ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

An Empirical Performance Analysis of Relative Keyword Search Techniques

Avinandan Kumar[1], Dipankar Nag[2], G.Michael[3]

UG Scholar, Department of CSE, Bharath Univerity, Chennai-73, Tamil Nadu, India[1,2]

Assistant Professor, Bharath Univerity, Chennai-73, Tamil Nadu, India[3]

ABSTRACT: In the EXISTING SYSTEM the Internet users increasingly demand for the keyword search interface to access the information effectively using the relational data. In the PROPOSED MODEL the relational keyword search were gathered together with different data models. We apply normalization Technique for the database to remove the redundancy of data. We will maintain a Index for every File or Content. File is stored in a Separate space and it's Index is stored for the Related Data. The MODIFICATION in this project ranking the files based on keywords present in the database and number of click done by different users.

KEYWORDS: Keyword search, on-line database, info retrieval, empirical analysis

I. INTRODUCTION

THE ubiquitous search text box has reworked the manner individuals move with info. Nearly 1/2 all net users use an enquiry engine daily [1], acting in way over four billion searches [2]. The success of keyword search stems from what it doesn't require—namely, a specialised search language or data of the underlying structure of the information. net users more and more demand keyword search interfaces for accessing info, and it's natural to increase this paradigm to relative information. This extension has been an energetic space of analysis throughout the past decade. Despite a major range of analysis papers being printed during this space, no analysis prototypes have transitioned from proof-of-concept implementations into deployed systems. the dearth of technology transfer in addition to discrepancies among existing evaluations indicates a desire for an intensive, freelance empirical analysis of projected search techniques. As a part of previous add this space, we have a tendency to created the primary benchmark to judge relative keyword search techniques [3]. This benchmark satisfies calls [4], [5] from the analysis community to standardize the analysis of those search techniques, and our analysis of search effectiveness [3] unconcealed that several search techniques perform comparably despite contrary claims within the literature. throughout our analysis of search effectiveness, we have a tendency to we have a tendency tore shocked by the problem we had looking out our information sets. specifically, easy implementations of the many search techniques couldn't scale to databases with many thousands of topples, that forced USA to put in writing "lazy" versions of their core algorithms and cut back their memory footprint. Even then, we have a tendency to were shocked by the excessive runtime of the many search techniques. different researchers have recently reported similar experiences. Baid etal. state [6], [...] current [keyword search] solutions have unpredictable performance problems. Specifically, whereas the systems turn out answers quickly for several queries, for several others they take associate intolerably durable, or perhaps fail to provide any answer when exhausting memory. Our shared expertise with existing search techniques suggests that the impromptu evaluations that seem within the literature area unit inadequate. This sentiment is supported by our survey of existing analysis [3] and by others WHO area unit acquainted with the practices established by the IR community for the evaluation of retrieval systems (e.g., see Webber [5]). during this paper, we have a tendency to augment our previous work [3] with associate analysis of existing search techniques' runtime performance. Our findings indicate that a lot of area for improvement exists.

II. RELATED WORK

Existing evaluations of relational keyword search techniques are ad hoc with little standardization. Webber summarizes existing evaluations with regards to search effectiveness. Our previous work [3] compares relational

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

keyword search techniques with regard to search effectiveness but does not consider runtime performance. Baid et al. [6] assert that many existing keyword search techniques have unpredictable performance due to unacceptable response times or fail to produce results even after exhausting memory. Our results—particularly the large memory footprint of the systems—confirm this claim. A number of relational keyword search systems have been published beyond those included in our evaluation. Chen et al. [4] and Chaudhuri and Das [31] both presented tutorials on keyword search in databases. Yu et al. [32] provide an excellent overview of relational keyword search techniques.

III.EXISTING SYSTEM

The Internet users progressively demand for the keyword search interface to access knowledge effectively mistreatment the relative data. In existing system, extending the keyword search paradigm to relative information has been a lively space of analysis at intervals the information and data retrieval (IR) community. an outsized range of approaches are planned and enforced, however despite varied publications, there remains a severe lack of standardization for system evaluations. .

