International Journal of Advance Research in Science and Engineering
Vol. No.5, Special Issue No. (02), March 2016
www.ijarse.com
IJARSE
ISSN 2319 - 8354

ROUTING QUERY USING KEYWORD SEARCH

Ashwini Wakchaure[1], Kirti Pawar[2], Snigdharani Pradhan[3]
Prof. Devdatta Bagul[4]
[1,2,3,4]Department of Computer Engineering, BVCOE & RI, Maharashtra, (India)

ABSTRACT

The number of accessible text data is growing rapidly in database, the need for hunting such information material is fiercely increasing. Even though the major RDBMSs have provided full-text search capabilities, but still there is a need of the whole knowledge of the data base to be known to the user and use a structured query language to hunt the information. This search paradigmatic is sophisticated for most traditional users. Keyword search is anperceptive model for searching associated data sources on the web. We recommend to route keywords only to consistent sources to decrease the high cost of processing key search queries over all sources.A novel method called top-k routing plans is used for computing, based on their potentials to contain results for a given key questioning. We introduce a key-element interrelationship summary that completely represents the interrelationships between keys and the data elements. A multilevel scoring structure is introduce for enumerating the relevancy of routing plans based on the level of keys, data bits, element sets, and sub graphs that connect these elements.

Keywords: Keyword search, Routing plan, Keyword query, RDF, Graph structured data, etc…

I. INTRODUCTION

Query processing over graph-structured data has appeal to much attention recently, as applications from a variety of areas continue to good large volumes of graph-structured data[7]. Keyword search is a proven and widely popular mechanism for querying document systems and the World Wide Web. newly, it has even been extensively applied to extract useful and important information from the Internet. Furthermore, the database research community has also recognized the benefits of keyword search and has been introducing keyword search capability into relational databases. [7]In this paper, keyword query routing plays important role. To investigate the problem of keyword query routing for searching keyword over a number of multiple linked data sources and structured data. Routing keywords only to relevant sources can decrease the high cost of searching for structured results that span multiple sources. [1][2] The existing system uses the keyword relationship (KR) collected individually for single databases. [1] [2] [3] Existing system shows the relationship between keywords as well as data elements. They are constructed for collection of linked data sources and then finally grouped as element of a compact summary called the set level keyword element relationship graph [KERG]. [3] This summary is important for address the scalability requirement of the linked data web scenario. [1][2] IR-style ranking has been proposed to incorporate relevance at the level of keywords. [1][2] To increased key ambiguity in the web-setting, a multiple level scoring mechanism is use for computing relevance of routing plans based on scores of the routing plans based on scores at the level of keywords data elements, element sets and sub graphs that link these elements. [1][2]The web is a source for searching information. The web is a collection at databases either it is text databases or relational databases. Also web is a collection at interlinked data sources that is linked data. Linked data is a connect related data using the web. Linked data compromise hundreds of sources containing billions of data which are connected by millions of links. [1][2] In database research the solution have been proposed either by given keyword query and retrieve the most relevant structured result or simply select most relevant databases [1][2][3]. But this solution is used only for single source. They are not directly applicable to the web at linked data. [1][2] Linked data produce the results including multiple data sources. The important thing here is to compute the most relevant combinations of sources from the database. The linked data on the web as shown in fig.1

