

# Scalable Key Termed Based Search System: Experimental Results and Analysis

Bhagyashree Deshmukh<sup>1</sup>, Shyam Gupta<sup>2</sup>

<sup>1</sup>PG Student, Department of Computer Engineering, Siddhant College of Engineering, Pune, Maharashtra State, India

<sup>2</sup>Professor, Department of Computer Engineering, Siddhant College of Engineering, Pune, Maharashtra State, India

**Abstract:** *Now a day, internet based applications are very popular and data on internet is massive and growing tremendously. Different methods are used for finding several data on internet, among which keyword find has definite to be a influential method to find out and get better information online as facts by several Internet search engines. Various techniques have been designed and executed, but even if various methods are available, there is difficulty that severe lack of normalization for system evaluations. But, many common information management systems do not maintain the familiar keyword search interface that people at this time need. Web sites, corporations, and governments all use relational databases to handling information, but keyword search in relational databases is complicated by reason of data transformations that dismiss redundancy and ensure consistency. The proposed system planned a thorough performance evaluation of keyword search over relational systems. The proposed system engages different additional approach for searching keywords from relational databases. Particularly, memory consumption describes a variety of procedures from being scaled from few datasets to thousands of vertices in database. The proposed system also decides the relationship or performance on basis of execution time and other evaluation factors. Thus the system calculates the performance for keyword search over relational databases and in that way helps in investigative the relation between the keyword search execution time against the a variety of factors like number of terms, frequency of search. Existing relational keyword search systems need extra execution time. So, the performance of existing relational keyword search systems is undesirable. The major reason behind this paper is to center of attention on evaluation of keyword search techniques inside relational database and to intend new schemes which assist to get better search efficiency. This implementation of our system presents better user interaction and organization to categorize both the objective behind queries and the expression of this intent.*

**Keywords:** keyword search, empirical performance, relational data, relational keyword

## 1. Introduction

Query search using keyword is including extra recognition inside information search area since while with keyword search techniques, user is not concern about query language. Currently, existing search engines provide query based keyword search on top of pages. Our paper demonstrates the solution to difficulty of searching keywords in relational databases. Organizations utilize relational databases to arrange data, but achieving keyword search processes in relational databases is extremely complex as relational database supports transformations of data that reduce duplication and defend consistency. Keyword search in relational database authorize users to remove data and to traverse the relationships between retrieved information and interface. The proposed system offers solution to overcome problems of present 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 various inconsistencies in the physical memory storage of data and a logical organization of the data. Relational databases are intended to remove the redundancy. It utilizes foreign keys to figure out relevant information.

Regardless of the major research interest in algorithms and

ranking strategies for relational key-terms based search, far less consideration has paying attention on the evaluation of these search techniques. One critical factor that has not been before examined in the literature is the construction of representative query workloads. It presents the most general empirical performance evaluation of relational keyword search techniques to appear to date in the literature. The outcomes point out that several existing search techniques do not offer satisfactory performance for realistic retrieval tasks. In particular, memory utilization precludes a lot of search techniques from scaling beyond small data sets with tens of thousands of vertices. It also investigates the connection among execution time and factors varied in previous evaluations; analysis indicates that most of these factors have relatively little impact on performance. In summing up, this work approves earlier claims concerning the undesirable performance of these search techniques and underscores the requirements for consistency in evaluations standardization exemplified by the IR community.

No such performance valuation techniques were before proposed in contrast to a variety of 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 demonstrate 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 their thesis proposed neither of planned search methods are included in the evaluations because to find search results both require an intermediate step [1]. Still, how to assess a system that returns a set of forms it is not understandable and to select the one(s) well-matched to the query desires the user. Over relational data to be applied in such a way that the system can assure a reasonable reply times it will think about methods that permit keyword. The key idea is to all the solutions are displays the classical keyword search, generate within some time limit. In the direction of constructing this type of system this work act as a first step, and for follow-on work that increase the performance and quality of such systems hope that it is a springboard. In common, between the form-based component analyzing the trade-offs and for future work keyword-based component is productive ground.

With use of 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 allows keyword-based search. In simple aspects not including any knowledge of the techniques or any requirement for writing complex queries BANKS allows users to recover information. As rooted trees outcomes to a query are modeled connecting tuples that match particular keywords in the query. By finding the graph rearwards BANKS compute 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 creating relational data on the Web BANKS highly reduce and making it searchable. For answering keyword queries it planned a framework, and to find query answers incrementally implemented an algorithm. With the help of academic and bibliographic databases the prototype calculated 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. With use of 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]. As stated by Chen et al., for developing extended frameworks for calculating the retrieval and ranking tactics of keyword search grant from the research community are greatly predictable on various structured data models [4].

A Framework for Evaluating Keyword based Search approaches Joel Coffman and Alfred C. Weaver (2010), for relational keyword search methods, presents an evaluation framework. The usefulness of relational keyword search methods maximize due to a new scheme [5]. For relational

database they construct the effectiveness of keyword search systems. For structured data the framework cautiously thinks about the exclusive necessities of keyword search systems and by the IR community follows the traditional definitions of significance developed. For this field this evaluation framework is the first planned 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), examine using a multi-granular graph representation to decrease I/O costs as compared to data structures exist in in virtual memory, keyword search on external memory graphs. For their multi-granular graph demonstration, which said that exploiting this graph representation is non-trivial they only execute the backward search heuristic. The trouble 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 build up a system that efficiently handle memory consumption, exchange of data to and from disk and use again datasets and query workloads to offer better consistency of results by way of the assist 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.

