



IMPLEMENTATION OF EFFECTIVE KEYWORD SEARCH ON DATABASE CONTENT

Garikamukkala Surekha¹, P.V.Ravikanth², P.Srinivasulu³

¹M.Tech Student, Dept of CSE, Malineni Lakshmaiah Engineering College, Singarayakonda, AP, India

²Assistant Professor, Dept of CSE, Malineni Lakshmaiah Engineering College, Singarayakonda, AP, India

³Associate Professor, Dept of CSE, Malineni Lakshmaiah Engineering College, Singarayakonda, AP, India

ABSTRACT:

Extending the keyword search concept towards relational data has been a dynamic area of study within database and information retrieval community for the past few years. Abundant techniques were projected, however irrespective of several publications; there remain lack of consistency for assessment of projected search techniques. Our knowledge with conventional methods of search techniques put forward that ad hoc evaluations that come into view in the literature are not enough. They were supported by survey of existing evaluations and by information retrieval community for assessment of retrieval systems. Our earlier efforts have compared methods of relational keyword search regarding search efficiency but do not imagine runtime performance. In our work we put forward most meticulous assessment of empirical performance concerning relational keyword search that has appeared so far in the literature. Altered from numerous evaluations that were reported in literature, ours examine overall, end-to-end performance of methods concerning relational keyword search. Unlike several evaluations that come into view in the literature, our benchmark make use of reasonable data sets and practical queries to inspect the numerous tradeoffs made in design of search techniques. It is the first effort to combine performance and search efficiency in evaluation of huge number of search techniques.

Keywords: *Keyword search, Ad hoc, Information retrieval, Data sets, Queries.*

1. INTRODUCTION:

Keyword search on relational as well as semi-structured data differs noticeably from conventional schemes of information retrieval. Relational databases are regularized to get rid of redundancy, and foreign keys recognize related information. For accessing information, users of internet are demanding interfaces of keyword search and it is likely to broaden this idea towards relational data. This expansion has been a dynamic area of research right the way through the past few years [1]. Regardless of a significant number of efforts were made in this area, no research prototypes have transitioned from proof-of-concept functioning into deployed systems. The lack of technology transfer fixed with discrepancies between existing evaluations specifies a requirement for a systematic, autonomous empirical evaluation of projected search techniques. The hidden assumption of keyword search is that, search terms are associated that make difficult the search process since there are numerous possible associations among search terms. Many techniques of relational keyword search estimate solutions to difficult problems. Researchers as a result make use of empirical assessment to find out benefits

of projected search techniques [2]. Numerous techniques were projected, but regardless of several publications, there remain lack of consistency meant for assessment of projected search techniques. In our work we present wide-ranging empirical performance estimation of methods concerning relational keyword search that has appeared so far in the literature.

2. AN OVERVIEW OF EXISTING WORKS:

The achievement of keyword search appears from a specialized query language or else information of fundamental structure of data. Straightforward implementations of numerous search methods may possibly not extent to databases with several tuples, which forced us decrease their memory footprint. Our experience with conventional methods of search techniques put forward that ad hoc evaluations that come into view in the literature are not enough. Altered from frequent evaluations that come into view, our benchmark makes use of reasonable data sets and practical queries to inspect the numerous tradeoffs made in design of search techniques. It was supported by survey of existing evaluations and those who are well-

known with practices established by information retrieval community for assessment of retrieval systems. Effectiveness metrics are moreover important towards assessment of retrieval systems since not every result is actually applicable to query's fundamental information requirement. Our result point towards that numerous existing search techniques do not make available satisfactory performance for practical retrieval tasks. Existing assessment of relational keyword search methods are ad hoc with minute standardization. Our earlier works compares methods of relational keyword search regarding search efficiency but does not imagine runtime performance [3][4]. Various relational keyword search systems have been available beyond those incorporated in our assessment. Different from many evaluations that were reported in literature, ours examine overall, end-to-end performance of methods concerning relational keyword search. Hence, we support a practical query workload rather than a well-built workload with queries that are not likely to be representative. Evaluations of projected search techniques do not explore significant issues related to performance. Numerous evaluations are also

differing, for reported performance of each system differ to a great extent between several evaluations [5]. Our experimental results question legitimacy of numerous previous evaluations, and we consider our benchmark is more strong and practical regarding retrieval tasks than the workloads employed in other evaluations.

