EXPERIMENTAL RESULTS AND ANALYSIS OF SCALABLE KEY TERMED BASED SEARCH SYSTEM

Bhagyashree Deshmukh[1], Shyam Gupta[2]

[1] *PG Student, Department of Computer Engineering, Siddhant College of Engineering, Pune, Maharashtra State, India*
[2] *Professor, Department of Computer Engineering, Siddhant College of Engineering, Pune, Maharashtra State, India*

ABSTRACT

Current era is internet based applications and data on internet is huge and increasing exponentially. For searching any data on internet various methods used, among which keyword find has confirmed to be a powerful technique to find out and recover information online as facts by a number of Internet search engines. Various methods have been planned and executed, but even though a lot of methods, there is problem that severe lack of normalization for system evaluations. But, many common information management systems do not support the familiar keyword search interface that people now require. Web sites, corporations, and governments all use relational databases to handling information, but keyword search in relational databases is difficult due to data transformations that dismiss redundancy and ensure consistency. The proposed system intended a thorough performance evaluation of keyword search over relational systems. The proposed system involves various other approaches for searching keywords from relational databases. Especially, memory consumption defines various techniques from being scaled from few datasets to thousands of vertices in database. The proposed system also determines the relationship or performance on basis of execution time and other evaluation factors .Thus the system evaluates the performance for keyword search over relational databases and thereby help in examining the relation between the keyword search execution time against the various factors like number of terms, frequency of search. Existing relational keyword search systems require more execution time. So, the performance of existing relational keyword search systems is unsatisfactory. The main purpose behind this paper is to focus on evaluation of keyword search techniques inside relational database and to propose novel schemes which help to improve search efficiency. This implementation of our system provides betters user interaction and organization to classify both the objective behind queries and the expression of this intent.

Keyword: - *keyword search, empirical performance, relational data, relational keyword*

1. INTRODUCTION

Query search using keyword is having more recognition within information search area because when using keyword search techniques, user is not worry regarding query language. Now a day, available search engines give query based keyword search on top of pages. Our paper shows the answer to problem of searching keywords in relational databases. Organizations use relational databases to organize data, but attaining keyword search methods within relational databases is very difficult as relational database supports transformations of data that decrease duplication and protect consistency. Keyword search in relational database permit users to extract data and to traverse the relationships between retrieved information and interface. The proposed system provides answer to conquer issues of current existing systems by datasets, experimental design and generation of query search load. Here we are going to conduct an empirical evaluation of existing relational keyword search methods using accessible standard to figure out their realistic searching performance for real world query search loads. Application of keyword search method on semi structured data differs from relational data greatly from existing IR (Information Retrieval). There exist some inconsistencies in the physical memory storage of data and a logical organization of the

data. Relational databases are design to eliminate the redundancy. It uses foreign keys to figure out relevant information.

Despite the significant research interest in algorithms and ranking strategies for relational key-terms based search, far less attention has focused on the evaluation of these search techniques. One critical factor that has not been previously examined in the literature is the construction of representative query workloads. It presents the most extensive empirical performance evaluation of relational keyword search techniques to appear to date in the literature. The results indicate that many existing search techniques do not provide acceptable performance for realistic retrieval tasks. In particular, memory consumption precludes many search techniques from scaling beyond small data sets with tens of thousands of vertices. It also explores the relationship between execution time and factors varied in previous evaluations; analysis indicates that most of these factors have relatively little impact on performance. In summary, this work confirms previous claims regarding the unacceptable performance of these search techniques and underscores the need for standardization in evaluations standardization exemplified by the IR community.

No such performance evaluation techniques were previously proposed in comparison to various factors such as frequency of terms, number of terms searched. The technique is a client server application where all the encoded data values and codes are fetched from the server. The performance evaluation of an information retrieval system is a decisive aspect for the measure of the improvements in search technology. In this paper, we show experimental results and performance analysis of scalable key termed based search system.

2. LITERATURE SURVEY

Over Relational Data toward Scalable Keyword Search, Baid et al. (2010), in this dissertation suggests neither of proposed search techniques is included in the evaluations because to obtain search results both need an intermediate step [1]. However, how to assess a system that returns a set of forms it is not clear and to select the one(s) compatible to the query needs the user. Over relational data to be implemented in such a way that the system can assurance a reasonable response time it will consider techniques that allow keyword. The main idea is to all the answers are displays the classical keyword search, produce within some time bound. Toward constructing this kind of system this work act as a first step, and for follow-on work that develop the performance and quality of such systems hope that it is a springboard. In general, between the form-based component analyzing the trade-offs and for future work keyword-based component is productive ground.

