International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Performance Measurement of Keyword Search Systems

Kaveri A. Dighe[1], M. M. Naoghare[2]

[1, 2]Department of Computer Engineering, Sir Visvesvaraya Institute of Technology, Nashik, Maharashtra, India

Abstract: ***In past few years many relational keyword system have been proposed. But the problem with them is that most of the system are defective or they do not give the exact search results. In this paper we are measuring the performance of all the keyword search systems, doing this will help to choose the correct keyword search system. The analysis of each and every system will be done. In this paper we will also seek the relationship between time needed for execution and factors changed in previous performances. Our analysis indicates that previous factors have less influence on performance.***

Keywords: Performance metrics, evaluation, keyword search, Retrieval system, schema-based, semantic performance.

1. Introduction

The ubiquitous search text box has changed the way people interact with information. Nearly half of all Internet users use a search engine daily, performing 2-3 billion searches . The success of keyword search systems from what it does not require—namely, a specialized query language or knowledge of the underlying structure of the data. Internet users' increasingly demanding for keyword search interfaces for accessing information, and it is common to extend this paradigm to relational data. This extension has been an active area of research throughout the past years We are not aware of any research projects that have changed from proof-of-concept implementations to deployed systems. Despite the significant number of research papers being published in this area, existing empirical evaluations ignore or only partially address many important issues related to search performance. Baidet assert that existing systems have unpredictable performance, which does not determine their usefulness for real world retrieval tasks. This claim has little support in the existing literature, but the failure for these systems to gain a foothold implies that robust, independent evaluation is necessary. In part, existing performance problems may be obscured by experimental design decisions such as the choice of datasets or the construction of query workloads. Consequently, we conduct an independent, evaluation of existing relational keyword search techniques using a publicly available benchmark to ascertain their real-world performance for realistic query workloads.

Keyword search on semi-structured data (e.g., XML) and relational data differs considerably from IR. A discrepancy exists between the data's physical storage and a logical view of the information. Relational databases are normalized to eliminate redundancy, and foreign keys identify related information. Search queries frequently cross these relationships (i.e., a subset of search terms is present in one tuple and the remaining terms are found in related tuples), which forces relational keyword search systems to recover a logical view of the information. The implicit assumption of keyword search—that is, the search terms are related—complicates the search process because typically there are many possible relationships between two search terms. It is almost always possible to include another occurrence of a search term by adding tuples to an existing result. This realization leads to tension between the compactness and coverage of search results.

2. Literature Survey

Baid, I. Rae, J. Li, A. Doan, and J. Naughton[1] Proposed , Keyword search (KWS) systems should return whatever answers they can produce quickly and then provide users with options for exploring any portion of the answer space not covered by these answers. Our basic idea is to produce answers that can be generated quickly as in today's KWS systems, then to show users query forms that characterize the unexplored portion of the answer space. Combining KWS systems with forms allows us to bypass the performance problems inherent to KWS without compromising query coverage. We provide a proof of concept for this proposed approach, and discuss the challenges encountered in building this hybrid system. Finally, we present experiments over real-world datasets to demonstrate the feasibility of the proposed solution [1].

Gaurav Bhalotia, Arvind Hulgeri, Charuta Nakhe, Soumen Chakrabarti S. Sudarshan [2] proposed, BANKS a system which enables keyword-based search on relational databases, together with data and schema browsing. BANKS enables users to extract information in a simple manner without any knowledge of the schema or any need for writing complex queries. A user can get information by typing a few keywords, following hyperlinks, and interacting with controls on the displayed results. BANKS models tuples as nodes in a graph, connected by links induced by foreign key and other relationships. Answers to a query are modeled as rooted trees connecting tuples that match individual keywords in the query. Answers are ranked using a notion of proximity coupled with a notion of prestige of nodes based on in links, similar to techniques developed for Web search [2].

