Everant.org/ETJ
Research Article
Engineering and Technology Journal ISSN: 2456-3358

Keyword Query Routing

Ziba Naaz

M.Tech Scholar Adusumilli Vijaya Institute of technological and Research Center
Ramanthapuram Hyderabad

ARTICLE INFO

corresponding Author:
Ziba Naaz
M.Tech Scholar Adusumilli Vijaya Institute of technological and Research Center
Ramanthapuram Hyderabad

ABSTRACT

Keyword search is an intuitive paradigm for searching linked data sources on the web. We propose to route keywords only to relevant sources to reduce the high cost of processing keyword search queries over all sources. We propose a novel method for computing top-k routing plans based on their potentials to contain results for a given keyword query. We employ a keyword-element relationship summary that compactly represents relationships between keywords and the data elements mentioning them. A multilevel scoring mechanism is proposed for computing the relevance of routing plans based on scores at the level of keywords, data elements, element sets, and subgraphs that connect these elements. Experiments carried out using 150 publicly available sources on the web showed that valid plans (precision@1 of 0.92) that are highly relevant (mean reciprocal rank of 0.89) can be computed in 1 second on average on a single PC. Further, we show routing greatly helps to improve the performance of keyword search, without compromising its result quality

I. INTRODUCTION

The Web today is not only a collection of textual data but also a collection of interlinked data sources (e.g., Linked Data). Linking Open Data is one such large project through which large amount of legacy data is transformed into the Resource Description Framework (RDF) and linked to other sources and published as linked data [1]. Linked data is comprised of many sources that contain billions of Resource Description Framework triples which are linked by millions of links like 'same As' links, which are published more frequently.

It would be difficult for a typical web-user to explore this linked data on the Web using any structured query languages. This is where the keyword search is applied. Unlike structured query languages, here, it is not necessary for the user to have any knowledge of the schema of the underlying data that he need to exploit. In the present scenario when a query is passed to the database through a keyword, it searches for the most relevant structured results [1], [2], [3] or a single relevant database. The issue with this approach is the Web of Linked Data is not directly applicable as a source may encompass may Linked sources of data. The main problem with this approach is not about finding the most relevant source, but computing most relevant combination of sources[6],[7].

II. LITERATURE SURVEY

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy n company strength. Once these things are satisfied, then next steps are to determine which

Published Vol. 2 Issue 1 Jan.2017
DOI: 10.18535/etj/v2i1.01
Page no.- 127-133

operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration are taken into account for developing the proposed system.

Effective Keyword Search in Relational Database

With the amount of available text data in relational databases growing rapidly, the need for ordinary users to search such information is dramatically increasing. Even though the major RDBMSs have provided full-text search capabilities, they still require users to have knowledge of the database schemas and use a structured query language to search information. This search model is complicated for most ordinary users. Inspired by the big success of information retrieval (IR) style keyword search on the web, keyword search in relational databases has recently emerged as a new research topic. The differences between text databases and relational databases result in three new challenges:(1) Answers needed by users are not limited to individual tuples, but results assembled from joining tuples from multiple tables are used to form answers in the form of tuple trees. (2) A single score for each answer (i.e. a tuple tree) is needed to estimate its relevance to a given query. These scores are used to rank the most relevant answers as high as possible. (3) Relational databases have much richer structures than text databases. Existing IR strategies are inadequate in ranking relational outputs. This paper, propose a novel IR ranking strategy for effective keyword search. The first that conducts comprehensive experiments on search effectiveness using a real world database and a set of keyword queries collected by a major search companies. This strategy is significantly better than existing strategies. This approach can be used both at the application level and be incorporated into a RDBMS to support keyword-based search in relational databases.

