International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 02 Issue: 06 | Sep-2015 www.irjet.net p-ISSN: 2395-0072

Keyword Search Approach used for Query Routing

Ashwini Wakchaure [1], Leena Nerkar [2], Kirti Pawar [3], Snigdharani Pradhan [4]

[1234] UG Student, Dept. of Comp. Engg., BVCOE & RI, Maharashtra, India
[5]UG Lecturer, Dept. of Comp. Engg., BVCOE & RI, Maharashtra, India

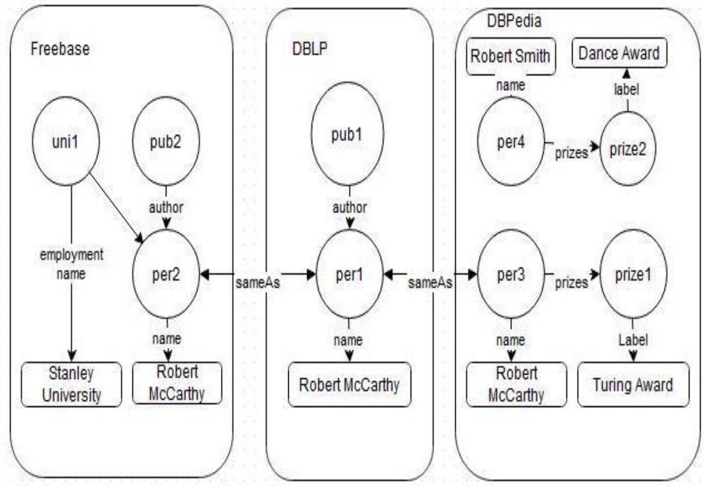
---***---

Abstract - ***The web is a source for searching information. But web is a operation that provides link about searching keyword. Normally, if retrieve any data from databases, then query can be formed for searching keywords. but this difficult for typical web uses. Because the query can be formed by means at structured queries using language like in SQL, MONGODB etc. Many at the approaches used only the single-source for database research. The important thing is computing the combination at most relevant sources for routing the keyword only for relevant sources, the novel method is proposed named as Top-k method for computing the routing plan & produce a result for given keyword search the keyword element relationship summary shows the relations between keywords & data element. Next, for computing the relevant at routing plan propose a multilevel scoring mechanism based on the level at keywords, data elements & lastly shows the performance at keyword search.***

Key Words*: Keyword search, Routing plan, Keyword query, RDF, Graph structured data, etc…*

1. INTRODUCTION

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]. Same as links describe two resource description framework that is RDF shows the same real world. The linked data on the web as shown in fig.1



like SQL, MONGODB etc. Technical users have knowledge about SQL language, so they can quickly exploit the web data but non-technical users do not have any knowledge regarding query language. So, this task is difficult for non-technical users. Each and every user can search the data by using keyword search for searching the keyword on the web do not required any knowledge regarding query languages, underlying or the schema.

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]. But this solution is used only for single source. They are not directly applicable to the web at linked data [1]. 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.

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]. The existing system uses the keyword relationship (KR) collected individually for single databases [1][2]. 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] [2]. This summary is important for address the scalability requirement of the linked data web scenario [1].

IR-style ranking has been proposed to incorporate relevance at the level of keywords [1]. 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].

The total paper arranged in four sections. First section provides the work related to existing system. In second section we provide the whole system architecture. Third section provides the methods for system and in last section shows the algorithm of system.

2. RELATED WORK

Fang Liu, Clement Yu et al., projected a unique IR ranking Strategy for effective keyword search. the primary that compact comprehensive experiments on search effectiveness employing a world info and a group of keyword queries collected by a significant search firms. This strategy is considerably higher than existing ways. [4]. Guoliang Li. et al., projected AN economical 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 to 1st model unstructured, semi-structured and structured knowledge as graphs, and so summarize the graphs and construct graph indices rather than exploitation ancient inverted indices.[5]. 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 practicality usually on the market in business RDBMSs. It develops query-processing methods that ride an important characteristic of IR-style keyword search: solely the few most relevant matches –according to some definition of "relevance"– area unit usually of interest. [6].
Yi Nilotic Xuemin sculpture dynasty Wang et al., studies the effectiveness and also the potency problems with respondent top-k keyword question in electronic database systems. It planned a brand new ranking formula by adapting existing IR techniques supported a natural notion of virtual document. Compared with previous approaches, this new ranking methodology is straightforward however effective, and agrees with human perceptions. It studied economical question process ways for the new ranking methodology, and propose algorithms that have stripped-down accesses to the information [7]. Quang Hieu Vu et al., proposes GKS, a completely unique technique for choosing the top-K candidates supported their potential to contain results for a given question. GKS summarizes every info by a keyword relationship graph, wherever nodes represent terms and edges describe relationships between them. Keyword relationship graphs area unit used for computing the similarity between every info and a American state question, so that, throughout question process, solely the foremost promising databases area unit searched. [8]