S. Chaudhuri and G. Das [3], With the proliferation of data sources exposed through web interfaces to consumers, simple ways of exploring contents of such databases are of increasing importance. Examples include users wishing to search catalogs of homes, cars, cameras, restaurants, and photographs. One approach that has been explored is to allow users to query such databases in the same ways as they

explore web documents. Thus, it is desirable to be able to use the paradigm of keyword querying and automated result ranking over contents of databases. However, the rich relationships and schema information present in databases makes a direct adaptation of information retrieval techniques inappropriate. This problem has attracted much attention in research as it presents a rich set of challenges from defining semantics of such querying model to developing algorithms that ensure adequate performance [3].

Y. Chen, W. Wang, Z. Liu, and X. Lin[4], give overview of the state of the art techniques for supporting keyword search on structured and semi-structured data, including query result definition, ranking functions, result generation and top-k query processing, snippet generation, result clustering, query cleaning, performance optimization, and search quality evaluation. Various data models will be discussed, including relational data, XML data, graph-structured data, data streams, and workflows. They describe the applications that are built upon keyword search, such as keyword based database selection, query generation, and analytical processing. Finally we identify the challenges and opportunities of future research to advance the field [4].

3. Proposed System

The performance of existing relational keyword search systems is somewhat disappointing, particularly with regard to the number of queries completed successfully in our query workload.

- The objective is to investigate not the underlying algorithms but the overall, end-to-end performance of these retrieval systems.
- 2. To underscores the need for Standardization
- To investigate the effectiveness of these retrieval systems.
- The goal is to investigate the scalability of the search techniques.

As shown in figure 1 is the block diagram for proposed system.

3.1 Block Diagram of the Proposed System

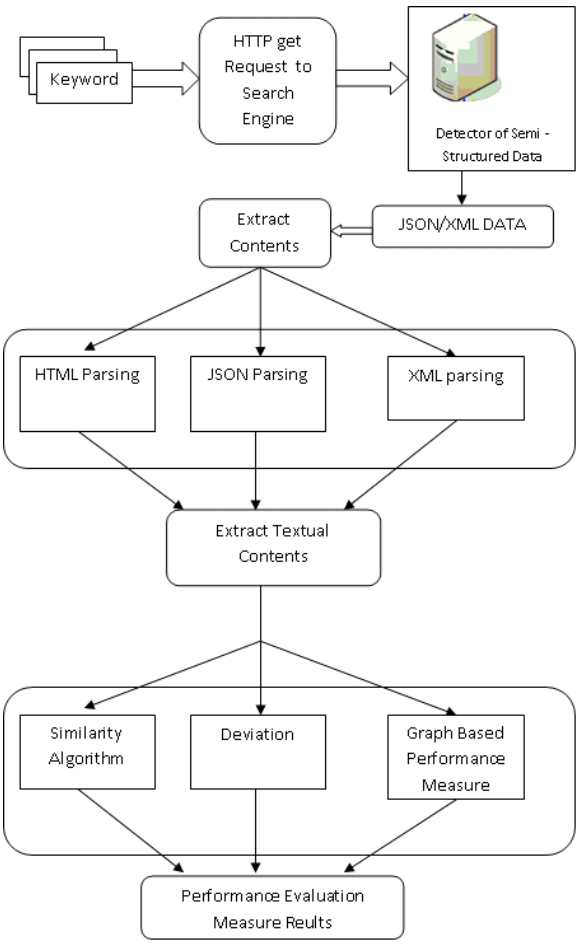

Figure 1: Block diagram of proposed system

4. Conclusion

Unlike many of the evaluations reported in the survey, ours is designed to find not the underlying algorithms but the overall, performance of these retrieval systems. Hence, we favor a realistic query workload instead of a larger workload with queries that are unlikely to be representative. The performance of existing relational keyword search systems is somewhat disappointing, particularly with regard to the number of queries completed successfully in our query workload. We were especially surprised by the number of timeout and memory exceptions that we witnessed. Because our larger execution times might only reflect our choice to use larger datasets, we focus on two concerns that we have related to memory utilization.