III. PROPOSED WORK

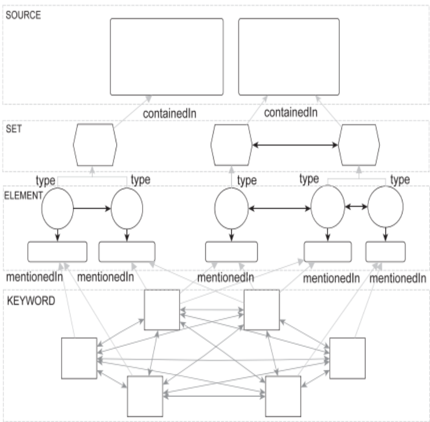
We propose a new method to solve the problem of keyword search over a large number of linked and structured data sources using keyword query routing. The high-cost of searching for keywords that span across different sources can be reduced by routing the keywords only to relevant sources. Unlike the existing system which only uses the relationships between the keywords, we employ the keyword element relationship graph [9] and apply routing plans over the obtained results. Then we apply Maximum Likelihood algorithm on the obtained results to the number of results by filtering the unwanted results we obtained from the keyword element relationship graph. It propose a novel method for computing top-k routing plans based on their potentials to contain results for a given keyword query. It employs a keyword-element relationship summary that compactly represents relationships between keywords and the data elements mentioning them. A multilevel scoring mechanism is proposed for computing the relevance of routing plans based on scores at the level of keywords, data elements, element sets, and sub graphs that connect these elements. Based on modeling the search space as a multilevel inter-relationship graph, it also proposes a summary model that groups keyword and element relationships at the level of sets, and developed a multilevel ranking scheme to incorporate relevance at different dimensions.

Advantages:

It reduce the high cost of processing keyword search queries over all sources. It improves the performance of keyword search.

Published Vol. 2 Issue 1 Jan.2017
DOI: 10.18535/etj/v2i1.01
Page no.- 127-133

SYSTEM ARCHITECTURE:



Modules:

The system is proposed to have the following modules along with functional requirements.

1. Keyword Search
2. Keyword Query Routing
3. Multilevel Inter-Relationship
4. Set - Level

1. Keyword Search

There are schema-based approaches implemented on top of off-the-shelf databases .A keyword query is processed by mapping keywords to elements of the database (called keyword elements).

Then, using the schema, valid join sequences are derived, which are then employed to join ("connect") the computed keyword elements to form so-called candidate networks representing possible results to the keyword query. Schema-agnostic approaches operate directly on the data. Structured results are computed by exploring the underlying data graph. The goal is to find structures in the data called Steiner trees (Steiner graphs in general), which connect keyword elements.

2. Keyword Query Routing

We propose to investigate the problem of keyword query routing for keyword search over large number of structured and Linked Data sources. Routing keywords only to relevant sources can reduce the high cost of searching for structured results that span multiple sources. To the best of our knowledge, the work presented in this paper represents the first attempt to address this problem. A solution to keyword query routing can address these problems by pruning unpromising sources and enabling users to select combinations that more likely contain relevant results. For the routing problem, we do not need to compute results capturing specific elements at the data level, but can focus on the more coarse-grained level of sources.

3. Multilevel Inter-Relationship

The search space of keyword query routing using a multilevel inter-relationship graph. The inter-relationships between elements at different levels are above *Fig*. A keyword is mentioned in some entity descriptions at the element level. Entities at the element level are associated with a set-level element via type. A set-level element is contained in a source. There is an edge between two keywords if two elements at the element level mentioning these keywords are connected via a path. We propose a ranking scheme that deals with relevance at many levels.

4. Set Level

We extract keywords and relationships from the data. Then, based on the elements and sets of elements in which they occur, we create keyword-element relationships. Precomputing relationships (i.e., paths) between data elements are typically performed for keyword search to improve online performance. These relationships are stored in specialized indexes and retrieved at the time of keyword query processing to accelerate the search for Steiner graphs. For database selection, relationships between keywords are also precomputed. This work neither considers relationships between keywords nor relationships between data elements but between keyword-elements that collectively represent the keywords and the data elements in which they occur.

