

Keyword Search Systems Performance Assessment

Kaveri A. Dighe¹, M. M. Naoghare²

¹Department of Computer Engineering, Sir Visvesvaraya Institute of Technology, Nasik, Maharashtra, India

²Professor, Department of Computer Engineering, Sir Visvesvaraya Institute of Technology, Nasik, Maharashtra, India

Abstract- *In back few years many keyword system have been proposed. But the problem with them is that most of the systems 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 system that already exists will be done. In this paper we will also seek the relationship between time needed for execution and factors changed in previous performances. The analysis shows that previous factors have less influence on performance. The results here indicate that many systems that are existing do not give the satisfactory or needed performance for realistic retrieval tasks. There is need of standardization.*

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

1. Introduction

The ubiquitous search text box has transformed the way people interact with information. Nearly half of all Internet users use a search engine daily, performing in excess of 4 billion searches. The success of keyword search stems from what it does not require—namely, a specialized query language or knowledge of the underlying structure of the data. Internet users increasingly demand keyword search interfaces for accessing information, and it is natural to extend this paradigm to relational data. This extension has been an active area of research throughout the past decade. However, we are not aware of any research projects that have transitioned from proof-of-concept implementations to deployed systems. We posit that the existing, ad hoc evaluations performed by researchers are not indicative of these systems' real-world performance, a claim that has surfaced recently in the literature.

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. Baid et al. assert that existing systems have unpredictable performance, which undermines 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, empirical 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 traditional 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 Proposed, Keyword search (KWS) systems should return the all answers they can produce fast and then provide users with options for exploring any portion of the answer space not covered by these answers. The basic idea is to generate answers that can be generated quickly as in today's keyword search systems, then to show users query forms that characterize the unshown portion of the answer space. Bringing together KWS systems with forms allows us to bypass the performance problems inherent to KWS without compromising query coverage. Here providing a proof of concept for this proposed approach, and discuss the challenges encountered in building this hybrid system. Finally, present experiments over real-world datasets to demonstrate the feasibility of the proposed solution.
- KWS systems should return whatever answers they can produce fastly and then provide users with options for show any part of the answer space not covered by these answers. The basic idea is to get the answers that can be generated fastly as in today's keyword search systems, then to show users query forms that characterize the portion of the answer space. Combining keyword search systems with forms allows us to detail scan the performance problems inherent to KWS without compromising query.
- Gaurav Bhalotia, Arvind Hulgeri, Charuta Nakhe, Soumen Chakrabarti S. Sudarsha proposed, BANKS, a system which enables keyword-based search on relational databases, together with data and schema browsing. BANKS enables users to get the information in a easy manner without any knowledge of the schema or any need for writing tough queries. A user can get information by typing a few keywords, following hyperlinks, and interacting with controls on the shown results. BANKS

models tuples as nodes in a graph form, 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 links, similar to methods developed for Web search.

3. Proposed System

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

- The objective is to investigate not only the available algorithms but the overall, end-to-end performance of these retrieval systems.
- 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 the figure 1, below is the block diagram for proposed system.

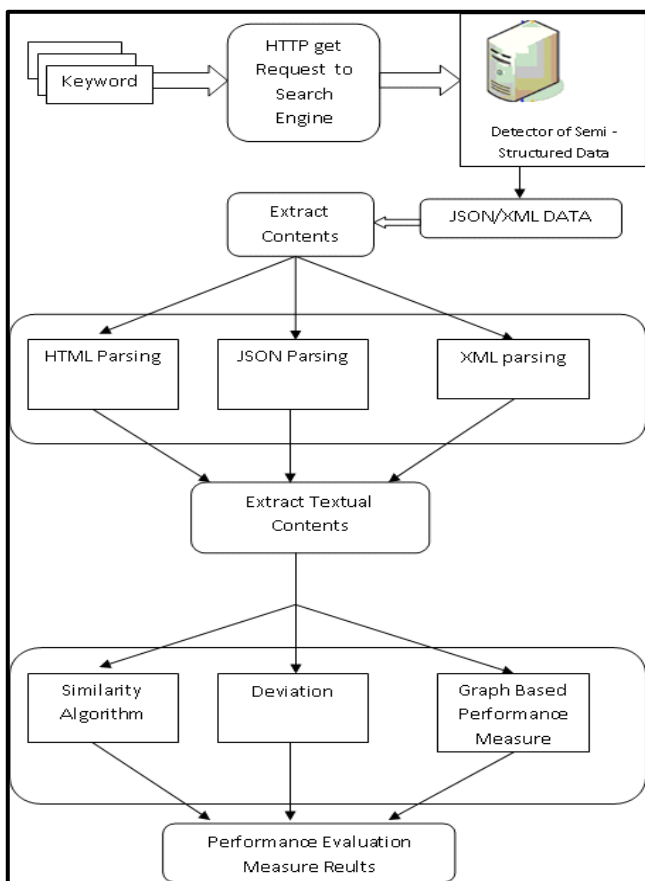


Figure 1: Block Diagram of Proposed System.

