Relational Keyword Search System

Pradeep M. Ghige[#1], Ruhi R. Kabra[*2]

[#]*Student, Department Of Computer Engineering, University of Pune, GHRCEM, Ahmednagar,Maharashtra, India.*

[*]*Asst. Professor, Department Of Computer Engineering, University of Pune, GHRCEM, Ahmednagar, Maharashtra, India.*

Abstract-Today's world searching any information on search engine is most famous and required for people . These processes havenumber of disadvantages like there is no standard way forinformation retrieval. For these lack of standardization that willnot clearly show the actual result also it displays keyword searchwithout ranking and Execution time is more in existing system.In proposed system, we implement a relational keyword searchsystems which increase the performance of a system. It is alsoworking on the re-ranking of the data and keyword. Our systemcannot display the tentative result, it shows the exact result.e.g suppose user search "APPLE", Google give APPLE-fruit as a result, but the fact APPLE is also an mobile and laptop. So this is the maindrawback. In this project we overcome this is aissues, we definethe category of the search word first then after user will selectthe appropriate word meaning that is going to search. Thenafter we can do the several operation. We will statically addthe database in our project. The objective of this technique isto manage Information and database. Proposed systems involvedindependently and developed their own unique systems to allowusers to access information. We also explore the relationshipbetween execution time and factors. The techniques are used in this system including clustering, re-ranking and searching.

Keywords- *Query Form, Indexing, Searching, Dynamic Query, Information extraction.*

I. INTRODUCTION

With the growing of internet more and more people searchrequired information on internet. New trends of internet itis possible to store huge amount of information, for usingthis information several techniques are available [1]. Keywordsearch is one of them and it is very famous, easy for user.Keyword search use number of techniques and algorithm for
storing and retrieving data [2], Keyword search is possible ongraph structure which enables relational, HTML and XML data[5]. But it have some disadvantages like less accuracy, doesnot giving a correct answer, required more time for searchingand large amount of storage space. In existing techniquessome algorithms are used including DISCOVER, BANKS,BLINKS, EASE and spark [7]. One important thing is thatexisting techniques for information retrieval are not capable tohandle real world databases.We propose a system to overcomethe disadvantages of existing techniques which discussed inkeyword search. Data mining or information retrieval is theprocess to retrieve data from dataset and transform it to userin understandable form [15]. One important thing is that itdoes not required the proper knowledge of database queries.

1.1 Problem Definition

In current trends of information retrieval there is not anystandardize process, the keyword ranking is also an important factor which are unavailable in existing method. It also requirelots of clicks and search time. The main problem in existingsystem is that it does not giving the proper answer as perinserted keyword. Proposed system overcome these limitationsand gives user proper system to satisfy all these requirements.In propose system combine schema based and graph basedapproaches which help to store and retrieve all types of dataset.Two important factors are involved in this system effectivenessand efficiency which gives the proper result in minimum space.
Data warehouse hold massive quantities of information,but accessing these information is typically restricted due tocomplex query languages. The structure of relational databasesof keyword search in databases is to provide users with asimple interface to discover and retrieve information. Theobjective of the work described to transition existing searchtechniques from research curiosities into real world requirement.The objective of this techniques is to manage databaseand information retrieval systems involved independently anddeveloped their own unique system to allow users for accessingthe information. We also explore the relationship betweenexecution time factors.

2. RELATED WORKS

Relational Keyword search are change for different applicationsand retrieval systems are different for that purposes.Requirement of applications change as per its use and alsochange algorithm and techniques, also vary as per requirement.One technique is not fulfilling the requirement of other dataset.In this section we will discuss all the research and techniqueswhich are available in existing approaches.

ISSN: 2278 – 7798

A. **Schema Based Approaches**

Schema based approaches support keyword search overrelational databases using execution of SQL commands. Thesetechniques are combine vertices and edges including tuples andkeys. Some techniques are as follows-

1) SPARK: With increasing of the text information storedin relational databases, there is increase a demand for RDBMSto support keyword query search on the text data.These techniquespropose new ranking formula using existing informationretrieval techniques [12]. The important thing of this techniquesis to work on large scale real databases, the example iscustomer relationship management. It uses a top-k join algorithmwhich includes two efficient query processing algorithmsfor ranking function. 1. Dealing with non-monotonic scoringfunction and 2.skyline sweeping algorithm.

2) DISCOVER: It allows user to issue keyword querieswithout any knowledge of the databases schema [15]. It returnsqualified joining network of tuples which is set of tuples thatare associated because they join on primary and foreign keys,collectively contain all the keywords of the query. This algorithmworks in two steps-1. The candidate network generatorgenerates all candidate networks of relation. 2. Plan generatorsbuild plants for the efficient evaluation of the set candidatenetworks. DISCOVER use a greedy algorithm that produces anear optimal execution plan with respect to the actual cost. Thekeyword search in this techniques will proceed in three steps-1. First it generates the smallest set of candidate networks.2.Then greedy algorithm creates a non-optimal execution plan toevaluate the set of candidate networks. 3. Finally the executionplan is executed by the DBMS. The main drawback of thisalgorithm is it works only for static optimization.