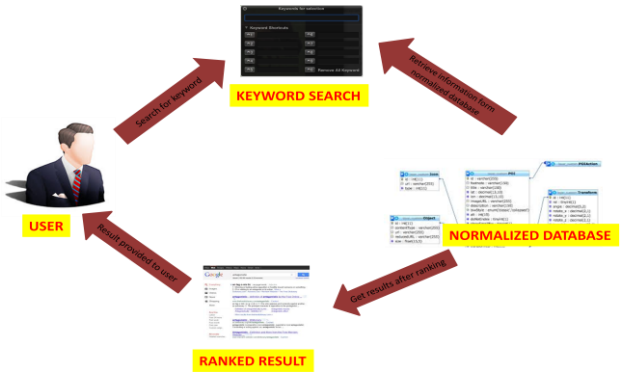
» Keyword Search without ranking.
» Execution time is more.

III. PROPOSED SYSTEM

the relational keyword search were gathered together with different data models. We apply normalization Technique for the database to remove the redundancy of data. We will maintain a Index for every File or Content. File is stored in a Separate space and it's Index is stored for the Related Data.

» Keyword Search with ranking.
» Execution Time consumption is less.
» File length and Execution time may be seen.
» Ranking can be seen by using chart.

V.ARCHITECTURE DIAGRAM



VI.MODULES

I. NETWORK CONSTRUCTION

Cloud servers are made with the files and also the index data are maintained within the main cloud server. the info are intercalary in every cloud servers, and network construction is formed with the complete information index gift in every cloud server. question is given to the most cloud server, in order that the most cloud server can verify the index data gift in it & divert the question to the corresponding cloud servers

II. FILTERING KEY WORDS

The words in the files are filtered and main keywords are filtered using Stemming Algorithm. The main keywords are extracted to filter the unwanted words. The Files names are updated in the corresponding cloud servers.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

III. ENCRYPTION MODULE

The query of the user is encrypted exploitation RSA algorithm; this cryptography method can forestall the information felony from the hackers. knowledge security is ensured exploitation RSA cryptography. In cryptography, RSA (which stands for Rivets, Shamir associated Adelman UN agency initial publically delineate it) is an algorithmic program for public-key cryptography. it's the primary algorithmic program identified to be appropriate for sign language in addition as cryptography, and was one in every of the primary nice advances publicly key cryptography. RSA is wide employed in electronic commerce protocols, and is believed to be secure given sufficiently long keys and therefore the use of up-to-date implementations.

IV. CLOUD SERVER

Cloud Server is the major main server which contains the index data of the entire data present in all sub Cloud Servers. The Cloud Server will act as the main server to receive the query from the user. The user query is encrypted using RSA algorithm, and sends to the main Cloud Server. The main Cloud Server decrypts the query and match with the index data present in it. The main Cloud Server will find the best match file using ranking algorithm.

V. RANKING ALGORITHM MODULE

In this module we rank the best file by calculating the ratio between term frequency with the total number of keywords. The value is calculated and compared with the rest of the values. The maximum valued files are ranked in order. The files are retrieved to the user as the index data of all the files are maintained in the index of the main cloud server.

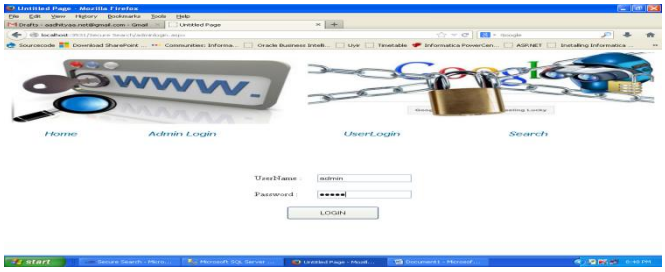
VI. BEST FILE IDENTIFICATION

The best file identification is achieved using Top k Query Algorithm. The maximum ranked values are obtained using Term frequency calculation. The files are kept in the ascending order. The best files are given as output to the main cloud server. The main cloud server retrieves top files and given as output to the user.