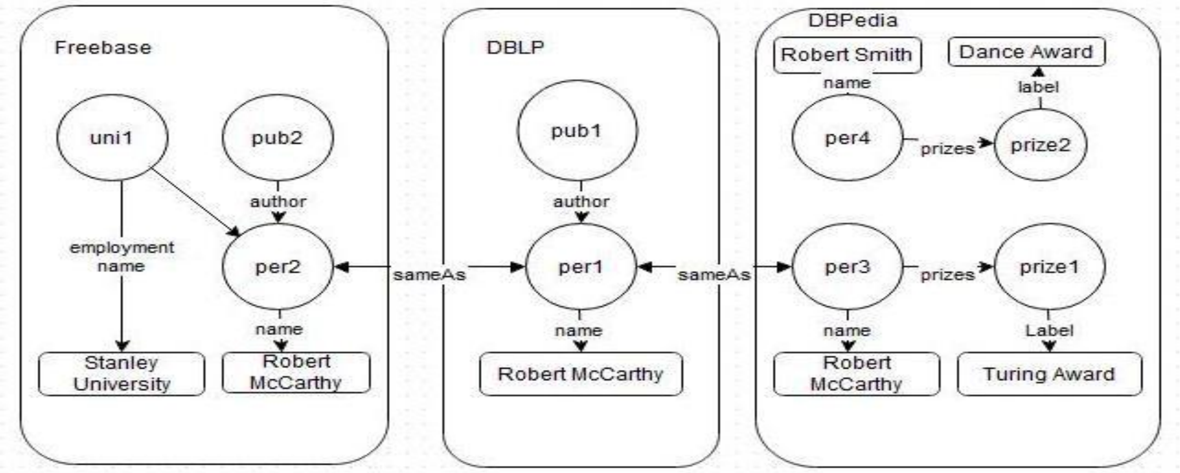

Fig1.Example of Linked data on web [1]

The most important queries are retrieved based on the keyword query; i.e., selects the single most important databases. The main issue here is to compute the most important combinations of sources from the database. The goal is to produce routing plans, which can be used to count results from multiple sources. We are focusing to the problem of keyword query routing over a huge number of data sources. Routing keywords only to relevant sources can reduce the high cost of searching for structured results that extent multiple sources. Relationships are represented between keywords and/or data elements. They are builds for the entire collection of linked sources, and then grouped as elements called the set-level keyword-element relationship graph (KERG)[11]. There are generic Linked Data browsers which allow users to start browsing in one data source and then cross along links into associated data sources. There are Linked Data search engines that crawl the Web of Data by among links between data sources and provide expressive query capabilities over aggregated data, similar to how a local database is problem today. The Web of Data also opens up new possibilities for domain-specific applications. Unlike Web 2.0 mashups which work across a fixed set of data sources, Linked Data applications operate on top of an free, global data space. This enables them to deliver more complete answers as new data sources appear on the Web[12]. We propose to investigate the problem of keyword query routing for keyword search over a large number of structured and Linked Data sources. Routing keywords only to important 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 show the first attempt to address this problem we use a graph-based data model to characterize individual data sources. In that model, we distinguish between an element-level data graph representing relationships between individual data elements, and a set-level data graph, which captures information about group of elements. This set-level graph essentially captures a part of the Linked Data schema on the web that is represented in RDFS, i.e., relations between classes.

II. LITERATURE SURVEY

Fang Liu, Clement Yu et al., projecteda unique IR ranking Strategy for effective keyword search. the primary that compact comprehensive experiments on search effectiveness employing aworldinfo and a group of keyword queries collected by a significant search firms. This strategy is considerablyhigher than existing ways. [8]. Guoliang Li. et al., projectedANeconomical and accommodative keyword search methodology, known as EASE, for classification and querying giant collections of heterogeneous knowledge. to realize high potency in process keyword queries, we tend to1st model unstructured, semi-structured and structured knowledge as graphs, and so summarize the graphs and construct graph indices rather thanexploitationancient inverted indices.[7]. V.Hristidis et al., adapts IR-style document-relevance ranking methods to the matter of process free-form keyword queries over RDBMSs. this question model will handle queries with each AND and OR linguistics, and exploits the delicate single-column text-search practicalityusuallyon the market in business RDBMSs. Yi Nilotic Xuemin sculpturedynasty Wang et al., studies the effectiveness and also thepotencyproblems withrespondent top-k keyword question in electronic database systems. It planneda brand new ranking formula by adapting existing IR techniques supported a natural notion of virtual document. Compared with previous approaches, this new ranking methodologyis straightforwardhowever effective, and agrees with human perceptions. It studied economicalquestionprocessways for the new ranking methodology, and propose algorithms that have stripped-down accesses to the information [6]. Quang Hieu Vu et al., proposes GKS, a completely uniquetechniquefor choosing the top-K candidates supported their potential to contain results for a given question. GKS summarizes everyinfo by a keyword relationship graph, wherever nodes represent terms and edges describe relationships between them. Keyword relationship graphs area unitused for computing the similarity between everyinfo and aAmerican statequestion, so that, throughoutquestionprocess, solelythe foremost promising databases area unit searched. [5]