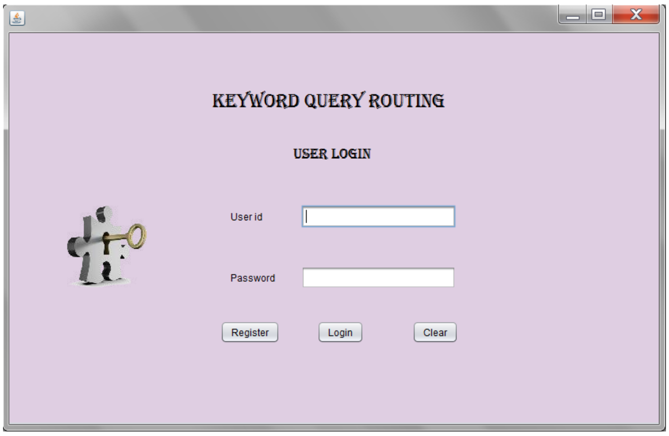
Published Vol. 2 Issue 1 Jan.2017
DOI: 10.18535/etj/v2i1.01
Page no.- 127-133

IV. PROJECT EXECUTION AND RESULT ANALYSIS

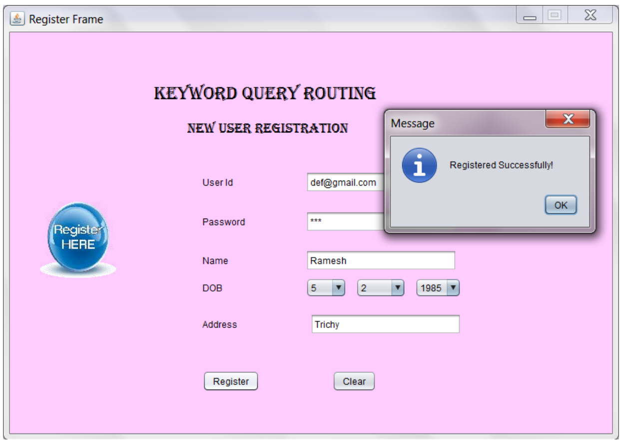
This chapter covers the snapshots that show the results of the project. The output of the project to various inputs is given in this chapter. The snapshots are self-explanatory. The result varies from input to input. The snapshot makes the user understand easily the working operations in the project. Below are snapshots of our project.

This chapter covers the snapshots that show the results of the project. The output of the project to various inputs is given in this chapter. The snapshots are self-explanatory. The result varies from input to input. The snapshot makes the user understand easily the working operations in the project. Below are snapshots of our project.

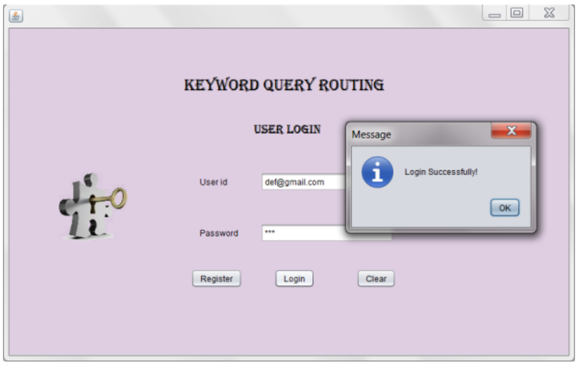
SNAPSHOTS OF THE PROJECT



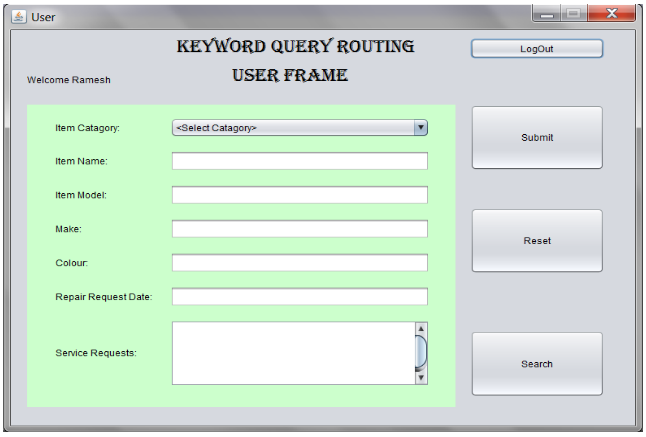
There is user login page in which we need to login by entering the user id and password



