

Efficient Multi-Dimensional Query Processing in Personal Information Management Systems

Wei Wang, Christopher Peery, Amélie Marian, Thu D. Nguyen
{ww, peery, amelie, tdnguyen}@cs.rutgers.edu

Technical Report DCS-TR-627
Department of Computer Science, Rutgers University
110 Frelinghuysen Rd, Piscataway, NJ 08854

April 3, 2008

Abstract

The relentless growth in capacity and dropping price of storage are driving an explosion in the amount of information users are collecting and storing in personal information management systems. This explosion of information has led to a critical need for complex search tools to access often very heterogeneous data in a simple and efficient manner. Such tools should provide both high-quality flexible scoring mechanisms and efficient query processing capabilities. In this paper, we focus on indexes and algorithms to efficiently identify the most relevant files that match multi-dimensional queries comprised of relaxed content, metadata, and structure conditions. We also adapted existing top- k query strategies to our specific scenario. Our work is integrated in Wayfinder, an existing fully functioning file system. We perform a thorough experimental evaluation of our file search techniques and show that our query processing strategies exhibit good behavior across all dimensions, resulting in good overall query performance and good scalability.