Mayssam Sayyadian et al., describes Kite, an answer to the keyword-search drawback over heterogeneous relative databases. Kite combines schema matching and structure discovery techniques to search out approximate foreign-key joins across heterogeneous databases. Such joins are crucial for manufacturingquestion results that span multiple databases and relations. Kite then exploits the joins – discovered mechanically across the knowledge bases – to alterquick and effective querying over the distributed data.[4] Bei Yu et al., study the informationchoicedownside for relativeknowledge sources, and propose a technique that effectively summarizes the relationships between keywords in a verycomputer databasesupported its structure. It develop effective ranking strategiessupported the keyword relationship summaries so asto pickthe foremosthelpful databases for a given keyword question. It enforcedthis technique on Planet workplace. thereinatmospherewe tend to use intensive experiments with real datasets to demonstrate the effectiveness of this projectedaccountmethodology.[3]

Table 1: Comparative Study [1]

Sr.no.	Paper Name	Technique/Existing Work
1	Effective Keyword Search In Relational Database	A novel IR Ranking strategy for effective keyword search
2	EASE: An Effective 3-in-1 Keyword Search Method for Unstructured, Semi-structured and Structured Data	An adaptive method, EASE, for indexing & querying large collection of heterogeneous data
3	Efficient IR-Style keyword search over Relational Database	It focuses on the Top-k matches for the query as well as adapts IR-style document relevance ranking strategies
4	SPARK: Top-k Keyword Query in Relational Databases	A new ranking formula adapted for existing IR techniques based on natural notion of virtual document
5	A Graph Method for Keyword based Selection of the top-K Databases	A novel GKS method is used for selecting the Top-k candidate based on their potential to contain result for a given query
6	Efficient Keyword Search Across Heterogeneous Relational Databases	It describes kite, a solution to the keyword search over heterogeneous RDBMS
7	Effective Keyword-based Selection of Relational Databases	It develop effective ranking methods based on keyword relationship

III. EXISTING SYSTEM

The existing system work on keyword search that relies on an upper level model named as element-level model i.e. data graphs. These information graphs are accustomed calculate keyword question results. parts mentioning keywords are retrieved from element-level model & methods between these parts are explored to calculate graphs. [1][2][11] To subsume the keyword routing drawback, the weather is hold on at the side of the sources. Hence, this data is retrieved & then derive the routing plans from the computed keyword question results. Thus, this answer is generally apply once range of keyword parts are little & thus exploring information graphs once range of keyword parts are massive, then it's pricey.

KRG (keyword relationship graph) captures the relationship between keywords at the keyword level. As opposite to keyword solutions, KRG are not captured direct edges between tuples but stand for paths between keywords.[1][2] & in database selection, KRG association are retrieved for all pairs of query keywords to construct a sub graph.[1][2] The main purpose here is to check whether, not only keywords but also tuples mentioning they are connected. KRG mainly focuses on database selection, it only needs for know that whether the two keywords are connected by using some join sequences. In the KRG, such type of information is stored as a relationships & this information can be retrieved directly. Such retrieved information can be paths between data elements. Retrieving & exploring paths that might be composed of several edges are clearly more expensive than retrieving relationships between keywords.[11] Multisource KRG models both relationships that are within & those that link between sources.[1][2] Keyword relationships are stored together with the elements are associated with source information. An element-level key-element relationship graph (E-KERG). [11]

International Journal of Advance Research in Science and Engineering
Vol. No.5, Special Issue No. (02), March 2016
www.ijarse.com
IJARSE
ISSN 2319 - 8354

3.1 Disadvantages of Existing System

The number of potential results may increase exponentially with the number of sources and links between them. Yet, most of the results may be not necessary especially when they are not relevant to the user.

IV.PROPOSED SYSTEM

The main purpose for keyword query is to identify the result from data sources. The result may produce the data from multiple data sources. All keyword search approach is used by users only for compact results, but issues here is to find the top-k keyword routing plans based on their relevance to a query. A appropriate plan should correspond to the information need as in intended by the user.

The search area of keyword question routing employing a construction inter-relationship graph. At all-time low level, it models relationship between keywords. The Figure two shows the inter-relationship between components at varied levels. At the part a keyword is mentioned in some entity descriptions. Entities at the part level area unit related to a set-level part supported the kind.[1][2] A set-level part is contained in a very supply.

