ScholarWorks@UMass Amherst

OPEN ACCESS DISSERTATIONS

Title

Query-Time Optimization Techniques for Structured Queries in Information Retrieval

Author

Marc-Allen Cartright, *University of Massachusetts Amherst* Follow

Date of Award

9-2013

Document Type

Open Access Dissertation

Degree Name

Doctor of Philosophy (PhD)

Degree Program

Computer Science

First Advisor

James Allan

Second Advisor

W. Bruce Croft

Third Advisor

David Smith

Subject Categories

Artificial Intelligence and Robotics | Computer Sciences

Abstract

The use of information retrieval (IR) systems is evolving towards larger, more complicated queries. Both the IR industrial and research communities have generated significant evidence indicating that in order to continue improving retrieval effectiveness, increases in retrieval model complexity may be unavoidable. From an operational perspective, this translates into an increasing computational cost to generate the final ranked list in response to a query. Therefore we encounter an increasing tension in the trade-off between retrieval effectiveness (quality of result list) and efficiency (the speed at which the list is generated). This tension creates a strong need for optimization techniques to improve the efficiency of ranking with respect to these more complex retrieval models

This thesis presents three new optimization techniques designed to deal with different aspects of structured queries. The first technique involves manipulation of interpolated subqueries, a common structure found across a large number of retrieval models today. We then develop an alternative scoring formulation to make retrieval models more responsive to dynamic pruning techniques. The last technique is delayed execution, which focuses on the class of queries that utilize term dependencies and term conjunction operations. In each case, we empirically show that these optimizations can significantly improve query processing efficiency without negatively impacting retrieval effectiveness.

Additionally, we implement these optimizations in the context of a new retrieval system known as Julien. As opposed to implementing these techniques as one-off solutions hard-wired to specific retrieval models, we treat each technique as a ``behavioral'' extension to the original system. This allows us to flexibly stack the modifications to use the optimizations in conjunction, increasing efficiency even further. By focusing on the behaviors of the objects involved in the retrieval process instead of on the details of the retrieval algorithm itself, we can recast these techniques to be applied only when the conditions are appropriate. Finally, the modular design of these components illustrates a system design that allows improvements to be implemented without disturbing the existing retrieval infrastructure.

Recommended Citation

Cartright, Marc-Allen, "Query-Time Optimization Techniques for Structured Queries in Information Retrieval" (2013). *Open Access Dissertations*. 779.
https://scholarworks.umass.edu/open_access_dissertations/779

Download

DOWNLOADS

Since November 25, 2013