International Journal on Future Revolution in Computer Science & Communication Engineering
Volume: 5 Issue: 5
ISSN: 2454-4248
19 – 23

An Improved Algorithm for Faster Multi Keyword Search in Structured Organization

[1.] Pradeep Kumar Gupta
M.Tech. Scholar, Deptt. Of CSE
KITE,Jaipur
Pradeep0141@gmail.com

[2.] Satish Kumar Alaria
Asst. Professor, Deptt of CSE
KITE,Jaipur
Satish.alaria@gmail.com

ABSTRACT: Searching is the major concern in the database operation. For accessing the information from database it is always required to perform the search. If the search process is efficient then time taken the get the required information will also be less. In this scenario, contain the information organized in the structured data for the big organization like any automobile industry maintaining information regarding its department. Now if perform the keyword based query, then on based of the keyword the queries will be formed.
So, in order to reduce the time involved in the formation of the queries from the table have suggested the use of the associative mapping table in the search mechanism which will reduce the time involved in the process. The main aim this work is save the CPU time and efficient utilization of CPU to solve the purpose of the green computing.

*****

1. INTRODUCTION

Big data is set to offer association's gigantic learning. Regardless, with terabytes and peta bytes of data filling affiliations today, standard outlines and structures are not up to the test. IT bunches are stacked with consistently creating requesting for data, uncommonly delegated examinations and sporadic reports. Pioneers wind up confounded since it takes hours or days to discover answers for questions, if by any methods. More customers are reckoning that self-advantage access should data in a shape they can without quite a bit of an extend understand and offer with others. In the present hypercompetitive business condition, associations not simply need to find and separate the essential data they require, they ought to find it quickly. Discernment empowers relationship to perform examinations and settle on decisions essentially more rapidly, yet the test is encountering the sheer volumes of data and getting to the level of detail required, all at a quick. The test just creates as the level of granularity increases. One possible course of action is hardware. A couple of merchants are using expanded memory and compelling parallel getting ready to crunch colossal volumes of data to an extraordinary degree quickly. Another system is putting data in-memory yet using a structure figuring approach, where various machines are used to handle an issue. The two philosophies empower relationship to examine tremendous data volumes and get business bits of learning in close continuous. Notwithstanding whether you can find and dismember data quickly and put it in the most ideal setting for the gathering of spectators that will use the data, the estimation of data for fundamental initiative purposes will be jeopardized if the data isn't correct or helpful.

1.1 Big data management and analytics

Big Data management and analytics [1] inquire about is continuing in three ways. They are:

(i) Building foundation and superior figuring strategies for the capacity of big data

(ii) Data management procedures, for example, incorporating numerous data sources (both big and little) and ordering and questioning big data;

(iii) Data analytics procedures that control and investigate big data to remove chunks.

In synopsis, Big Data management and analytics systems incorporate expanding current data management and mining strategies to handle monstrous measures of data and additionally growing new methodologies including chart data management and digging procedures for keeping up and investigating extensive organized data.

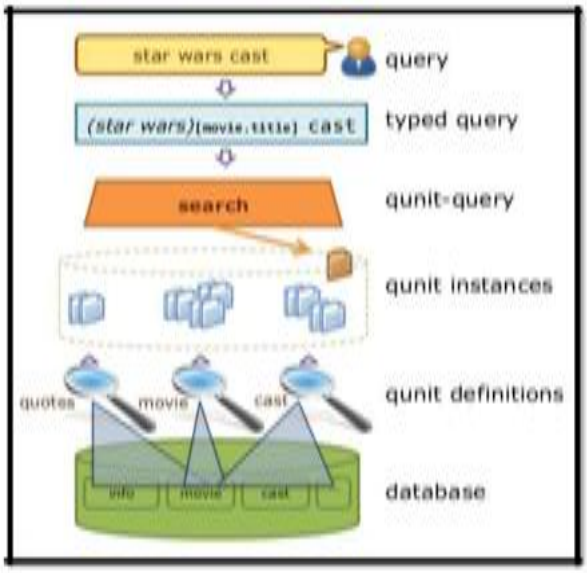

Fig 1.1 Searching Concepts in Database

2. LITERATURE SURVEY

Sonia Bergamaschi Elton Domnori Francesco Guerra [1], Keyword inquiries offer a helpful other option to customary SQL