B. Graph Based Approaches

Graph based approaches assume that the database is modeledas a weighted graph where the weight of the edges indicatethe importance of relationships. This weighted tree with edgesis related to steiner tree problem [19]. Graph base searchtechniques is more general than schema based techniquesincluding XML, relational databases and internet.

1) BLINKS: In query processing over graph-structured isa top-k keyword search query on a graph finds the top kanswered according to some ranking criteria. To reduce indexspace BLINKS partition a data graph into blocks. The bilevelindex stores the summary information at the blocklevel [10]. The main advantages of this algorithm is bettersearch strategy which combining indexing with search andpartitioning based indexing. It will conclude implementingranked keyword searches on graph structure data.

2) BANKS: BANKS is a system which enables that thedatabase is modeled as a weighted graph where the weightof the edges indicate importance of relationship [17]. Thisweighted tree with edges is related to steiner tree problem.Graph based search techniques is more general than schema based techniques including XML, relational databases andinternet.

C. Bidirectional expansion for keyword search on graph databases.

In relational databases XML and HTML data can bepresented as a graph with entities as node and relationshipas edges [13]. This techniques focuses on backward expandingsearch by allowing forward search from potential roots towardsleaves. It will handle partial specification of schema andstructure using tree pattern with approximate matching. Infuture work it can be reduce the number of nodes touched.

D. Queried unit for database search

Structure queries are hard to express to overcome this disadvantagesQUNITS focus on create a clear separation betweendatabase queries and ranking [4]. In QUNITS based searchhigh ranking segmentation is used which join between insertedquery words. This techniques structured information can beconsider as one source information rather than multiple. Thismake system easier to expand and enhanced with additionalIR methods.

E. EASE

The efficiency of keyword search on both structured andsemi-structured data is a challenging problem. EASE techniquesis model structured, unstructured and semi-structureddata as graph [9]. Existing search engine cannot integrate informationfrom multiple interrelated pages to answer keyword.EASE develop an efficient keyword search method basis on topk-style processing of large amounts of data for discovery ofrich structural relationship.

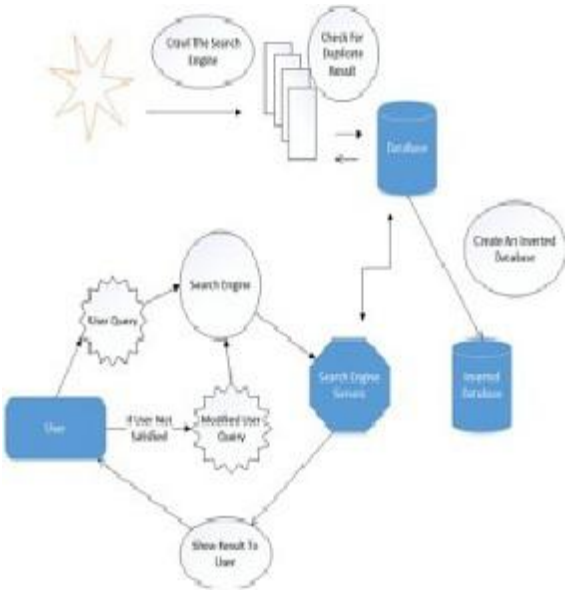
F. Steiner tree based search

A relational database can be modeled as graph G=(V,E). In this case there is one to one mapping between a tuple in the database and a node in V. G is directed graph with two edges[19]. Most existing method of keyword search over relationaldatabases find the steiner tree composed of relevant tuples asthe answer and identify a single tuple unit to answer keywordqueries. To overcome this problem this techniques integratemultiple related tuple units for effectively answer keywordqueries.

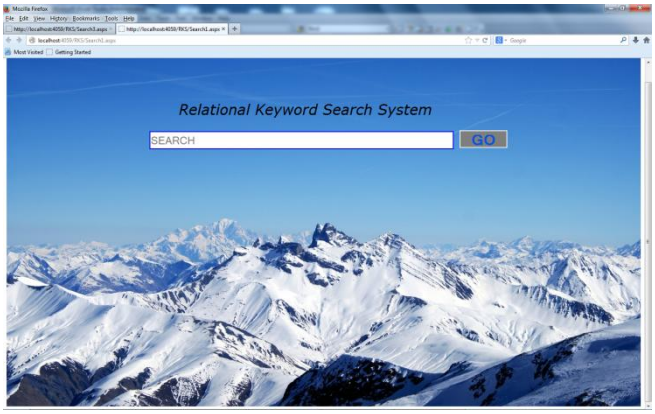
3. SYSTEM ARCHITECTURE

User insert query or keyword for retrieving an information then search engine fetch that query and give it to search engine server, search engine server searches an information in database also check the duplicate result. The crawler will process the data in database and give result to search engine. After all the processing information will give to user. If user will not satisfy then user insert modified query for searching an information. Same time inverted database will create which stored the current querying information. It helps toretrieve data efficiently with less time. The re-ranking process is simultaneously executed.

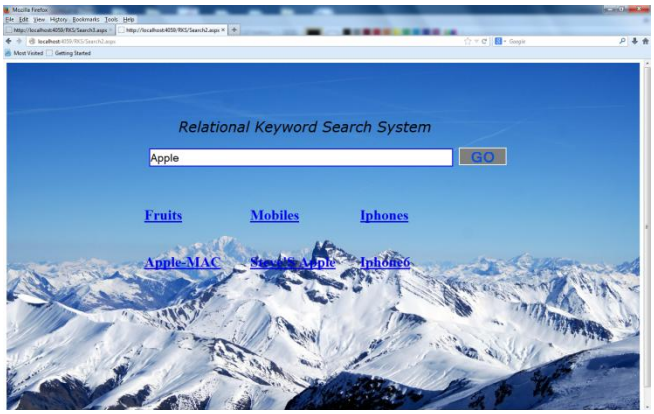
ISSN: 2278 – 7798



Inquisitive Module:
This module is worked at user side,it gives input content or query to system and process that query.The Information retrieval is depends on this module at userside. User simply insert query or keyword in search field andsee the result. The result is in the form of URL. User does notrequired the knowledge of database query, it will simply enterany keyword and get a result which is related to input. Thesynonyms of this model is searching model, which providesfor user simple and user-friendly searching option.



Clustering Module:
The main implementation of proposedsystem is based on clustering module which helpfor storing and retrieving information. According to enteredquery system will gives cluster information for selecting theappropriate cluster. After selecting the cluster it will searchinformation in that selected cluster and gives result to user. Atthe time of information storage selecting the matched cluster isimportant. The information is stored in exact matching clusterwhich will help to increase the efficiency and effectiveness ofsystem.



Admin Module:
Admin module is responsible for insertionof data as well as making clusters and manipulation.The database management is the major responsibility of adminmodule. All the type of data operations like insertion, updatingand deletion are responsible to admin module.

4. Mathematical Module

1: S be a system that is an search engine
S=f::::g
2:S = fI;O; P; S; F;Cg
3:Q—where Q is input query for search
4: O=O— O is output which is in form of URL or text
5: Identify function as P
P=searc(C)
P=(f1,f2,f3) is a number of function
f1=function of cluster
f2=function of searching
f3=function of re-ranking
6: C=constraint of which is used for dataset
7: Identify failure case a F
failure occurs when O!=user expectation
8: Identify success case as S
success is defined as O=user expectation

5. ANALYSIS AND RESULTS

A. Data sets
In order to confirm and implement proposedsystem work a rough data set is collected from World WideWeb that consist of one keyword related information which isa combination of data. If we are taking an example of applethen apple is a mobile, apple is a fruit and also apple is alaptop. When apple is a keyword then all information relatedto that keyword which collect from internet and stored in onesystem using clustering. The rough data is processed and storedin appropriate location i.e. in cluster is important task in ourproposed work. Each cluster store the information which havea similar features.

ISSN: 2278 – 7798

Performance: We attempt to utilize two basic algorithmfor performance measurement which are Precision and Recall.

Precision:

In the field of information retrieval, precision is the fraction of retrieved documents that are relevant to find.

$$precision \frac{relevant\ document\ \cap retrieved\ document}{retrieved\ document}$$

Precision takes all retrieved documents into account, but it can also be evaluated at a given cut-off rank, considering only the topmost results returned by the system. This measure is called precision at n or P@n. Precision is also used with Recall the percent of all relevant documents that is returned by the search.

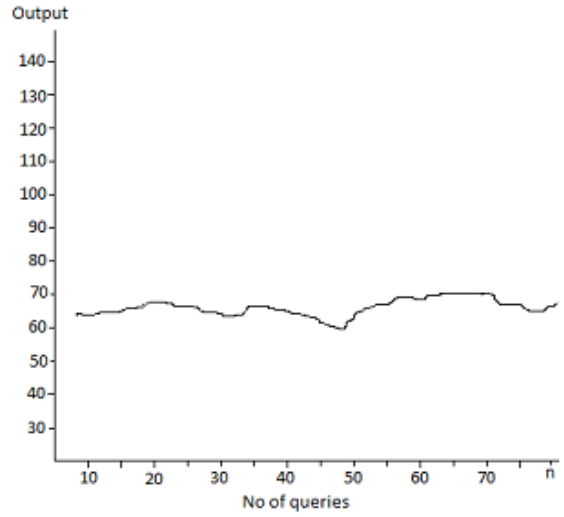
Recall:

Recall in information retrieval is the fraction of the documentsthat are relevant to the query that are successfully retrieved.

$$Recall \frac{relevant\ document\ \cap retrieved\ document}{relevant\ document}$$

PRECISION AND RECALL PERFORMANCE

Performance	@10	@20	@30	@40
Precision	0.675	0.656	0.649	0.643
Recall	0.679	0.688	0.721	0.696



Overall we will study all the existing techniques which isavailable in market. Each system has some advantages andsome issues. We compare all the techniques and checked the

performance. So finally conclude that any existing systemcannot fulfill all the requirement of keyword query search.They require more space and time; also some techniques are limited for particular dataset. The Proposed technique issatisfying number of requirement of keyword query search using different algorithms. The performance of keyword searchis also the better to compare other and it shows the actual resultrather than tentative. It also shows the ranking of keyword andnot requires the knowledge of database queries. Compare toexisting algorithm it is a fast process.To making this proposed system more powerful and accuratesome enhancements are required including the first buildsincrementally on the research presented in this work. Theseextensions address more recent research that re-examines assumptionsthat underlay the evaluation of relational keywordsearch techniques. The second area for future work is unquestionablythe most critical. It addresses how to visualize andbrowse through relational keyword search results; visualizationwill likely be a central component of these systems when theyare deployed in real-world settings. As a future work we cansearch the techniques which are useful for all the datasets,means only single technique can be used for MONDIAL,IMDb etc., also implement a PMSE(Personal Mobile SearchEngine) for mobile search and this system apply for multipledatabase.

REFERENCES

1. Joel Coffman, Alfred C. WeaverAn Empirical Performance Evaluationfor Relational Keyword Search Systems, IEEE transaction on Knowledgeand Data Engineering, 2014.

2.Sharmili C.,RexieJ.A.M.,"Efficient Keyword Search Methods in RelationalDatabases",IJERA,2013 International Journal of EngineeringResearch and General Science.

3. William Webber,"Evaluating the Effectiveness of Keyword Search",IEEEComputer Society Technical Committee on Data Engineering,2010.

4.Arnab Nandi,H.V.Jagdish,"Qunits:queried units for database search" ,4th Biennial Conference on Innovative data system research(CIDR) Asilomar,California,USA,2009

5.Yi Chen,WeiWang,"Keyword Search On Structured and Semi-Structured Data" ,SIGMOD'09 Rhode Island,USA,2009.

6. Li Qin, Jeffrey Xu.Yu.,Keyword search in Databases:The Power ofRDBMS",SIGMOD'09 Rhode Island,USA,2009.

7. Konstantin Golenberg, Benny Kimelfeld,"Keyword proximity Search inComplex Data Graphs", SIGMOD'08,Vancouver,BC,Canada,2008.

8. BhavanDalvi, MeghaKshirsagar,"Keyword Search on External MemoryData Graph",VLDB'08, Auckland, New Zealand,2008.

9. Guoling Li, Beng Chin Ooi,"EASE: An Effective 3-in-1KeywordSearch Method For Unstructured, Structured and Semi- StructuredData,SIGMOD'08,Vancouver, BC,Canada,2008.

10. Hao He, HaixunWang,"BLINKS: Ranked Keyword Searches OnGraph",SIGMOD'07,Beijing,China,2007.

11. Fang Liu, Clemet Yu,"Effective Keyword Search in RelationalDatabases,SIGMOD,Chicago,USA,2006.

12. YiLuo, XueminLin,"SPAR

13. K:Top-k Keyword Query in Relational Databases",SIGMOD'07,Chicago,China,2007

ISSN: 2278 – 7798

14.VarunKocholia, ShashankPandit,"Bidirectional Expansion For KeywordSearch on Graph Databases",University of California,USA,2005
15. M.L.Shore, L.R.Foulds,"An Algorithm for the Steiner Problem InGraph",(National Chung-Cheng University,China,2004
16. VegelisHristidis, YannisPapakonstantinou, "DISCOVER : KeywordSearch in Relational Database",28thVLDB Conference, HongKong,China,2002
17. AmitSinghal, "Modern Information Retrieval: a brief overview",IEEEcomputer Society Technical Committee on Data Engineering,2001
18.GauravBhalotia,ArvindHulgeri,"Keyword Searching and Browsing inDatabase using BANKS",University of California, Berkeley.
19.VagelisHristidis, Luis Gravano,"Efficient IR-Style Keyword SearchOver Relational Databases".
20. E.W.Dijkstra,"Two Problems in Connexion with Graph"1959 .

ISSN: 2278 – 7798