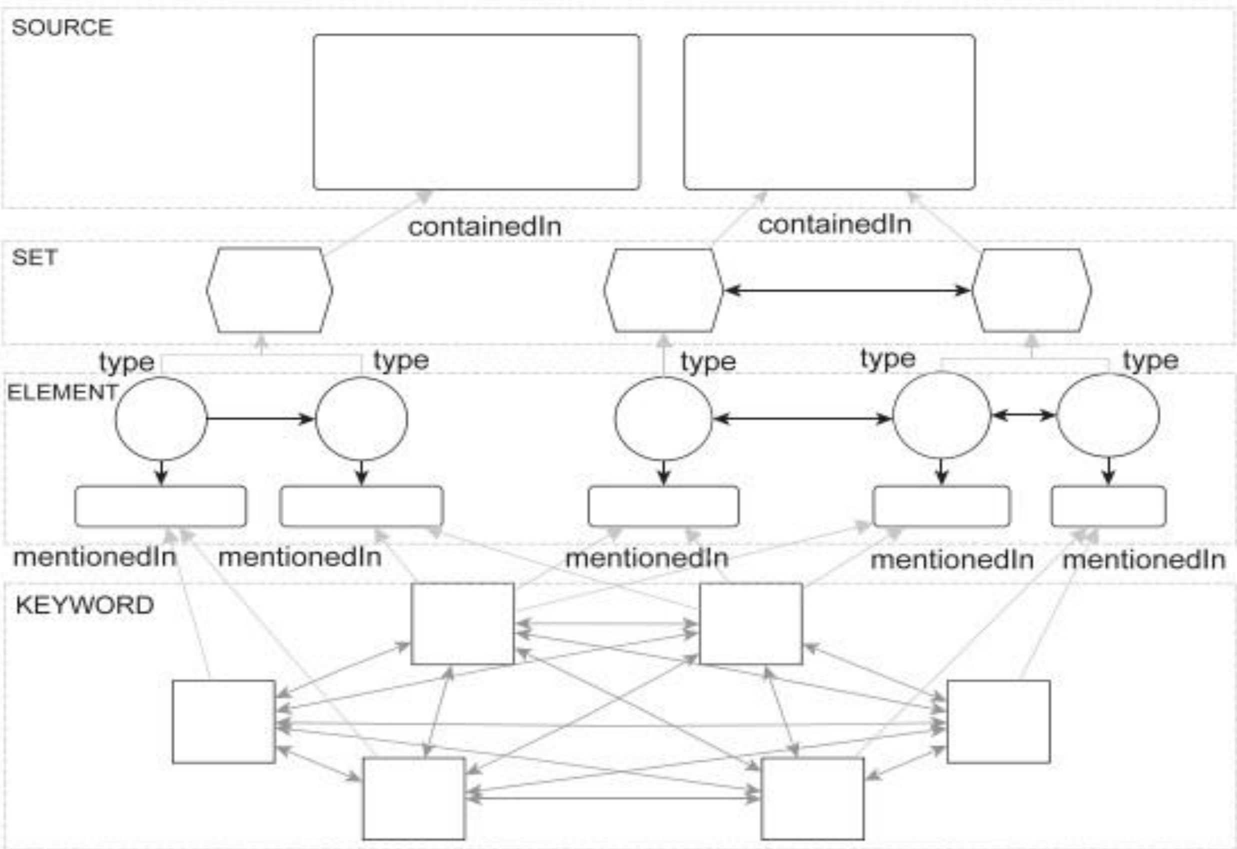

Fig 2: Proposed system

Advantages Of Proposed System:

1) Routing keywords only to relevant sources can reduce the high cost of searching for structured results that

span multiple sources.

2) The routing plans, produced can be used to compute results from multiple sources.

Algorithm**:**

- **Input**: The query K, the summary W'k (N'k, k)
- **Output**: set of routing plans [RP]

JP ← a join plan that contains all (ki , kj) 2k ;

T ← a table where every tuple captures a join sequence of KERG relationships e'k 'k , the score of each e'k , and the combined score of the join sequence; it is initially empty;

While JP .empty() **do**

(ki , kj) ← JP .pop() ;

(ki , kj) ← retrieve(k , (ki , kj));

If T .empty() **then**

T ← (ki , kj);

Compute scores of tuples in T via

Score (K , W'k s);

[RP] ← Group T by sources to identify unique combinations of sources;

Compute scores of routing plans in [RP] via

SCORE (K , RP);

Sort [RP] by score;