3. AN OVERVIEW OF EVALUATION FRAMEWORK:

Different from numerous evaluations that come into view in the literature, our benchmark make use of reasonable data sets and practical queries to inspect the numerous tradeoffs made in design of search techniques. Our benchmark is only one up to now in literature that assures minimum criteria that was established by community of information retrieval for assessment of retrieval systems. It is the initial work to merge performance and search efficiency in evaluation of huge number of search techniques. Our evaluation benchmark comprises of three data sets such as MONDIAL, IMDb, as well as Wikipedia. The size of datasets varies extensively such as MONDIAL is excess by two orders of magnitude lesser than IMDb data set. Wikipedia lies in between. The schemas as

well as content also differ very much. MONDIAL contain a complex schema whereas IMDb subset has lesser than MONDIAL. Wikipedia moreover has only some relations; however it contains the complete text of articles, which highlight complicated ranking schemes for results. Our data sets approximately span range of data set sizes that were used in other evaluations although IMDb and Wikipedia data sets are subsets of original databases. By means of a database subset possibly overstates effectiveness and efficacy of assessed search techniques. The query workload does not employ real user queries that are extracted from a search engine log for two reasons such as Internet search engine logs do not hold queries in support of data sets not derived from websites and second reason is that, numerous queries are intrinsically unclear and knowing the user's original information requirement is necessary for precise relevance assessments. We separately obtain several information needs for each data set. The gold standard in support of relevance judgments was attained by building SQL queries that has recovered the entire promising applicable results for every information need. The results that are returned by SQL queries were manually

judged for significance where in keeping with definition of relevance recognized by information retrieval community. The applicable results must deal with query's information requirement not just enclose all search terms. We make use of two metrics to compute runtime performance. The first is execution time, which denotes the time elapsed from issuing a query until the termination of algorithm. Our second metric is response time, which was described as time elapsed from issuing query until results have been returned. An overview of MAP across a variety of search methods and data sets was shown in fig1. Effectiveness metrics are moreover significant to evaluation of retrieval systems since not every result is actually applicable to query's fundamental information requirement. There is no precedent from information retrieval community to assess retrieval systems by means of an entirely objective metric since retrieval systems openly answer subjective information requirements [6].

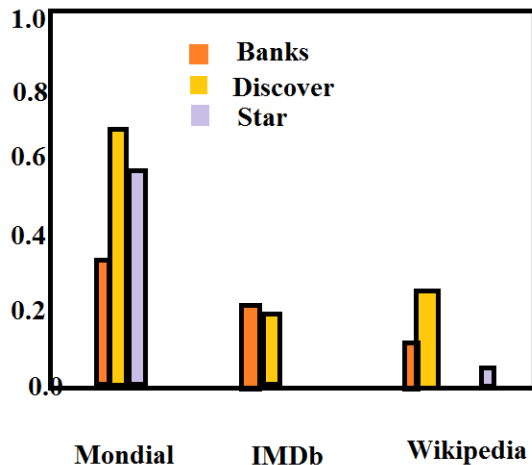


Fig1: An overview of MAP across a variety of search methods and data sets.

4. CONCLUSION:

The concealed supposition of keyword search is that, search terms are associated that make difficult the search process since there are numerous possible associations among search terms. The ability of keyword search appears from a specialized query language or else information of fundamental structure of data. Researchers thus utilize empirical assessment to find out benefits of projected search techniques. In our work we put forward most meticulous assessment of empirical performance concerning relational keyword search that has appeared so far in the literature. Our results point towards that numerous existing search techniques do not make available satisfactory performance for practical retrieval tasks. Existing assessment of relational keyword search methods are ad

hoc with minute standardization. Unlike numerous evaluations that were reported in literature, ours examine overall, end-to-end performance of methods concerning relational keyword search. We support a convenient query workload rather than a well-built workload with queries that are not likely to be representative. Our experimental results question authenticity of several previous evaluations, and we consider our benchmark is more strong and practical regarding retrieval tasks than the workloads employed in other evaluations. Our benchmark is only one up to now in literature that assures minimum criteria that was established by community of information retrieval for assessment of retrieval systems.

REFERENCES

- [1] H. He, H. Wang, J. Yang, and P.S. Yu, "BLINKS: Ranked Keyword Searches on Graphs," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '07), pp. 305-316, June 2007.
- [2] G. Kasneci, M. Ramanath, M. Sozio, F.M. Suchanek, and G. Weikum, "STAR: Steiner-Tree Approximation in Relationship Graphs," Proc. Int'l Conf. Data Eng. (ICDE '09), pp. 868-879, Mar. 2009.
- [3] G. Bhalotia, A. Hulgeri, C. Nakhe, S. Chakrabarti, and S.Sudarshan, "Keyword Searching and Browsing in Databases Using BANKS," Proc.

18th Int'l Conf. Data Eng. (ICDE '02), pp. 431-440, Feb. 2002.

[4] L. Qin, J. Yu, L. Chang, and Y. Tao, "Querying Communities in Relational Databases," Proc. IEEE Int'l Conf. Data Eng. (ICDE '09), pp. 724-735, Mar. 2009.

[5] G. Li, J. Feng, X. Zhou, and J. Wang, "Providing Built-in Keyword Search Capabilities in RDBMS," The VLDB J., vol. 20, pp. 1-19, Feb. 2011.

[6] V. Hristidis and Y. Papakonstantinou, "DISCOVER: Keyword Search in Relational Databases," Proc. 28th Int'l Conf. Very Large Data Base (VLDB '02), pp. 670-681, Aug. 2002.