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 manufacturing question results that span multiple databases and relations. Kite then exploits the joins – discovered mechanically across the knowledge bases – to alter quick and effective querying over the distributed data.[9] Bei Yu et al., study the information choice downside for relative knowledge sources, and propose a technique that effectively summarizes the relationships between keywords in a very computer database supported its structure. It develop effective ranking strategies supported the keyword relationship summaries so as to pick the foremost helpful databases for a given keyword question. It enforced this technique on Planet workplace. therein atmosphere we tend to use intensive experiments with real datasets to demonstrate the effectiveness of this projected account methodology.[10]

TABLE 1: COMPARATIVE STUDY

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	An adaptive method, EASE, for indexing &

	Search Method for Unstructured, Semi-structured and Structured Data	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

3. 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] A set-level part is contained in a very supply. there's a grip between 2 keywords if 2 components at the part level mentioning these keywords area unit connected supported a path.[1][3]

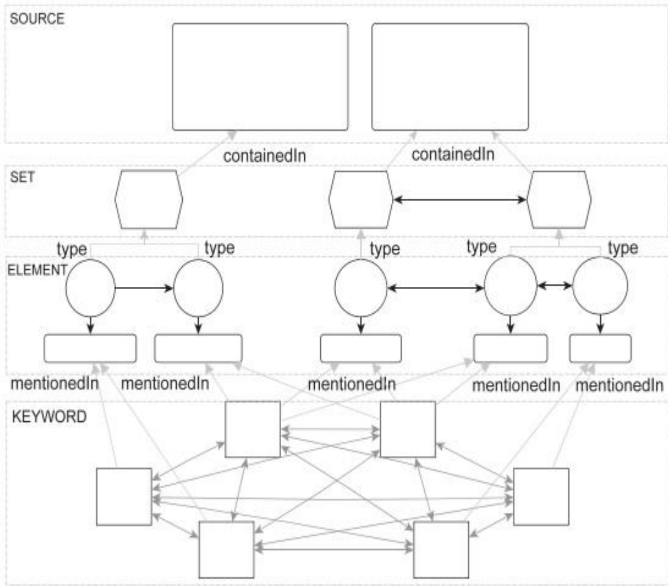


Figure 2. System Architecture [1]

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][3] 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] & in database selection, KRG association are retrieved for all pairs of query keywords to construct a sub graph.[1] 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.[3] Multisource KRG models both relationships that are within & those that link between sources.[1] Keyword relationships are stored together with the elements are associated with source information. An element-level key-element relationship graph (E-KERG).[3]

4. APPROACHES FOR KEYWORD QUERY ROUTING

There are four approaches for Keyword Query Routing :

1) Upload Details to Linked Data Sources
2) Keyword Search using multilevel inter relationship
3) Compute Routing Plans
4) Get Search Results

4.1 Upload Details to Linked Data Sources

First User transfers his own details to linked data sources. Linked data sources area unit connected info. Existing work uses keyword relationships (KR) collected on an individual basis for single databases. This paper represents relationships between keywords in addition as those between information parts. The goal is to provide routing plans, which may be wont to figure results from multiple sources.

S = {s , e , X , Y}

Where,

s = Start state of module.

e = End state of module.

X = Input parameters

Y = output of module

$X = \{w^*(G^*, N^*, \varepsilon_i^* \uplus \varepsilon_e^*)\}$

Where,

G^* = {set of all data groups}

$G^* = \{g_1(N_1^*, \varepsilon_1^*), g_2(N_2^*, \varepsilon_2^*), \ldots\ldots\ldots, gn(N_n^*, \varepsilon_n^*)\}$

N^* = {set of all nodes }

$N^* = \{U^{n}{}_{l} = 1N^*_{L}\}$

ε_i^* = {set of all internal edges that connects element within a particular source}

$\varepsilon_i^* = \{U^{n}{}_{l=1}\, \varepsilon^*_{il} \cup \varepsilon^*_{el}\}$

ε_e^* = {set of all external edges which establish between elements of two different sources}

$\varepsilon_e^* = \{e(ni , nj) \mid ni\ eNi^*, nj\ eNj^*, Ni^* \neq N^*j\}$

4.2 Keyword Search Using Multilevel Inter Relationship

A keyword question is processed by mapping keywords to parts of the info. Then, victimization the schema, valid is part of sequences area unit derived that area unit then used to affix the computed keyword parts to create alleged candidate networks representing potential results to the keyword question. Schema-agnostic approaches operate directly on the information. Structured results area unit computed by exploring the underlying information graph. The goal is to seek out structures within the information referred to as Steiner trees (Steiner graphs in general), that connect keyword parts.