 ISSN: 2454-4248
19 – 23

in querying relational databases with substantial, regularly obscure, patterns and occurrences. The test in anstheyring such questions is to find their proposed semantics, develop the SQL inquiries that depict them and utilized them to recover the separate tuples. Existing methodologies normally depend on files constructed from the earlier on the database content. This truly restrains their materialness if from the earlier access to the database content isn't conceivable. Illustrations incorporate the on-line databases got to through theyb interface, or the sources in data combination systems that work behind wrappers with particular query abilities. Besides, existing writing has not concentrated to its full expand the between conditions over the ways the distinctive keywords are mapped into the database esteems and outline components. In this work, they portray a novel strategy for making an interpretation of keyword inquiries into SQL in light of the Munkres (a.k.a. Hungarian) calculation. Our approach handles the over two constraints, as well as it offers huge changes in the distinguishing proof of the semantically significant SQL questions that portray the planned keyword query semantics. They give subtle elements of the procedure usage and a broad trial assessment.

K.Arun, Dr. L.Jabasheela"[2]. Worldwide Journal of Innovative Research in Information Security (IJIRIS).This paper the creator has given the data with respect to the big data and effect on the World Wide Web. This paper additionally portrays about the approach of Big Data, its Architecture and Characteristics. Here creator talked about the characterizations of Big Data to the business needs and how for it will help us in basic leadership in the business condition.

KalyaniShirudkar, DilipMotwani [3]. In this paper creator has exhibited challenges in the field of "Big data Security". Big data experiences number of difficulties which are identified with security like-calculation in conveyed programming, security of data stockpiling and exchange log, input separating from customer, versatile data mining and analytics, get to control and secure correspondence. Keeping in mind the end goal to handle the issues identified with the security the creator have talked about in the paper different composes security procedures for instance Type Based keyword search for security of big data, utilization of half breed cloud to give privacy in big data.

Nishu Arora, Rajesh Kumar Bawa[4] . The thought behind the Cloud processing is that client can utilize the administration whenever, anyplace through the Internet, straightforwardly using the program. And in the field of the cloud registering data is spared in virtual space as it utilizes the programs to utilize arrange administrations. And additionally the systems are related so principle concern is with respect to the security of data. In this paper the creator has portrayed about the big data , its writes and how it can be valuable with the assistance of cloud figuring.

A B M Moniruzzaman, Syed Akhter Hossain [5]. NoSQL, for "Not just SQL," which relates gathering of non-relational data management systems; where databases are not constructed mailnly on tables, and likewise don't utilize SQL for data control. NoSQL database management systems are likewise useful when working with a gigantic amount of data when the data's tendency does not require a relational model.

3. PROPOSED WORK

The time of big data is currently coming. Be that as it may, the customary data analytics will be unable to handle such huge amounts of data. The inquiry that emerges presently is, the way to build up an elite stage to proficiently break down big data and how to plan a proper mining algorithm to locate the helpful things from big data.

The main objective of the search is the explore the big data and to find out the best and the most effective algorithm to perform the search in the big data in the most effective way.

3.1 Proposed Concept

Big Data analysis for the structured data:

In this case we will implement the key word search algorithm for fetching the data from the structured data i.e.

Then it will search for all the keywords in all the columns of the table and then find out the frequency count i.e. the number of matches in the columns of each of the table.

And the columns having the bigger match will for the part of query.

Consider the following data structure,

Person

Name	Area	Phone	Address	Email
Watson	Database	(520) 4631234	30 Bloor	watson@aaa.bb
Lenzerini	Database	(390) 6987654	Ariosto 25	lenzerini@bbb.cc
Date	Database	(817) 1937842	107 GACB	date@ccc.dd
Hunt	Inf. Systems	(343) 2920812	17 Helix	Hunt@ddd.ee

Affiliated

Professor	Department
Watson	x123
Lenzerini	cs34
Date	cs34
Hunt	m111

Department

id	DName	Address	Director
x123	CS	25 Blicker	Watson
cs34	IE	15 Tribeca	Hunt
ee67	EE	5 Charles	Date
m111	ME	2 Cottle	Hunt

Author

Name	Publication
Lenzerini	Data Integration
Date	Foundation Matters

Publication

Title	Year	Resource
Data Integration	2002	ACM DL
Foundation Matters	2002	DBLP

Database

Name	Address
DBLP	http://www.informatik.uni-trier.de
ACM DL	http://portal.acm.org/dl.cfm

Fig 3.1 Sample Database Structure

	P.N	P.A	P.P	P.Ad	P.E	D.I	D.D	D.A	D.Di
workers	0	0	0	0	0	0	0	0	0
department	0	0	0	0	0	0	0	0	0
CS	30	18	0	18	0	50	75	50	50

Fig 3.2 Frequency Distribution of Matching

The table in **fig 3.2** specifies number of times the keyword is present in the table column. P.N means that Persons table Name column and workers row means that the number of times the worker keyword appears in the particular columns of the table.

Now we are proposing that we are we will place a simple mapping table in between the main database and search mechanism, with the following structure,

Now suppose we search for the Keyword John CS and the result of matching is that person table Name and Department Table Dname store that

 ISSN: 2454-4248
19 – 23

TABLE 3.1
KEYWORD MATCHING TABLE

FieldName	TableName	KeyWordmatch
Name	Person	John
Dname	Department	CS

So if we again search for the same keyword, then we have to not go through the whole frequency calculation process.
Thus we help us to reduce the time required for the search and building up the main query.
Using the set-associative mapping strategy for keyword searching in which we will maintain a table log for columns name and its matching keywords so that we can directly look into this table to form the query, this will reduce the time taken to form the query or retrieve the matched details.
In our proposed work we are taking the keyword search algorithm for searching for the big data and we are wanted to perform the secure search.

3.2 Keyword Search Algorithm in Big Data:
Keyword search was examined in the database setting before being reached out to data joining. Multi-keyword Boolean search enables the clients to enter multiple query keywords to ask for suitable documents. Among these works, conjunctive keyword search plots just restore the documents that contain the greater part of the query keywords. Disjunctive keyword search plans restore the majority of the documents that contain a subset of the query keywords.
Ranked search can empower brisk search of the most significant data. Sending back just the best k most significant documents can adequately diminish network activity.
In our proposed approach we have make use of two algorithms,

Algorithm for Keyword Search
Step1: Capture the Keyword String user entered for Searching
Step 2: Split the multi-keyword string into an array. Now each element of array is the keyword to be searched.
Step 3: In the keyword search, we will maintain the following data structures,

Structure 1:
tablename
fieldname
Keyword matched
By making this structure we will get access the table name and fieldname of the table containing the keyword.
Step 4: If the keyword search in not found then the statistical analysis of the counting the occurrence of the keyword in the columns of the table is performed and then the information is transferred to the table with structure 1.
Step 5: If the search keyword is found in the structure 1 , then the navigation in the tables is not required so the quick access to the related table query will be achieved and thus faster access of the data.

3.3 Security Concept in Big Data Search:
A trapdoor work is a capacity that is anything but difficult to figure one way, yet hard to register the other way (discovering its reverse) without uncommon data, called the "trapdoor". Trapdoor capacities are broadly utilized as a part of cryptography.
In numerical terms, if f is a trapdoor work, at that point there exists some mystery data y, with the end goal that given f(x) and y, it is anything but difficult to register x. Think about a latch and its key. It is minor to change the latch from open to shut without utilizing the key, by pushing the shackle into the bolt component. Opening the lock effortlessly, in any case, requires the way to be utilized. Here the key is the trapdoor.

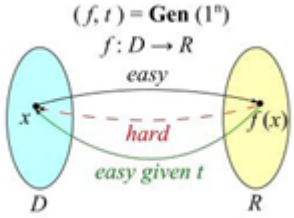

Fig: 3.4 a trapdoor function

The possibility of trapdoor work. A trapdoor work f with its trapdoor t can be produced by an algorithm Gen. f can be productively registered, i.e., in probabilistic polynomial time
In the accompanying two illustrations, we generally expect it is hard to factorize an extensive composite number.
RSA Assumption
In this illustration, having the opposite of e modulo φ(n), the Euler's totient capacity of n, is the trapdoor:

$$f(x) = x^e \mod n$$

On the off chance that the factorization is known, φ(n) can be figured, so then the reverse d of e can be registered d = e-1 mod φ(n), and then given y = f(x) we can discover x = yd mod n = xed mod n = x mod n. Its hardness takes after from RSA presumption.

4. RESULT ANALYSIS

Case I Single Word Search
The single word “Jaipur” is searched, then the Base works and proposed comparison is show in the table 4.4

TABLE 4.4
KEYWORD “CS” SEARCH COMPARED IN BASE AS WELL AS PROPOSED IMPLEMENTATION

Base Work	Proposed Work
2344 ms	28 ms

 ISSN: 2454-4248
19 – 23

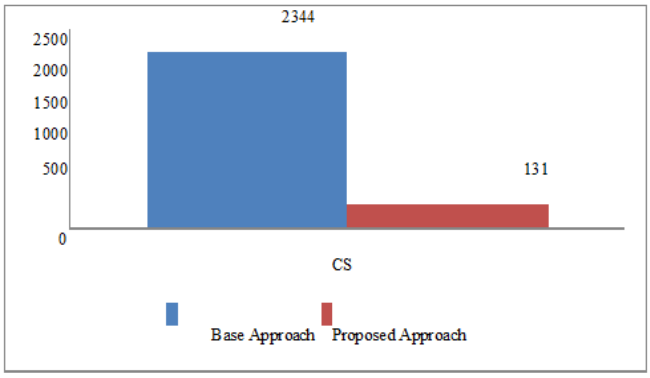

Fig 4.11 Comparison Graph for Keyword "CS"

Case II Multiple Word Search

The single word "CS Jaipur" is searched, then the Base works and proposed comparison is show in the table 4.5

TABLE 4.4
KEYWORD "CS JAIPUR" SEARCH COMPARED IN BASE AS WELL AS PROPOSED IMPLEMENTATION

Base Work	Proposed Work
2967 ms	231 ms

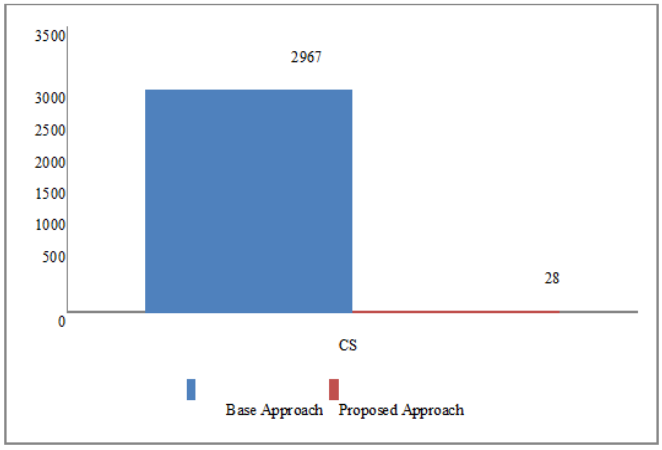

Fig 4.12 Comparison Graph for Keyword "CS Jaipur"

5. CONCLUSION & FUTURE WORK

The big data is a very vast field and much work is required to be done in this field and there is always requirement for the better algorithm for the searching of any information in the big data and in the future will try to extend this work on the cloud computing and parallel processing of the big data.

The main aim this work is save the CPU time and efficient utilization of CPU to solve the purpose of the green computing.

The algorithm of novel keyword search, which is proposed by us will make use of the mapping technique to store the more frequent search index, so the future search can be performed with the mapping table and will in return save the CPU time and in overall will reduce the load on the CPU, thus resulting in the conservation of the energy.

In future, will try to extend this research towards the search of the images and other complex data and in segment of security will try to extend our work with the DNA Passwords, ECG cryptography and similar concepts.

REFERENCES

[1]. Sonia Bergamaschi Elton Domnori Francesco Guerra.Keyword Search over Relational Databases: A Metadata Approach, 2011.