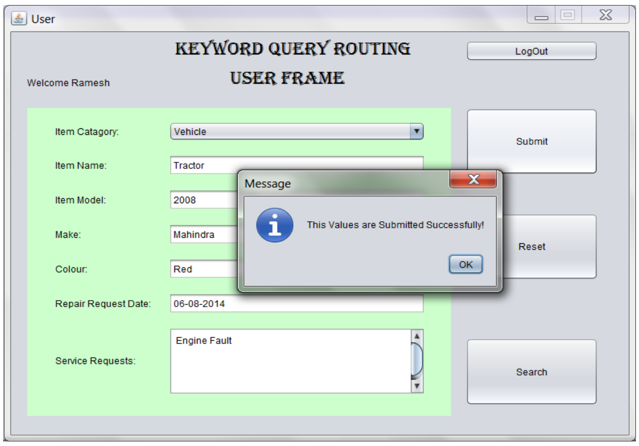
After login user name and id is registered by clicking on to register button. And it shows registered successfully



This shows that the user id n password has been successful.

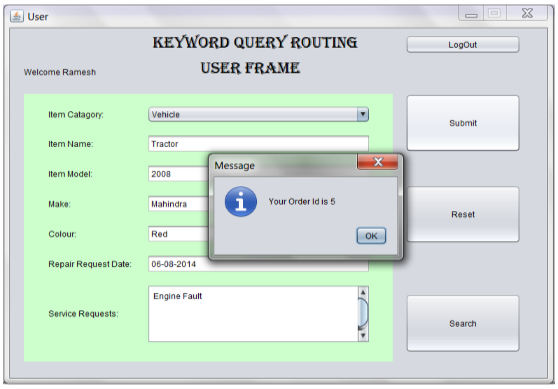


It shows user frame in which we submit the details of item that the user needs

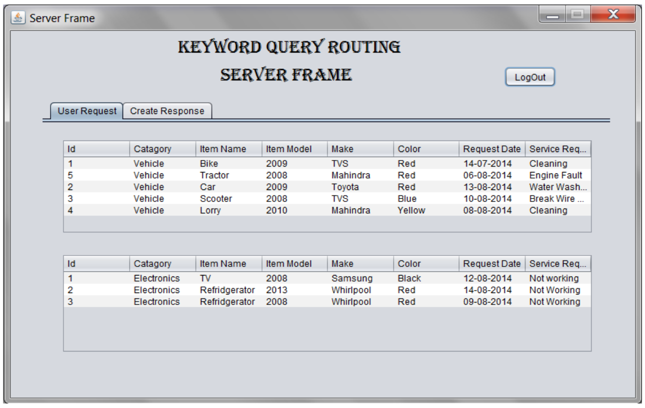


We can order many other items also by clicking on to reset button. Again enter the details what u want to order and click on to submit .It shows submission successful