Using BANKS in Databases Searching and Browsing the Keyword by G. Bhalotia et. al. (2002) popularized for specifying results, the backward expanding search heuristic [2]. On relational databases schema browsing BANKS is a system which enables keyword-based search. In simple aspects without any knowledge of the methods or any need for writing complex queries BANKS enables users to retrieve information. As rooted trees results to a query are modeled connecting tuples that match particular keywords in the query. By finding the graph rearwards BANKS calculate results from vertices that enclose keywords of query. To query and browse relational database with ease BANKS accept users with no familiarity of database systems. The intensions involved in producing relational data on the Web BANKS highly minimize and making it searchable. For answering keyword queries it proposed a framework, and to find query answers incrementally implemented an algorithm. Using academic and bibliographic databases the prototype evaluated in terms of speed and meaningfulness of answers.

Keyword Querying and Ranking in Databases Chaudhary and Das (2009), and Keyword Querying and Ranking in Databases Chen et. al. (2009), the tutorials on keyword find in databases displayed by both. Using simple keywords empowering users to access databases, from the steep research curve of mastering which can relieve the users a structured query language and knowing complex and possibly fast emerging data schemas. The state-of-the-art methods are used, along with query result definition, ranking functions, Result clustering, query cleaning and performance optimization [3]. According to Chen et al., for developing expanded frameworks for evaluating the retrieval and ranking strategies of keyword search grant from the research community are greatly expected on various structured data models [4].

A Framework for Evaluating Keyword based Search Strategies Joel Coffman and Alfred C. Weaver (2010), for relational keyword search methods, presents an appraisal framework. The effectiveness of relational keyword search techniques maximize due to a novel scheme [5]. For relational database they produce the effectiveness of keyword search systems. For structured data the framework carefully considers the unique requirements of keyword search systems and by the IR community follows the traditional definitions of relevance developed. For this field this evaluation framework is the first designed and for evaluating current and future systems provides common workloads. Direct comparison of the effectiveness and performance of different search techniques enables the evaluation framework.

On External Memory Data Graphs Keyword Search B. B. Dalvi, et. al.(2008), investigate using a multi-granular graph representation to reduce I/O costs as compared to data structures residing in virtual memory, keyword search on external memory graphs. For their multi-granular graph representation, which suggests that exploiting this graph representation is non-trivial they only implement the backward search heuristic. The problem of searching graphs addresses by Researchers that are considerably larger than main memory [6]. To save I/O costs and direct the search toward portions of the data graph most likely to yield results existing search techniques can take advantage of the multi-granular representation.

3. PROPOSED SYSTEM

3.1 Problem definition

To develop a technique that efficiently manages memory utilization, swapping data to and from disk and reuse datasets and query workloads to provide greater consistency of results with the help of document retrieval using relational key-terms based search system depending on existing search techniques.

3.2 Main modules

Proposed system consist of following main modules

- Admin Module: Admin is central user of system. He can able to view User Details present in system. He can also upload files for searching from the users.
- User Module: User module is controlled under admin module. He may search files by using query or keywords. After searching gained result files may be view by user. This contains file length, ranking of files and execution time of the files.
- Keyword Search Module: This module searches files by using keywords which have been generated by the admin.
- View Ranking of files: For file ranking purpose we are using graph which can be referred by the chart.
- View File Length and Execution time: All file uploaded by user or admin having length which is read in KB format as well as these are stored it in relational database. From this we can easily calculate execution time of files.

3.3 Plan of system execution

i. Extract Data from IMDB dataset
ii. Load User data, Movie data and Rating data
iii. Process semi-structured data to structured data
iv. Generate Candidate network
v. Generate SQL statement and SQL Queries
vi. Process the correct query
vii. Evaluate the performance
viii. Show graphical presentation

4. RESULTS AND ANALYSIS

Results are shown in the form of charts between execution time required for proposed system and current existing IR systems.