V. SYSTEM OVERVIEW

It is for searching linked information sources on the computer network. To route keywords only to the relevant sources to reduce high cost of processing keyword search queries over all sources. Aim is to improve the performance of keyword search, without compromising its result quality. In the proposed system, query expansion takes place using correlated, linguistic and semantic features. The goal of keyword expansion is to improve precision and recall. Relation between Keywords and the elements of data takes place by using a keyword-element relationship. Here, two types of search techniques. One is element level search and another is set level search technique. The proposed system uses routing keyword search for queries having many keywords. This improves the performance of keyword search. [14] This way can greatly reduce time and space costs. The general architecture of the proposed system is elaborated using given figure. [2]

Basic Blocks:

A. Keyword Expansion

B Element-level Model

C Set-level Model

D Ranking

International Journal of Advance Research in Science and Engineering
Vol. No.5, Special Issue No. (02), March 2016
www.ijarse.com
IJARSE
ISSN 2319 - 8354

A. Keyword Expansion:

Query Expansion is the measure of codifying a seed problem to improve retrieval performance in information retrieval execution. In the content of web hunt, query expansion involves evaluating users input and expanding the search query to match additional documents. The aim of the element expansion is to revamp precision and recall. [13]

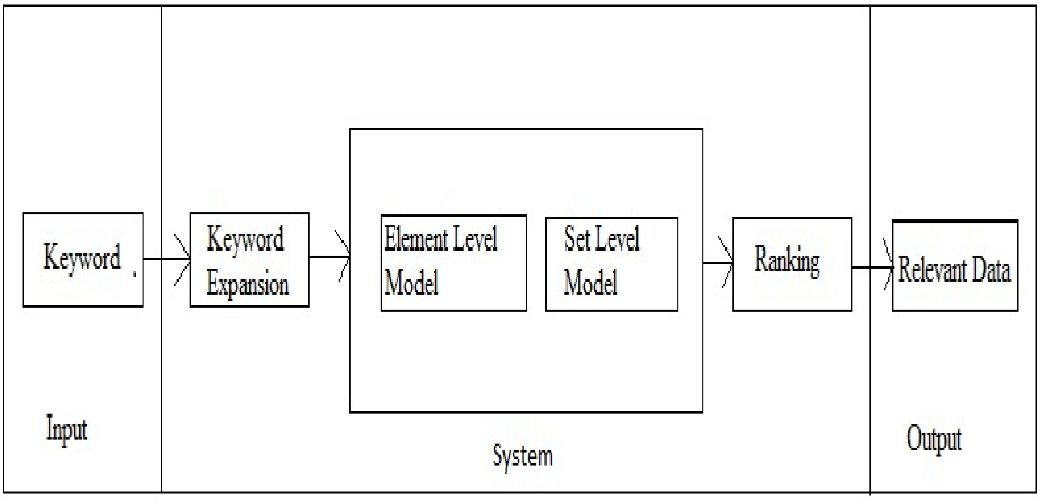

Fig3: System Architecture

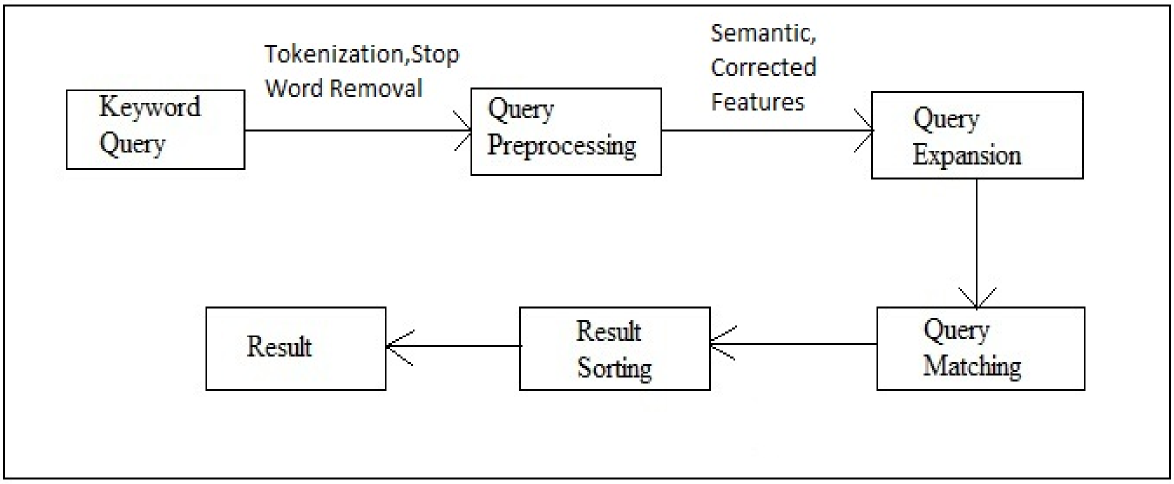

Fig4: Query Expansion

B. Element Level Model:

In this paradigmatic, we mainly concentrate on IR technique of data retrieval. This technique allows users to search unstructured information, and do not need users to understand any database schemas. For Element level search, we used LSI(Latent Semantic Indexing)technique.LSI is an indexing and retrieval method that uses a mathematical technique called single value decomposition (SVD)to identify patterns in the relationship between terms and concepts contained in an unstructured collection of text. It called LSI because of its ability to correlate semantically related terms that are latent in a collection of text. LSI begins by constructing Term Document

International Journal of Advance Research in Science and Engineering
Vol. No.5, Special Issue No. (02), March 2016
www.ijarse.com
IJARSE
ISSN 2319 - 8354

Matrix, to identify occurrence of the m unique terms within a collection of n documents. Each term represent by a row and each document is represented by column.

C. Set-level Model:

Set-level model perform set level search, which captures information about group of elements. This set-level graph substantially take a part of the Linked Data schema on the web that is represented in RDFS, i.e., relations between classes.

D. Ranking

It depends on user hunting a product. Depending upon how many times a product is hunt, a grade number is given to it.

VI. EXPERIMENTAL RESULTS

Data set is nothing but the collection of data used to test the results. Data for this project is fetched online from DBPedia, Freebase and LOD (Linked Open Data) Cloud.

A. Precision:

It is the ratio of the number of significant records regain to the total number of Irrelevant and relevant records retrieved.

It is usually expressed as a percentage.

Precision =(Number of relevant results retrieved)/(Number of results retrieved)

The following table shows precision of direct search (no query expansion) and our approach on 10 search tasks.

Table 1

Sr. No.	Keyword Query	Precision of Direct search	Precision of Our Approach
1	DU	77%	84%
2	SPPU	56%	76%
3	MIT	55%	73%
4	NIT	74%	94%
5	IIT	69%	74%
6	BRAU	84%	89%
7	SNDT	77%	86%
8	JNU	53%	81%
9	MVV	80%	71%
10	TMV	81%	81%