4.3 Compute Routing Plans

Routing plans are computed by looking for Steiner graphs. Given a question K and also the outline, the algorithmic program computes a collection of routing plans. For this, it 1st determines the part of set up JP. Supported this set up, KERG relationships are retrieved for each keyword, pair, and joined with the intermediate result table. This table contains candidate routing graphs, together with the legion their constituent parts and their combined score. Once the part of set up is worked off, the combined score is computed for each tuple in T, i.e., for each routing graph W0S K.

S = {s , e , X , Y }

Where,

s = Start state of module.

e = End state of module.

X = Input parameters

Y = output of module

X = {The web graph W = (G , N , ε) & keyword query K}

The mapping $\mu : K \rightarrow 2^G$ that associates a query with a data graphs of set is called a keyword routing plan.

Y = {Set of routing plans}

A plan RP is considered valid w.r.t. K when the union set of its data groups contains a result for K.

4.4 Get Search Results

Routing graphs represent constant set of sources, are aggregate into one single result. This can be as a result of we wish to output solely those plans that capture distinctive combination of sources. Keyword search is associate degree intuitive paradigm for looking joined knowledge sources on the online. Finally we tend to get the relevant results through routing plans.

S = {s , e , X Y}

Where,

s = Start state of module.

e = End state of module.

X = Input parameters

Y = Output of module

X = { A web graph W (N , ε) contains a result for a query i.e K}

K = {K1,K2,K3,………,Kn}

Y = {($n_i \leftrightarrow n_j$) for all n_i, nj ∈ N^s}

Note : Path between n_i and n_j for all n_i, $n_j \in N^s$.

5. COMPUTING ROUTING PLANS

Routing plans are computed by searching for Steiner graphs, a routing graph contains a set of data sources. This data sources contains information that enables user to access this information & then user check whether it is relevant or not. A routing plan is relevant only if the nodes represent the keywords related to the intended information need.

Algorithm: Compute Routing Plan(K, W'_k) [1]

Input: The query K, the summary W'_k (N'_k, $_k$)

Output: set of routing plans [RP]

JP ← a join plan that contains all (k_i , k_j) 2^k ;

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**

(k_i , k_j) ← JP .pop() ;

(k_i , k_j) ← retrieve($_k$, (k_i , k_j));

If T .empty() **then**

T ← (k_i , k_j);

else

T ← (k_i , k_j) T ;

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;

6. CONCLUSIONS

It gives a solution to the different problem of database query routing. It is also based on modeling the search space as a multiple level inter-relationship graph, the summary model is presented that groups the keyword and element association at the level of sets, and established a multiple level line up scheme to incorporate relevance at variance dimensions. More ever, essential performance gain can be achieved when routing is applied to an existing keyword search system to snip sources.

ACKNOWLEDGEMENT

We wish to express our honest gratitude to Prof. C.K. Patil, principle and Prof. H.D. Sonawane, H.O.D of Computer Department for providing me on opportunity for presenting Paper on "Keyword Search Approach used for Query Routing". We sincerely that to our paper guide Lecturer K.R. Wagh for her guidance and encouragement for completing the paper work. We wished to express our gratitude to the officials and especially our staff members who render their help during the period of our paper last but not least we wish to avail our self of this opportunity, express a sence of gratitude and love to our friends and our parents for their manual support, strength, help for everything.

REFERENCES

[1] Thanh Tran and Lei Zhang, "Keyword Query Routing". IEEE Transactions On Knowledge And Data Engineering, Vol. 26, No. 2, February 2014.
[2] Prachi M. Karale, Natikar S.B., "A Survey on Keyword Query Routing", International Journal of Advance Research in Computer Science & Management Studies, Vol. 2, Issue 11, Nov.2014
[3] N.Saranya, R.Rajeshkumar, et al., "A Survey on Keyword Query Routing in Databases", International Journal of Engineering Sciences & Research Technology, March 2015.
[4] F. Liu, C.T. Yu, W. Meng, A. Chowdhury, "Effective Keyword Search in Relational Databases," Proc. ACM SIGMOD Conf.,pp. 563-574, 2006.
[5] 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.
[6] 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.
[7] 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.
[8] 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.
[9] 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.
[10] 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.