VII.OUTPUT

I. HOME PAGE

II. ADMIN LOGIN



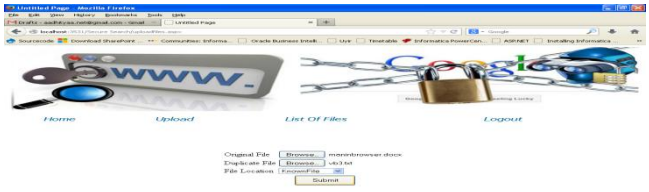
ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

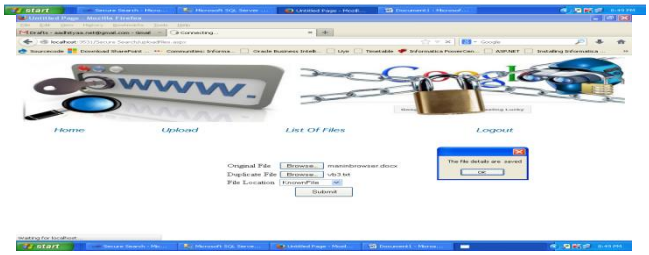
International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

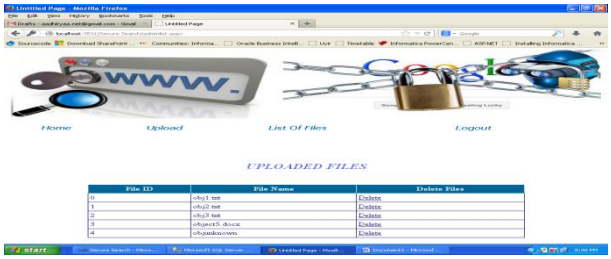
Vol. 3, Issue 3, March 2015

III.FILE UPLOADING

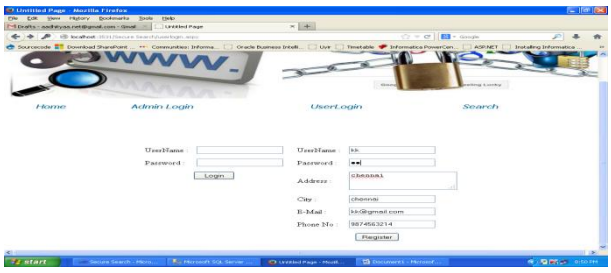




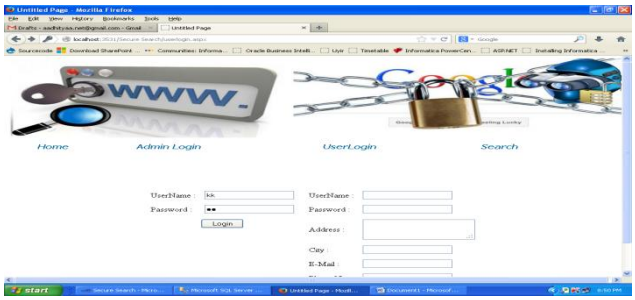
III. VIEW THE UPLOAD FILES



IV. USER REGISTRATION



V. USER LOGIN



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

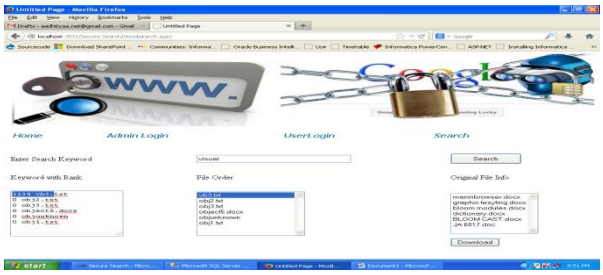
(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