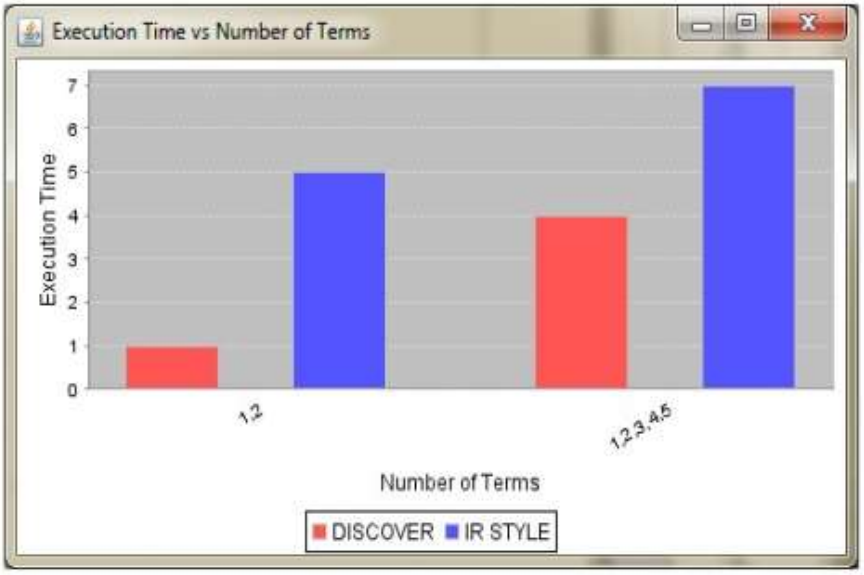

Chart -1: Execution time vs Number of terms

Above chart-1 shows that execution time of proposed system is very less than existing IR systems.

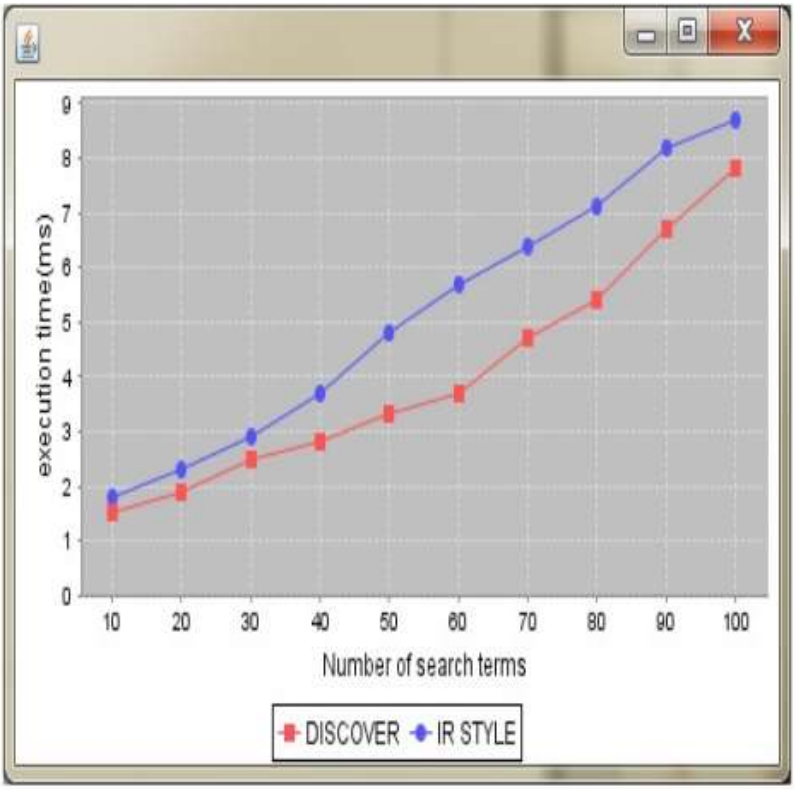

Chart -2: Execution time vs Number of search terms

From chart-2, we got result as proposed system efficient than existing IR system with respect to time.

Fig -1: Precision and Recall

5. CONCLUSIONS

After analyzing results, we concluded that if system has data objects and require high correctness then strong consistency model should be used or else for other applications weaker form of data consistency should be used. In our system, we characterized consistency as a service (CaaS) model and a two-level auditing structure that useful for users for checking whether the cloud service provider (CSP) is offering the agreed consistency, and to investigate the seriousness of the infringements, if any. Using CaaS model, the users can judge the value of cloud services and prefer a right CSP amongst a choice of applicants, for example the smallest amount costly one that still offers sufficient consistency for the users' applications.

Our system introduced a method that proficiently handles memory utilization, swapping data to and from disk with the help of document retrieval using relational key-terms based search. The proposed system reuses datasets and query workloads to provide greater consistency of results depending on which dataset is used. The proposed system explores the relationship between execution time and factors varied in previous evaluations.

6. ACKNOWLEDGEMENT

I would like to express my sincere thanks to my guide Prof. Shyam Gupta, Siddhant college of Engineering for his motivation and useful suggestions which truly helped me in improving the quality and effectiveness of this paper. I take this opportunity to express my thanks to my teacher, family and friends for their encouragement and support and valuable suggestions.

7. REFERENCES

[1]. 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.
[2]. 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.
[3]. S. Chaudhuri and G. Das, Keyword Querying and Ranking in Databases, Proceedings of the VLDB Endowment, vol. 2, pp. 1658-1659, August 2009.
[4]. Y. Chen, W. Wang, Z. Liu, and X. Lin, Keyword Search on Framed and Semi-Framed Data, in Proceedings of the 35th SIGMOD International Conference on Management of Data, ser. SIGMOD 09, June 2009, pp. 1005-1010.
[3]. 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.
[4]. 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.