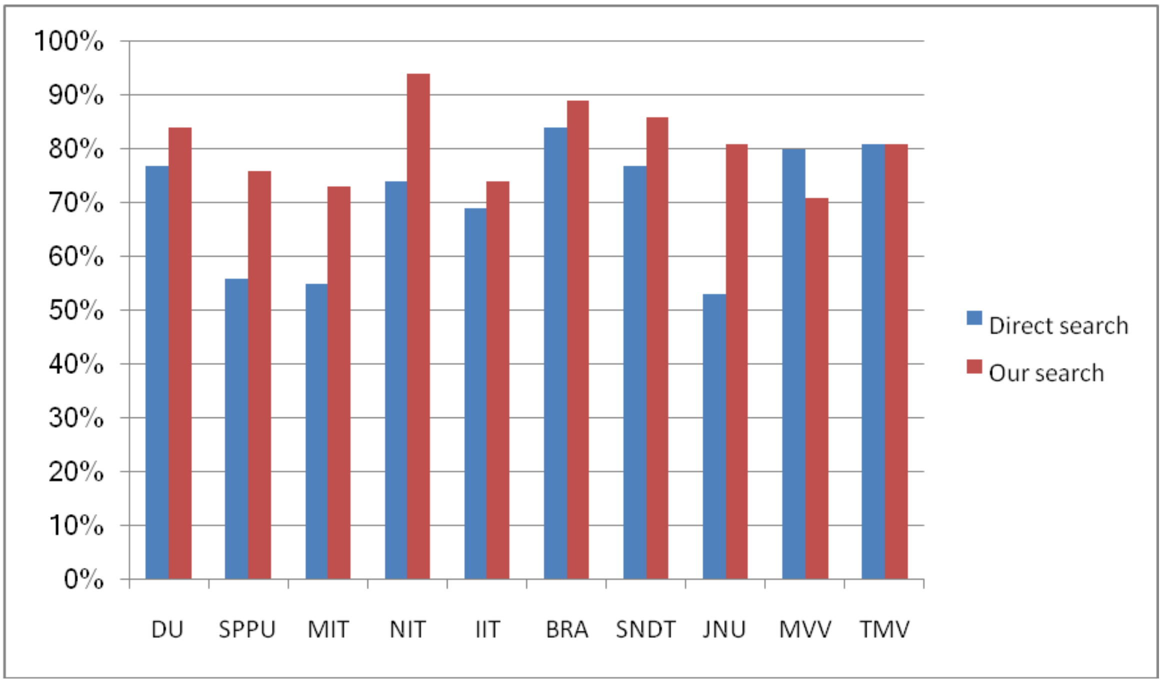

Fig.5: Precision of direct search (no query expansion) and our approach

B. Recall

Recall is the ratio of the number of significant records regain to the total number of relevant records in the database.

It is usually expressed as a percentage.

Recall =(Number of relevant results retrieved)/(Number of relevant docs)

The following table shows recall of direct search (no query expansion) and our approach on 10 search tasks.

Table 2

Sr. No	Keyword Query	Recall of Direct Search	Recall of Our Approach
1	DU	82%	92%
2	SPPU	89%	97%
3	MIT	82%	94%
4	NIT	74%	90%
5	IIT	91%	93%
6	BRAU	51%	84%
7	SNDT	71%	81%
8	JNU	79%	91%
9	MVV	88%	78%
10	TMV	81%	83%

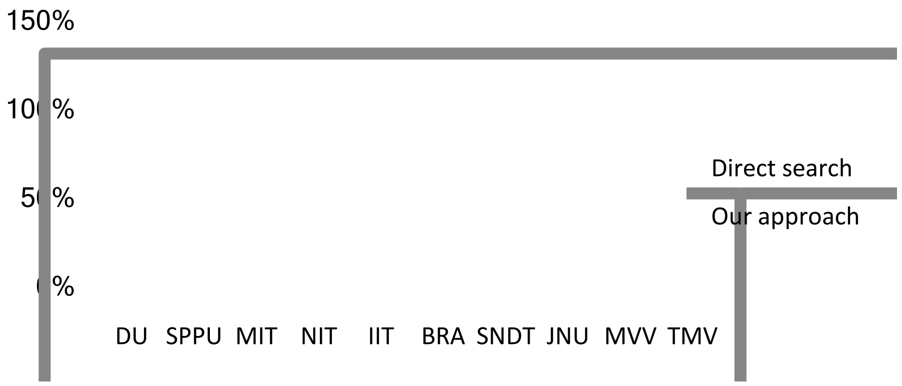

Fig. 6 Recall of direct search (no query expansion) and our approach

B. Time Complexity

The following graph shows results on the basis of time it shows the how much time is required for existing system and how much time proposed system takes for Keyword Search.

Fig.7 Time comparison between Existing and proposed system

Sr. No	Keyword Query	Time required for keyword search with routing(time)	Time required for keyword search without routing(time)
1	University	1613	8837
2	SPPU	3348	13330
3	MIT	3407	6628
4	IIT	2414	6235
5	BRT	1702	2847

Above tables shows that the proposed system is more effective than existing one as it requires less time to get desired query result. Precision and recall values are also improved.

