Efficient Prediction of Difficult Keyword Queries over Databases

ABSTRACT:

Keyword queries on databases provide easy access to data, but often suffer from low ranking quality, i.e., low precision and/or recall, as shown in recent benchmarks. It would be useful to identify queries that are likely to have low ranking quality to improve the user satisfaction. For instance, the system may suggest to the user alternative queries for such hard queries. In this paper, we analyze the characteristics of hard queries and propose a novel framework to measure the degree of difficulty for a keyword query over a database, considering both the structure and the content of the database and the query results. We evaluate our query difficulty prediction model against two effectiveness benchmarks for popular keyword search ranking methods. Our empirical results show that our model predicts the hard queries with high accuracy. Further, we present a suite of optimizations to minimize the incurred time overhead.

EXISTING SYSTEM:

- There have been collaborative efforts to provide standard benchmarks and evaluation platforms for keyword search methods over databases. One effort is the data-centric track of INEX Workshop. Queries were provided by participants of the workshop. Another effort is the series of Semantic Search Challenges (SemSearch). The results indicate that even with structured data, finding the desired answers to keyword queries is still a hard task. More interestingly, looking closer to the ranking quality of the best performing methods on both workshops.

PROBLEM DEFINITION:

- Suffer from low ranking quality.
- Performing very poorly on a subset of queries.
PROPOSED SYSTEM:

- We set forth a principled framework and proposed novel algorithms to measure the degree of the difficulty of a query over a DB, using the ranking robustness principle.
- Based on our framework, we propose novel algorithms that efficiently predict the effectiveness of a keyword query.

ADVANTAGES OF PROPOSED SYSTEM:

- Easily mapped to both XML and relational data.
- Higher prediction accuracy and minimize the incurred time overhead.

SYSTEM REQUIREMENTS:

HARDWARE REQUIREMENTS:

- System: Pentium IV 2.4 GHz.
- Hard Disk: 40 GB.
- Floppy Drive: 44 Mb.
- Monitor: 15 VGA Colour.
- Ram: 512 Mb.

SOFTWARE REQUIREMENTS:

- Operating system: Windows XP/7.
- Coding Language: JAVA/J2EE
- IDE: Netbeans 7.4
- Database: MYSQL
REFERENCE: