AN APPROACH TOWARDS SECURE MULTIKEYWORD RETRIEVAL

[1] MERCY PAUL SELVAN , [2]DR. A. CHANDRASEKAR, [3]K. KOUSALYA,

[1]Research Scholar, Dept. of CSE, Faculty of Computing, Sathyabama University, Chennai.

[2]Prof., Department of Computer Science Engineering, St.Joseph's College of Engineering, Chennai.

[3]P.G. Student, Dept. of CSE,Faculty of Computing,, Sathyabama University, Chennai.

E-mail: [1]mercypaulselvan@gmail.com, [2]drchandrucse@gmail.com [3]kousi2012@yahoo.in

ABSTRACT

Extracting information from the database has become tedious process. Because redundant data will be extracted sometimes, so we are approaching a new method to extract the data from database. Giving single word as an input, we may not get relevant data, so multiple data can be given as input to extract the relevant information from the database, so that efficiency will be higher when compared to other methods. For security of data, we can encode the data on the admin side. Observations are done while queries were given to retrieve the data. The information was retrieved, joined and sorted using a top - k algorithm. The information are connected by a graph and status are reported.

Keywords: *Extract, data, graph, top-k, relevant,retrieval.*

1. INTRODUCTION

A Database contains information in the text format, so retrieving the data is complex process. If we give single word as an input, the word will be pattern matched in the index of the table and the same information may be retrieved repeatedly. If multiple data are given as input and for that particular word synonym word also indexed and that also retrieved.

A user was using the web to retrieve the data, they may not know the exact synonym for the word. If the user query more than one data as an input. We can get more relevant information when compared to single data as an input.

We try to retrieve the data in multiple catalogs. The multiple inputs are only to retrieve the needed data to the user.

Existing system implements keyword pattern matching in a federal database.using multi-level keyword seek,we propose   synonym of the particular data and these data are also retrieved so that the information we extract from multidatabase will be highly related information to the user.

2. LITERATURE SURVEY

Various methods have been proposed to extract the most relevant information Existing system can be split into two types:first type is extracting the data from the top of the database [3].In this method a keyword is given as an input and the data are retrieved and information is joined or linked shape the output of the requested user.In the second method joining will be done straightly on the data. Here each attribute is linked with a graph. Based on the time taken to retrieve the data graph will be drawn. [1] proposed a novel method for computing top-k routing plans based on their potentials to contain results for a given keyword query. [4] presented an adjustable framework to answer keyword queries using the authority transfer paradigm. [11] proposed a software framework to improve the quality of search engines and decrease time to search process.[12] proposed a utility and frequency dependent algorithm for path traversal using prefix tree structure.[13] proposed a data mining technique which helps in paatern matching using association rules.

In IR style ranking logical operators were used to retrieve the data.In this method information retrieved not highly relevant because it should satisfy either AND or OR condition. [5]

In our proposed system,we try to extract the relevant information and ensuring strong security to the data stored at the admin side.we don't need information from the top of the database,focusing on the bottom of the database.

There are two methods to select the data from the database,one is using matrix search and other one is graph based information search.

In matrix search it considers only binary values, there are many chances for repetition of data.while in graph based search each attribute is a node and each node is connected to a graph.based on the distance between the inputs given graph will be drawn for the time taken to retrieve the data and time taken to link the data.

Comparing with the matrix and graph based retrieving, graph based extracting the data is Tuff process and gives relevant information,reduces redundant information.

In our methodology we compute scores for time, distance to retrieve the data. Based on the score the information is linked and sorted using hash sort, information retrieved are highly relevant data.
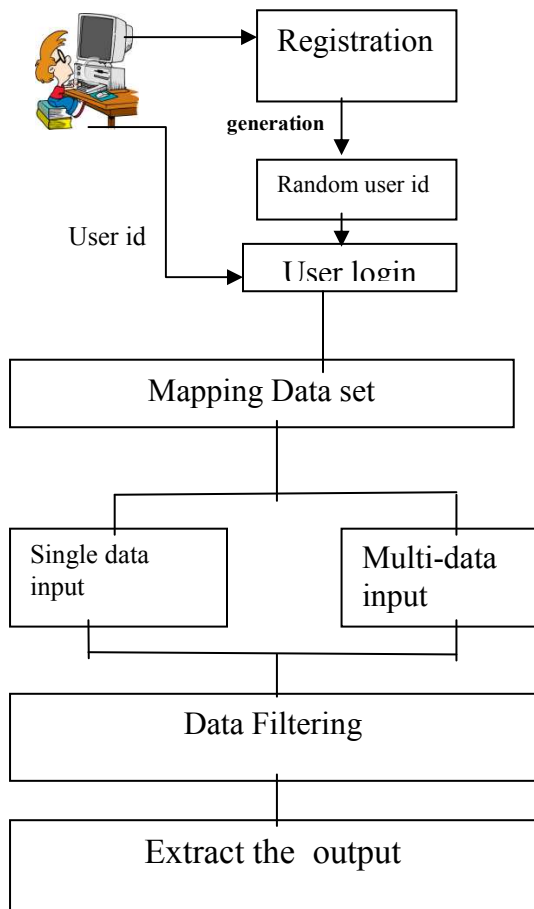
3 SYSTEM ARCHITECTURE:

Figure 1 System Architecture

In figure 1 the following process are done while extracting the output.

If the inputs are given data sets will be mapped in index of the table.And data are extracted from different tables using the attributes.

4.MODULE DESCRIPTION

Our system architecture will have the following modules.
4.1 User Registration
In this module user has to give their details such as their name, userid, password and some other needed information by the server side.A random user id will be sent to user mail-id and
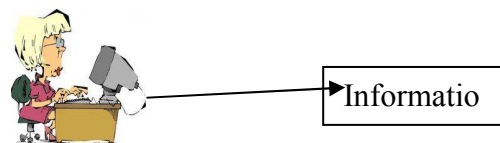the information will be submitted by the admin.

Figure 2 Registration

4.2 Encryption
A random five digit user id will be sent to user mail id.Here the admin has to approve the data.There are three process in the server side.New application, Approved application and cibil checked.In the new application the last submitted data will be in the admin.The admin has to approve the data.Here Encryption of data will be done to secure the data.To reduce redundancy the admin has to approve the data.Here admin will check the data two times..
If the data already present in the table admin reject the data else approved data will be stored as a row in the table.
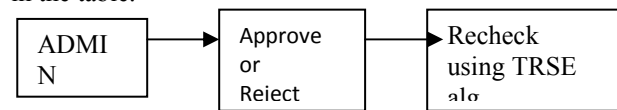

Figure 3 Audit By Admin

4.3.Single Data Search
The user has to give input to search the file in the database.By giving single data as an input, User may not get relevant information as an output.The single data input will be pattern matched and the files are extracted based on the data.
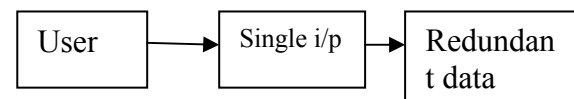

Figure 4 Single Data Search

4.4.Multi Data Search

In multi data search more than one word will be given as input and random userid or tuple id will also be given as input and the related words also indexed and highly relevantinformation are extracted by linking multiple file using join operation and the data are sorted using Top-k algorithm.Here each aatribute is taken as input and a graph was drawn, and status of the data were reported.
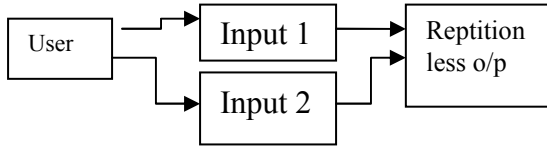

Figure 5 Multi Data Search

5 ALGORITHM

Given the data D and the synopsis Sd'.Algorithm calculates a set of combine method |CM|

Algorithm1:calculating set of values(D,Sd')

Input:the data D,the synopsis Sd'(Ad',ed')
Output:combined data sets
CP←a calculating method that consists of all data(di',dj')
Td←table which contains numerous row
Initially CP=NULL
While→CP.NULL() do
(di',dj')←CP.delete();
E(di,dj)←extract(E'(di',dj'))
If Td.null then
Td←E(di,dj)
Else
Td←E(di,dj) join Td
Calculate values for rows in Td to make out grouping of data
Calculate values of combine method in |CP| by means of (D,CP)
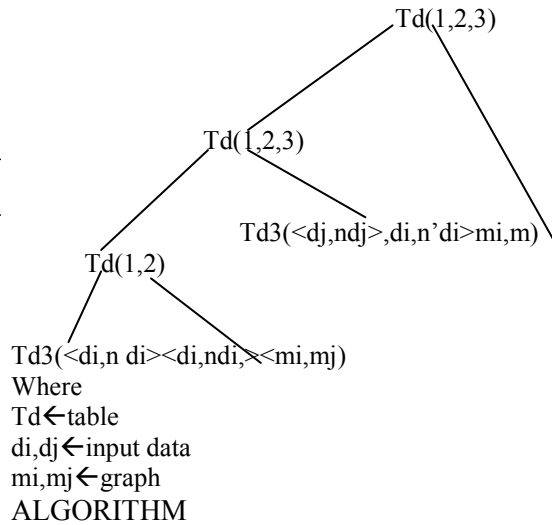order|CP| by value.
Where
Ad←attribute in a row
E←Edges
D←data
Example:
For the input data D={satyabama,mercy,lect} we will extract the multi factor relations {(satyabama,engineering,table1),(mercy,staff, table1)},{(satyabama,engineering,table1), (mercy,staff,table2)}and{(satyabama,engineering,table1),(mercy,name,table2)} for the couple of

data's(satyabama,mercy) and continue in the sameway(mercy,lecturer)and(satyabaa,lecturer).
By joining the above data our result will be

1. {(satyabama,engineering,table1),(mercy,stafftable3))join{(mercy,staff,table3),(lecturer, staff,table3)}
2. Join   the   extracted   data   with {(satyabama,engineering,table1),(lecturer,staff,table3)on the data extracted for satyabama and 3)
3. At   last   join {(satyabama,engineering,table1),(lecturer, staff,table3)with the 2 point.
4. The   result   for   the   input   data   is {satyabama,engineering),(mercy,staff,table) (staff,lecturer,table3)

Where
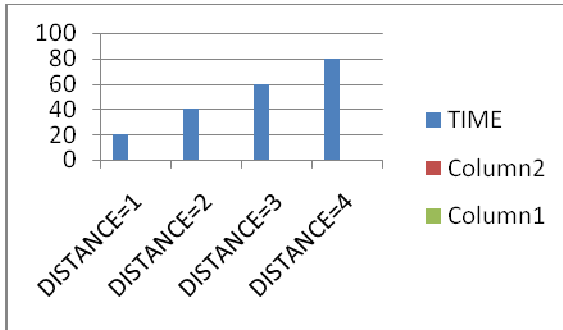Td←table
di,dj←input data
mi,mj←graph
ALGORITHM

Algorithm 2   To calculate top element in the database

1. Insert the top-k indexed item.
2. If the length of top-k is lesser than value k then
3. Insert the row in the ascending order.
4. Else for all top k value <zero.
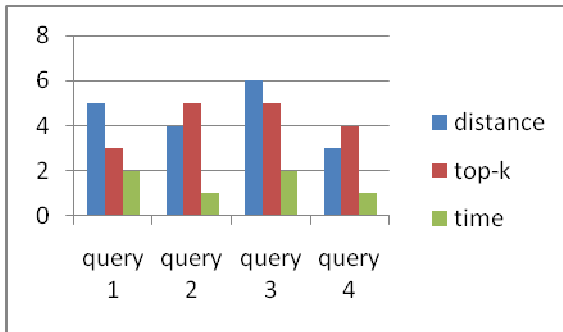5. Discard the top-k item and index in ascending order.

6 GRAPH

6.1 Graph Based On Time And Distance

Graph 1

If the distance between the nodes of the data rises, time taken to join the data and time taken to retrieve the data will be higher.
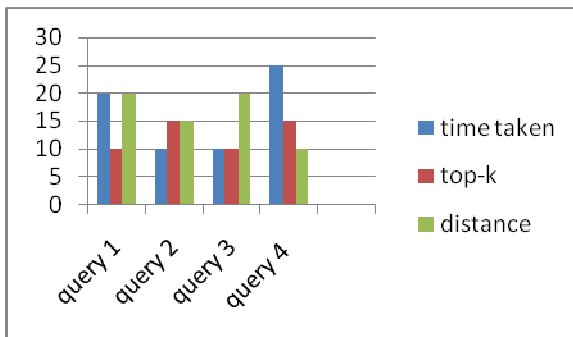
6.2 Graph On Single Data As An Input

Graph 2

If the input was a single data time taken to retrieve the is lesser when compared to multi data as an input.but there are many redundant information comparing with multi data input.

6.3 Graph For Multi Data Input

Graph 3

7. CONCLUSION

We provided a solution for efficient retrievel of data in relational database.extracted were highly releated data.Time taken while computing more then one data as input will higher but the output will efficient.Time will higher while computing distance is higher.While comparing with the existing system efficiency will be high.

REFRENCES:

[1] Than Tran and Leizhang "Keyword query routing", *IEEE Transactions on Knowledge and Data Engineering,* Vol. 26, No. 2, February 2014, pp. 363-374.

[2] L. Guo, F.Sha et al., "XRANK – Ranked Keyword Search Over XML Documents", SIGMOD 2003,June 9-12 2003,ACM 1-58113-634-X/03/06.

[3] F.Liu, W.Meng and A. Chowdhury, "Effective Keyword Search in Relational Databases", *SIGMOD 2006 June 27-29 2006,*ACM 1-59593-256-9/6/0006.

[4] A. Balmin et. al., **"**ObjectRank: Authority - Based Keyword Search in Databases ", *Proceedings of the 30th VLDB Conference,*Toronto, Canada, 2004.

[5] Mercy Paul Selvan, A.Chandrasekar "An Approach To The Pagerank Based on Dwelling Time and Frequency of Visiting Links", *Indian Journal of Applied Research(IJAR),* vol.4 no.10, October 2014, pp. 154,155.

[6] D.E. Rose and D.R. Cutting, "Ranking for Usability: Enhanced Retrieval for Short Queries," *Apple Technical Report* 163, 1996.

[7] T.K. Sellis, "Multiple-Query Optimization," *ACM Trans. Database Systems,* vol. 13, no. 1, pp. 23-52, 1988.

[8] MercyPaul Selvan and A.Chandrasekar "Ranking websites by its own features", *International Journal of Applied Engineering Research(IJAER),* ISSN 0973-4562 vol. 9 no. 22,2014pp. 12049-56

[9] Y.Luo "SPARK2: Top-k Keyword Query in Relational Databases", *IEEE Transactions on Knowledge and Data Engineering,* Vol. 23, No. 12, December 2011, pp. 1763-80.

[10]   A. Viji Amutha Mary and Dr. T. Jebarajan, "A Novel Optimized Hybrid Perturbation Technique to improve Security level", *International Journal of Applied Engineering Research (IJAER)*, ISSN 0973-4562 Volume 9, Number 21 (2014) pp. 10555-10566.

[11] R. Subhashini and V.Jawahar Senthil Kumar, "A Framework for Efficient Information Retrieval Using NLP Techniques", CCIS 142,Springer-Verlag Berlin Heidelberg, cnc 2011, pp. 391-393, 2011.

[12]   L.K. Joshila Grace and Dr. V. Maheshwari, "Utility, Importance and Frequency Dependent Algorithm for Web Path Traversal Using Prefix Tree Data Structure", *Journal of Theoretical and Applied Information Technology(JATIT),* ISSN 1992-8645 Volume 60, Number 1(2014) pp. 154-164.

[13] Kavitha Esther Rajakumari and   Dr. S. Srinivasan, "Software Development and Reuse using Functional clone", *International Journal of Applied Engineering Research(IJAER)*, ISSN 0973-456 Volume 10, Number 5(2015) pp. 13287-13295.