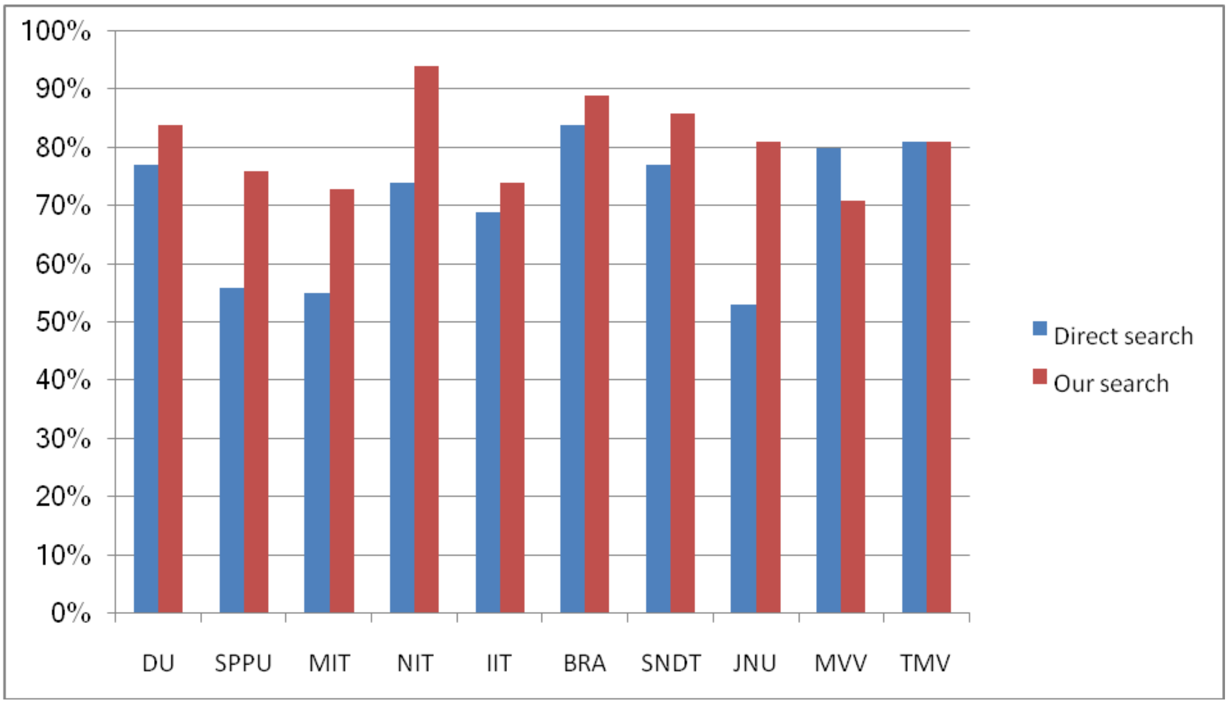

Fig.7 Time comparison between Existing and proposed system

VII. CONCLISION

Efficient keyword query routing system helps to better the performance of keyword search, without composed quality of result. Examine the difficulty of keyword query routing for keyword search over a huge number of structured and Linked Data sources. Routing keywords only to applicable sources can reduce the high cost of searching for structured results that span multiple sources. System can be good results in minimum time, while not compromising too much on quality. The proposed system uses routing keyword search for the queries having multiple keywords. Here, comparative study between proposed system and existing system is studied. From the above results we conclude that, without routing the keyword search is problematic when the number of keywords is large. In Future, our aim is to route audio video multimedia data and reduce searching time for multimedia data.

REFERANSES

[1] A. Wakchaure, K. Pawar and S. Pradhan, “Keyword Search Approach used for Query Routing”, BVCOE &RI, India, vol. 02 Issue: Sep-2015.

[2] Thanh Tran and Lei Zhang, “Keyword Query Routing”. IEEE Transactions On Knowledge And Data Engineering, Vol. 26, No. 2, February 2014.

[3] 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.

[4] M. Sayyadian, H. LeKhac, A. Doan, and L. Gravano, "Efficient Keyword Search Across Heterogeneous Relational Databases," Proc. IEEE 23rd Int'l Conf. Data Eng. (ICDE), pp. 346-355, 2007

[5] 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.

[6] 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.

[7] Guoliang Li,Beng Chin Ooi et al.,"EASE: An Effective 3-in-1 Keyword Search Method for Unstructured, Semi-structured and Structured Data", Proc. ACM SIGMOD Conf., pp. 903-14, 2008.

[8] F. Liu, C.T. Yu, W. Meng, A. Chowdhury, "Effective Keyword Search in Relational Databases," Proc. ACM SIGMOD Conf.,pp. 563-574, 2006.

[9] Suwarna G and Srujana.p, "Approaches for keyword query routing", vol. 4, oct-2014.

[10] Pawar P and N. Bhale, "Effective keyword query routing for search engine", vol. 6(1),2015.

[11] N. Saranya, R.Rajeshkumar and S.Saranya, "A Survey on Keyword query routing in database" vol. 2277-9655, 2015.

[12] K. Sudheer and P. Jagdeesh, "An effective keyword Query Routing in web application", vol. 2348-4845.

[13] Pawar P, "Query expansion for efficient keyword query routing system", vol. 5, Jully 2015.