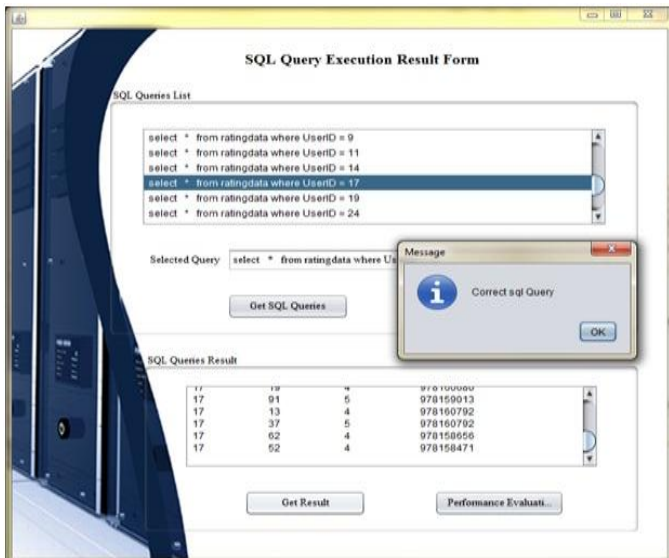
### 3.3 Plan of system execution

- 1) Extract Data from IMDB dataset
- 2) Load User data, Movie data and Rating data
- 3) Process semi-structured data to structured data
- 4) Generate Candidate network

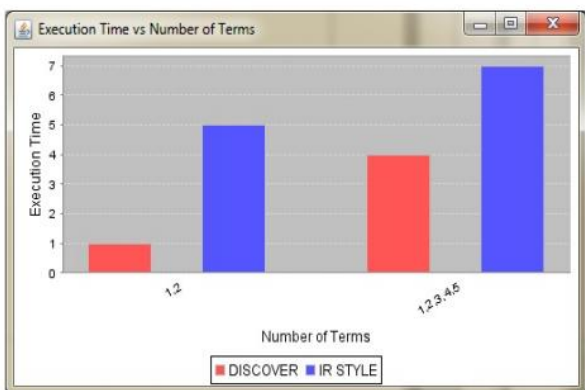
- 5) Generate SQL statement and SQL Queries
- 6) Process the correct query
- 7) Evaluate the performance
- 8) Show graphical presentation

#### 4. Results and Analysis

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

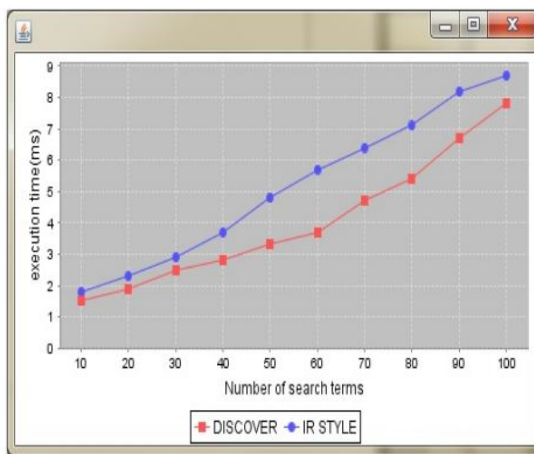


**Figure 1:** Execution of system



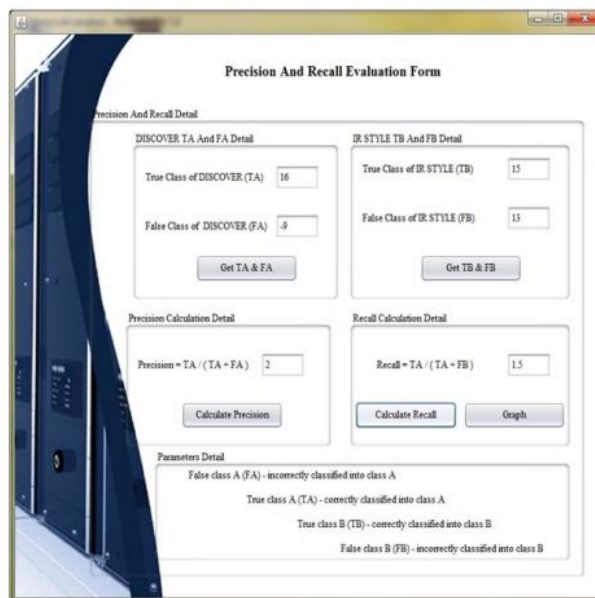
**Figure 2:** Execution time vs Number of terms

Above Figure-2 shows that execution time of proposed system is very less than existing IR systems.



**Figure 3:** Execution time vs Number of search terms

From Figure-3, we got result as proposed system efficient than existing IR system with respect to time.



**Figure 3:** Precision and Recall

#### 5. Conclusions

After examining outcomes of our proposed system, 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 proposed system, we differentiate consistency as a service (CaaS) model and a two-level auditing structure that useful for users for inspection whether the cloud service provider (CSP) is contributing the decided consistency, and to explore the significance of the infringements, if any. With CaaS model, the users can evaluate the value of cloud services and favor a right CSP in the middle of a alternative of applicants, for example the smallest amount costly one that still offers sufficient consistency for the users' applications.

Our system established a technique that expertly deal with memory utilization, exchange of data to and from disk with the help of document retrieval using relational key-terms



based search. The proposed system reprocesses datasets and query workloads to give better consistency of results depending on which dataset is used. The proposed system discovers the relationship between execution time and factors varied in previous evaluations.

## 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 35<sup>th</sup> SIGMOD International Conference on Management of Data, ser. SIGMOD 09, June 2009, pp. 1005-1010.
- [5] J. Coffman and A. C. Weaver, A Framework for Evaluating Database Keyword Search Strategies, in Proceedings of the 19<sup>th</sup> ACM International Conference on Information and Knowledge Management, ser. CIKM 10, October 2010, pp. 729-738.
- [6] 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.