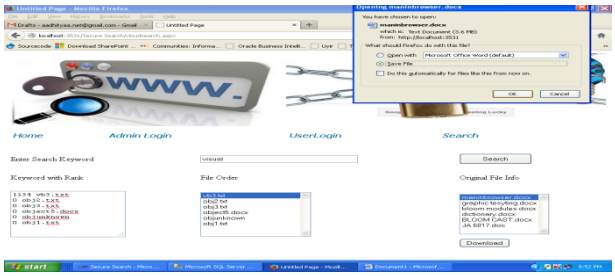
VI. USER PROFILE

VII. SEARCH

VIII. FILE ORDER LIST



IX. FILE DOWANLOAD

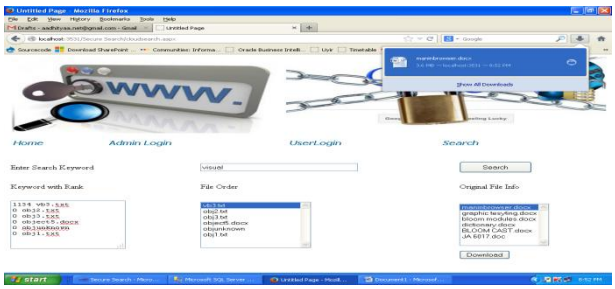


ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015



VIII.CONCLUSION

Unlike many evaluations reported in the literature, ours investigates the overall, end-to-end performance of relational keyword search techniques. Hence, we favour 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 data set). Our experimental results do not reflect well on existing relational keyword search techniques. Runtime performance is unacceptable for many search techniques. Memory consumption is additionally excessive for several search techniques. Our experimental results question the scalability and improvements claimed by previous evaluations. These conclusions are consistent with previous evaluations that demonstrate the poor runtime performance of existing search techniques as a prelude to a newly-proposed approach.

IX.FUTURE ENHANCEMENTS

Further analysis is definitely necessary to research the myriad of experimental style choices that have a major impact on the analysis of relative keyword search systems. for instance, our results indicate that existing systems would be unable to look the complete IMD b info, that underscores the requirement for a progression of information sets that may permit researchers to create progress toward this objective. making a set of the first information set is common, however we tend to aren't tuned in to any work that identifies the way to confirm if a set is representative of the first information set. additionally, totally {different| completely different} analysis teams typically have different schemas for a similar information (e.g., IMD b), however the impact of various info schemas on experimental results has additionally not been studied. Our results ought to function a challenge to the present community as a result of very little previous work has acknowledged these challenges. Moving forward, we tend to should address many problems. First, we tend to should style algorithms, information structures, and implementations that acknowledge that main memory is proscribed. Search techniques should manage their memory utilization with efficiency, swapping information to and from disk as necessary. Such implementations area unit unlikely to possess performance characteristics that area unit almost like existing approaches however should be used if relative keyword search systems area unit to scale to massive information sets (e.g., many various tuples). Second, evaluations ought to reprocess information sets and question workloads to produce bigger consistency of results, for even our results vary wide betting on that information set is taken into account. Having the community coalesce behind reusable take a look at collections would facilitate higher comparison among systems and improve their overall analysis [5]. as luck would have it, our analysis benchmark is starting to gain traction during this space as proved by others' adoption of it for his or her evaluations [34], [36], [39]. Third, the follow of researchers re implementing search techniques could account for a few analysis discrepancies. creating the first ASCII text file (or a binary distribution that accepts a info computer address and question as input) on the market to alternative researchers would greatly cut back the chance that ascertained variations area unit implementation artefacts.

REFERENCES

[1] D. Fallows, "Search Engine Use," technical report, Pew Internet and Am. Life Project, ttp://www.pewinternet.org/Reports/ 2008/Search-Engine-Use.aspx. Aug. 2008.
[2] com Score, "Global Search Market Grows 46 Percent in 2009," http://www.comscore.com/Press_Events/Press_Releases/2010/1/Global_Search_Market_Grows_46_%_in_2009, Jan. 2010.
[3] J. Coffman and A.C. Weaver, "A Framework for Evaluating Database Keyword Search Strategies," Proc. 19th ACM Int'l Conf. Information and Knowledge Management (CIKM '10), pp. 729-738, Oct. 2010.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

[4] Y. Chen, W. Wang, Z. Liu, and X. Lin, "Keyword Search on Structured and Semi-Structured Data," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '09), pp. 1005-1010, June 2009.
[5] W. Webber, "Evaluating the Effectiveness of Keyword Search," IEEE Data Eng. Bull., vol. 33, no. 1, pp. 54-59, Mar. 2010.
[6] A. Baid, I. Rae, J. Li, A. Doan, and J. Naught on, "Toward Scalable Keyword Search over Relational Data," Proc. VLDB Endowment, vol. 3, no. 1, pp. 140-149, 2010.

BIOGRAPHY

Avinandan Kumar & Dipankar Nag is a Student in computer science engineering Department , College of Bharath University ,He received Bachelor of Technology (B.TECH) degree in 2015 from Bharath university, Chennai Tamilnadu, India.