References

[1] Joel Coffman, Alfred C. Weaver, “An Empirical Performance Evaluation of Relational Keyword Search Systems”, IEEE Transactions on Knowledge and Data Engineering, vol: 26, Issue: 1) Year: 2014.
[2] Baid, I. Rae, J. Li, A. Doan, and J. Naughton, “Toward Scalable Keyword Search over Relational Data,” Proceedings of the VLDB Endowment, vol. 3, no. 1, pp. 140–149, 2010.
[3] G. Bhalotia, A. Hulgeri, C. Nakhe, S. Chakrabarti, and S. Sudarshan, “Keyword Searching and Browsing in

Databases using BANKS," in Proceedings of the 18th International Conference on Data Engineering, ser. ICDE '02, February 2002, pp. 431–440.
[4] S. Chaudhuri and G. Das, "Keyword Querying and Ranking in Databases," Proceedings of the VLDB Endowment, vol. 2, pp. 1658–1659, August 2009. [Online]. Available: http://dl.acm.org/citation.cfm?id=1687553. 1687622
[5] Y. Chen, W. Wang, Z. Liu, and X. Lin, "Keyword Search on Structured and Semi-Structured Data," in Proceedings of the 35th SIGMOD International Conference on Management of Data, ser. SIGMOD '09, June 2009, pp. 1005–1010.
[6] J. Coffman and A. C. Weaver, "A Framework for Evaluating Database Keyword Search Strategies," in Proceedings of the 19th ACM International Conference on Information and Knowledge Management, ser. CIKM '10, October 2010, pp. 729–738. [Online]. Available: http://doi.acm.org/10.1145/1871437.1871531
[7] B. B. Dalvi, M. Kshirsagar, and S. Sudarshan, "Keyword Search on External Memory Data Graphs," Proceedings of the VLDB Endowment, vol. 1, no. 1, pp. 1189–1204, 2008.
[8] V. Hristidis, L. Gravano, and Y. Papakonstantinou, "Efficient IR-style Keyword Search over Relational Databases," in Proceedings of the 29th International Conference on Very Large Data Bases, ser. VLDB '03, September 2003, pp. 850–861.
[9] H. He, H. Wang, J. Yang, and P. S. Yu, "BLINKS: Ranked Keyword Searches on Graphs," in Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data, ser. SIGMOD '07, June 2007, pp. 305–316.
[10] C. D. Manning, P. Raghavan, and H. Sch¨utze," Introduction to Information Retrieval." New York, NY: Cambridge University Press, 2008.
[11] Tan P-N ,Steinbach M. and Kumar V.,Introduction to Data Mining,Addison Wesley,2006
[12] SoumenChakrabarti ,Morgan Kaufmann;1 edition (November 26,2008),"Data Mining"
[13] "Global Search Market Grows 46 Percent in 2009," http: //www.comscore.com/Press Events/Press Releases/2010/ Global Search Market Grows 46 Percent in 2009, January 2010.
[14] S. E. Dreyfus and R. A. Wagner, "The Steiner Problem in Graphs," Networks, vol. 1, no.3, pp.195–207, 1971. [Online].Available: http://dx.doi.org/10.1002/net.3230010302.
[15] D. Fallows, "Search Engine Use," Pew Internet and American Life Project, Tech. Rep., August 2008, http://www.pewinternet.org/Reports/ 2008/Search-Engine-Use.aspx.

Author Profile

Ms. Kaveri A. Dighehas completed her B.E in Computer Engineering from Pune University and currently pursuing Master of Engineering from SVIT Chincholi, Nashik, India

Prof. M. M. Naoghare has completed her B.E in Computer Engineering from College of Engineering, Badnera, Amravati University and M.E in Computer Science & Engineering from P.R.M.I.T & R, Badnera, Amravati. She is presently working as an Associate Professor in SVIT Chincholi, Nashik, India