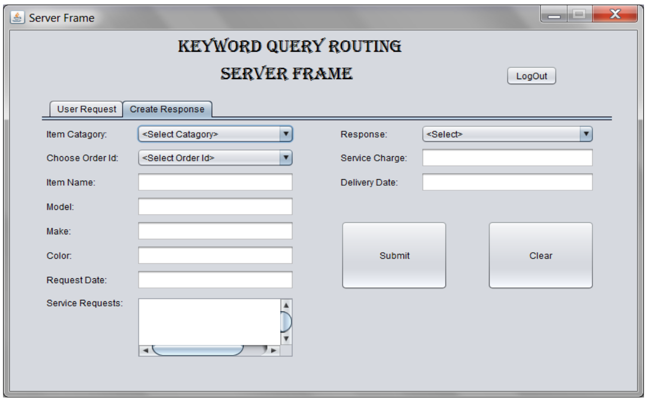
Published Vol. 2 Issue 1 Jan.2017
DOI: 10.18535/etj/v2i1.01
Page no.- 127-133



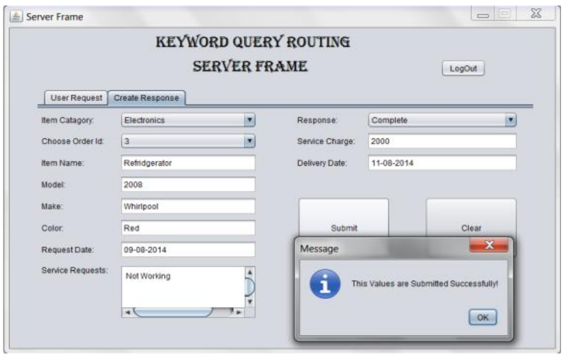
It shows the order id



This show the server frame in which details of items that have been registered

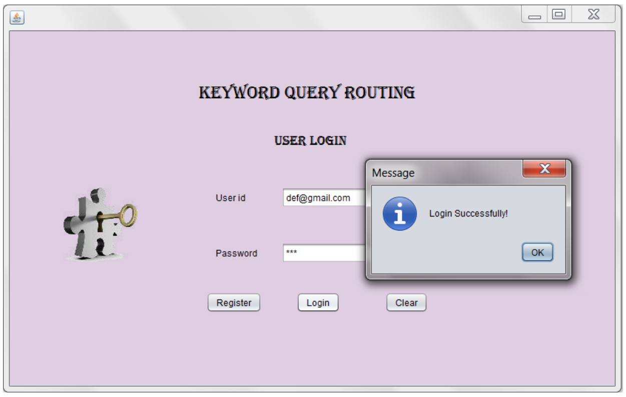


Enter the details for the server frame

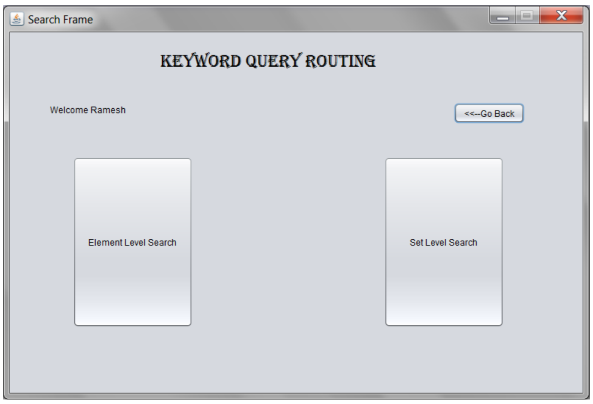


After submitting it displays a message that the value is submitted successfully

Clicking on to the clear button user can submit other values n submit.It displays a message value submitted successfully. After submission logout from the frame

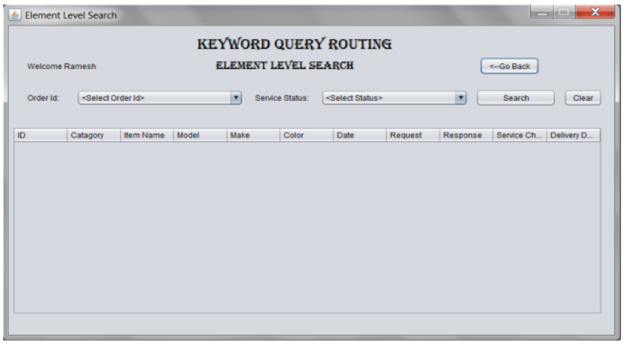


Enter the user name and password and clock on to login

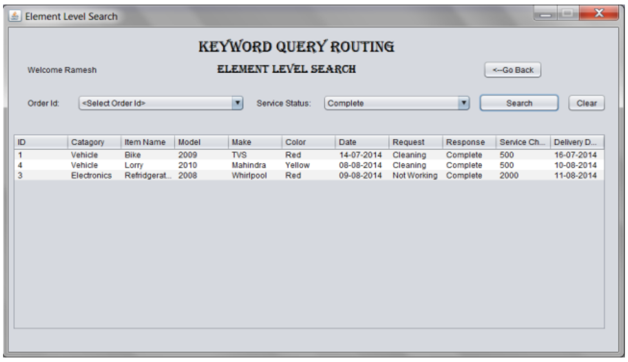


We can search any item click on to element level search

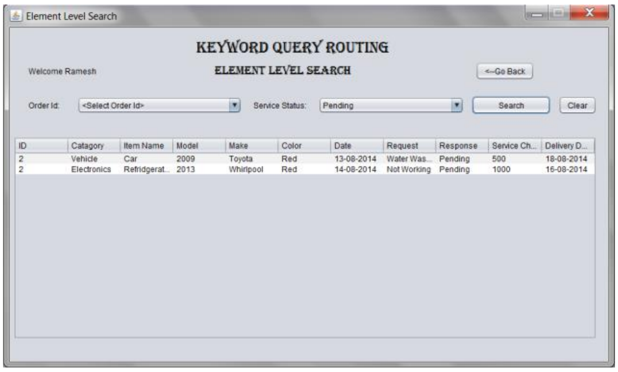
Published Vol. 2 Issue 1 Jan.2017
DOI: 10.18535/etj/v2i1.01
Page no.- 127-133



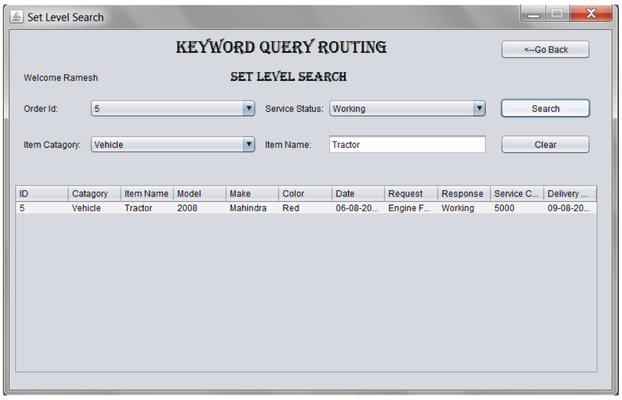
Select the oder id need to search



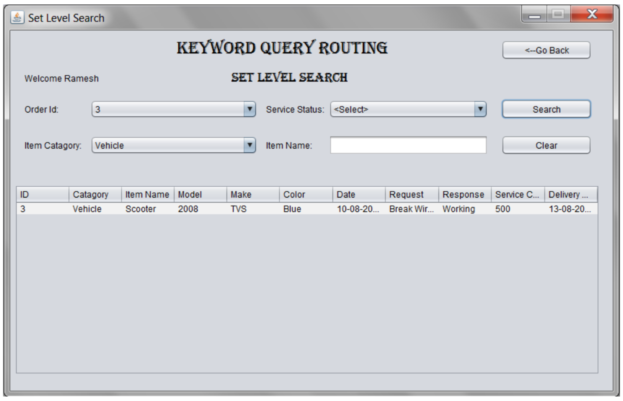
This show the element level of search which is complete



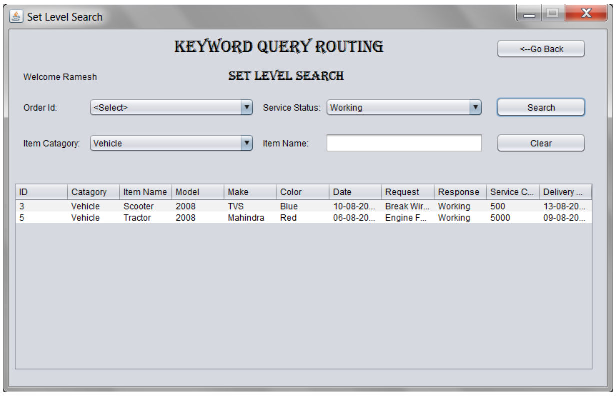
Shows the service level which is pending



Shows the order id with its service status as working



This shows the oder id with no service status



This shows the service status which is working.

V. CONCLUSION AND FUTURE WORK

This project proposes the idea of routing keyword query to produce more relevant results by implementing relationship graphs between the keywords at different levels. This idea proposes to reduce the high cost of searching for structured data spanning across multiple resources by routing the keywords only to the relevant sources. A correct routing plan will be selected by using graphs developed based on the relationships between keywords in the query at different level. This project is tested with a database having 856 records in four different datasets. The records in the datasets are created such that each dataset will some kind of data about the keywords. By this we created a web of data similar to the Linked data on the internet where information about a keyword may encompasses on different sources. Queries with more keywords would also generate effective results, but they cannot be handled efficiently. For example, if we give a query with more keywords as

Published Vol. 2 Issue 1 Jan.2017
DOI: 10.18535/etj/v2i1.01
Page no.- 127-133

a query in the existing system, it would also give effective results, but it might take a higher time which is not desirable in a present day's demand of high responsiveness. Keyword search without routing is problematic when the query has many words. That is the reason for routing of queries having more number of keywords. We plan to explore more. This query rewriting framework gives the better results compare to all previous frameworks environments. Its retrieves the high quality results compare to all previous approaches. All users are following these approaches and improve the recall for retrieve the results information. In future we control more dirty queries with new keyword query routing approaches.

BIBLIOGRAPHY AND REFERENCES

1. V. Hristidis, L. Gravano, and Y. Papakonstantinou, "Efficient IR-Style Keyword Search over Relational Databases, "Proc. 29th Int'l Conf. Very Large Data Bases (VLDB), pp. 850-861, 2003.
2. F. Liu, C.T. Yu, W. Meng, and A. Chowdhury, "Effective Keyword Search in Relational Databases," Proc. ACM SIGMOD Conf., pp. 563-574, 2006.
3. Y. Luo, X. Lin, W. Wang, and X. Zhou, "Spark:Top-K Keyword Query in Relational Databases," Proc. ACM SIGMOD Conf., pp. 115-126, 2007.
4. Pawar Prajakta Bhagwat et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (1), 2015, 434-437
5. Oya Y. Rieger, " Search engine use behavior of students and faculty: User perceptions and implications for future research"
6. B. Yu, G. Li, K.R. Sollins, and A.K.H. Tung, "Effective Keyword-Based Selection of Relational Databases," Proc. ACM SIGMOD Conf., pp. 139-150, 2007.
7. Q.H. Vu, B.C. Ooi, D. Papadias, and A.K.H. Tung, "A Graph Method for Keyword-Based Selection of the Top-K Databases," Proc. ACM SIGMOD Conf., pp. 915-926, 2008.
8. http://firstmonday.org/article/view/2716/2385
9. Thanh Tran And Lei Zhang, "Keyword Query Routing Ieee Transactions On Knowledge And Data Engineering, Vol. 26, No. 2, February 2014. 46
10. K.Collins-Thompson, Reducing the risk of query expansion via robust constrained optimization. In CIKM. ACM, 2009.

Published Vol. 2 Issue 1 Jan.2017
DOI: 10.18535/etj/v2i1.01
Page no.- 127-133