4. Snapshots

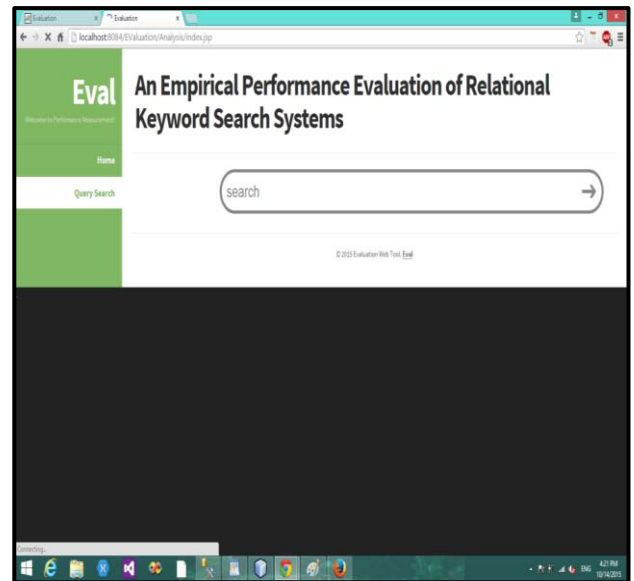


Figure 2: Query Search

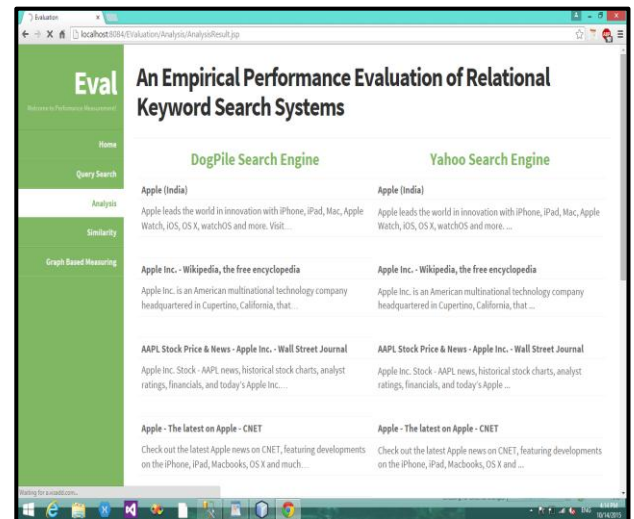


Figure 3: Analysis

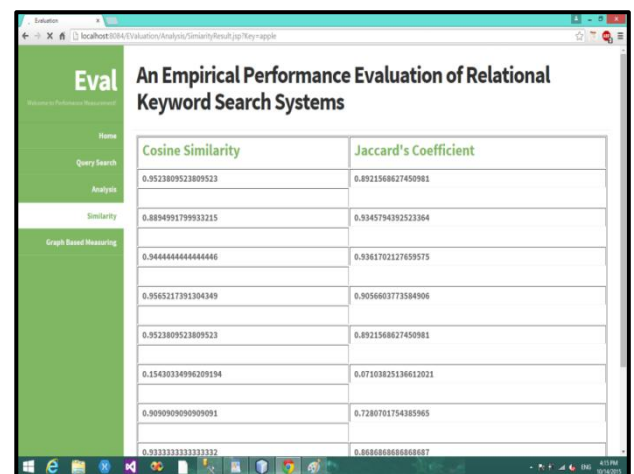


Figure 4: Similarity

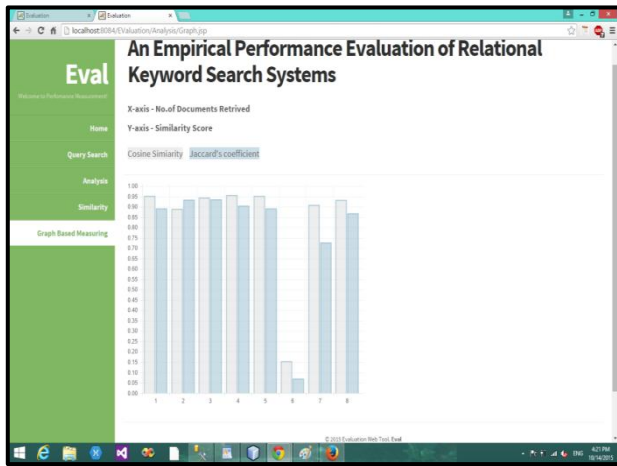


Figure 5: Graph Based Measuring

5. Conclusion

Unlike many of the evaluations reported in the literature, here is designed to investigate not the underlying algorithms but the overall, end-to-end performance of these retrieval systems. Hence, favor a realistic query workload instead of a larger workload with queries that are unlikely to be representative (e.g., queries created by randomly selecting terms from the dataset). Overall, the performance of existing relational keyword search systems is somewhat disappointing, particularly with regard to the number of queries completed successfully in query workload. Given previously published results, were especially surprised by the number of timeout and memory exceptions that we witnessed. Because larger execution times might only react choice to use larger datasets, here focus on two concerns that have related to memory utilization.

6. Acknowledgement

It is a great pleasure to acknowledge those who extended their support, and contributed time and psychic energy for the completion of this project work. At the outset, I would like to thank my project guide Prof. M.M.Naoghare, who served as sounding board for both contents and programming work. Her valuable and skillful guidance, assessment and suggestions from time to time improved the quality of work in all respects. I would like to take this opportunity to express my deep sense of gratitude towards her, for her invaluable contribution in this project work. I am also thankful to Prof. S. M.Rokade, Head of Computer Engineering Department for his timely guidance, inspiration and administrative support without which my work would not have been in process.

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] A. 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>.
- [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. Schütze, —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] Soumen Chakrabarti, 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](http://www.comscore.com/Press%20Events/Press%20Releases/2010/Global%20Search%20Market%20Grows%2046%20Percent%20in%202009), 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](http://www.pewinternet.org/Reports/2008/Search-Engine-Use.aspx).

Author Profile

Ms. Kaveri A. Dighe has 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.