[2]. K.Arun, Dr. L.Jabasheela "Big Data: Review, Classification and Analysis Survey" International Journal of Innovative Research in Information Security" Volume 1, Issue 3, Sep 2014.

[3]. KalyaniShirudkar, DilipMotwani "Big-Data Security" International Journal of Advanced Research in Computer Science and Software Engineering Volume 5, Issue 3, March 2015.

[4]. Nishu Arora, Rajesh Kumar Bawa "A Review on Cloud to Handle and Process Big Data" International Journal of Innovations & Advancement in Computer Science IJ IACS ISSN 2347 – 8616 Volume 3, Issue 5 July 2014

[5]. A B M Moniruzzaman, Syed Akhter Hossain "NoSQL Database: New Era of Databases for Big data Analytics-Classification, Characteristics and Comparison" International Journal of Database Theory and Application Vol. 6, No. 4, August, 2013.

[6]. Shilpa, Manjit Kaur "BIG Data and Methodology-A review" Volume 3, Issue 10, October 2013 ISSN: 2277 128X.

[7]. Gary Pan, Seow Poh Sun, Calvin Chan and Lim Chu Yeong,"Analytics and Cybersecurity: The shape of things to come",CPA ,2015

[8]. Erol Gelenbe and Omer H. Abdelrahman,"Search in the Universe of Big Networks and Data",IEEE ,2014

[9]. Shengli Wu,Chunlan Huang,Jieyu Li,"Combining Retrieval Results for Balanced Effectiveness and Efficiency in the Big Data Search Environment",IEEE International Conference on Computer and Information Technology,2014.

[10]. Ajeet Lakhani,Ashish Gupta,K. Chandrasekaran,"IntelliSearch: A Search Engine based on Big Data Analytics integrated with Crowdsourcing and category-based search",International Conference on Circuit, Power and Computing Technologies ,2015

[11]. Zhihua Xia, Member, Xinhui Wang, Xingming Sun, and Qian Wang,"A Secure and Dynamic Multi-keyword Ranked Search Scheme over Encrypted Cloud Data",IEEE,2015

[12]. Bing Wang, Wei Song, Wenjing Lou Y. ,Thomas Hou,"Inverted Index Based Multi-Keyword Public-key Searchable Encryption with Strong Privacy Guarantee",IEEE Conference on Computer Communications (INFOCOM),2015.

[13]. Lu, Yue & Tan, Chew Lim.," Keyword searching in compressed document images". DEC,2003.

[14]. S. S. Pawar, A. Manepatil, A. Kadam and P. Jagtap, "Keyword search in information retrieval and relational database system: Two class view," International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), 2016.

[15]. Q. Dong, Z. Guan and Z. Chen, "Attribute-Based Keyword Search Efficiency Enhancement via an Online/Offline

 ISSN: 2454-4248
19 – 23

Approach," IEEE 21st International Conference on Parallel and Distributed Systems (ICPADS), 2015.

[16]. Kehinde K. Agbele, Kehinde Daniel Aruleba, Eniafe F. Ayetiran,"Efficient schema based keyword search in relational databases." University of Computer Studies, Mandalay, Myanmar, International Journal of Computer Science, Engineering and Information Technology (IJCSEIT) 2.6 (2012).

[17]. Sanjay Agrawal,SurajitChaudhuri,Gautam Das,"DBXplorer: enabling keyword search over relational databases",SIGMOD,2002.

[18]. A. Karapakula, M. Puramchand and G. M. Rafi, "Coordinate matching for effective capturing the similarity between query keywords and outsourced documents," IET Chennai 3rd International on Sustainable Energy and Intelligent Systems (SEISCON 2012), Tiruchengode, 2012.

[19]. W. Tang, L. Yan, Z. Yang and Q. H. Wu, "Improved document ranking in ontology-based document search engine using evidential reasoning," in IET Software, vol. 8, no. 1, pp. 33-41, February 2014.

[20]. ShengliWu,JieyuLi,"Merging Results from Overlapping Databases in Distributed Information Retrieval",PDP,2013.