatically improves the performance of queries with complex full-text predicates.

Request

To obtain a copy of this UCSD technical report please send an email or a letter request to:

Emiran Curtmola
ecurtmola@cs.ucsd.edu

University of California, San Diego - CSE
9500 Gilman Dr.
La Jolla, California 92093

1