International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 11, November 2015
Available at http://internationaljournalofresearch.org

Keyword Query Routing in Linked Database

G. Brahmanaidu[1]& P. Anusha [2]

[1]PG Scholar,Dept of CSE, Qis College of Engineering and Technology, Ongole,Prakasam Dist,Andhra Pradesh

[2]Assistant Professor,Dept of CSE, Qis College of Engineering and Technology, Ongole,Prakasam Dist,Andhra Pradesh

ABSTRACT:

A symbol of operation that looks for twin documents that carry one or more than line specified by the individual is called keyword search. Detect the content we demand. 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. This way can greatly reduce time and space costs.

Keywords-- Data mining; keyword search; keyword query; keyword query routing; routing plan; query expansion

I. INTRODUCTION

Recently web has evolved from linked documents. Linked data is an approach to publishing and sharing data on the web. Data from different domains i.e. companies, people, books, films, scientific publications, television ,music, and radio programs, proteins, genes, clinical trials and drugs, online communities, scientific data and statistical, and reviews. We propose to enquire the difficulty of keyword query routing over a huge number of Linked and structured Data sources for keyword search. To route keywords to appropriate sources reduces the cost of searching. We use a graph-based information model to characterize separate collection sources. In that role model, we describe between an element-level data graph, which represents relationships between separate data elements, and a set-level graph of data, which takes data about group of elements. This set-level graph takes a part of the Linked Data schema that is in RDFS, i.e., relations between classes.

The web is not only text data but also interlinked data. Querying large amount of data is challenging. Linked Data consists of hundreds of sources. Those sources having billions RDF triples, which are connected by millions of links course. While dissimilar forms of links can be formed, the often published are same As, which shows that two RDF resources show the same real-world object.

The linked data contains data in different areas, such as ecommerce, the biosciences, and e-government. To search linked data we use this keyword search method which use keyword query routing. To reduce the more cost in searching structured results that span many sources, we propose to route the keywords to the appropriate databases. The aim is to create routing plans, which can be used to calculate results from many different sources. For selecting the true routing plan, we use graphs that are based on the relationships among different keywords which are present in the keyword query. This is considered at the many different levels such as keyword, element, set level etc.

In contrast to internet two.0 mashups that work against a set set of knowledge sources, coupled knowledge applications operate high of AN unbound, world knowledge area. this permits them to deliver a lot of complete answers as new knowledge sources seem on the online. we tend to propose to analyze the matter of keyword question routing for keyword search over an outsized range of structured and coupled knowledge sources. Routing keywords solely to relevant sources will scale back the high value of finding out structured results that span multiple sources. To the most effective of our data, the work given during this paper represents the primary conceive to address this downside. We use a graph-based information model to characterize individual information sources. in this model, we tend to distinguish between Associate in Nursing element-level information graph representing relationships between individual information parts, and a set level information graph, that captures data concerning cluster of parts. This set-level graph basically captures a region of the joined information schema on the net that's described in RDFS, i.e., relations between categories. Often, a schema could be incomplete or just doesn't exist for RDF information on the net. In such a case, a pseudo schema may be obtained by computing a structural outline like a dataguide.

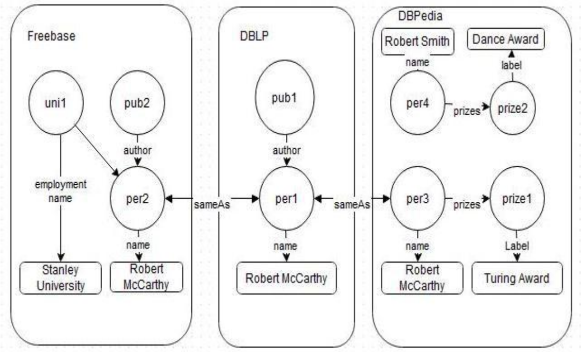

Fig.1. Example of linked data on web.

The web is not any longer a set of matter information however conjointly an online of interlinked information sources. One project that for the most part contributes to the current development is Linking Open information. Through this, an enormous quantity of structured data was created publically accessible. Querying that vast quantity of knowledge in Associate in nursing intuitive means is difficult. Jointly, joined information comprises many sources containing billions of RDF triples, that area unit connected by lots of links. While completely different sorts of links may be established, those oftentimes printed area unit same As links, that denote that 2 RDF resources represent constant real-world object. The illustration of the joined information on the net. The joined information net already contains valuable information in numerous areas, like e-government, ecommerce, and therefore the biosciences. to boot, the amount of accessible datasets has fully grown solidly since its origin [2]. so as to go looking such information we tend to use keyword search techniques that use keyword question routing as shown in Fig.1. To decrease the high price incurred in looking structured results that span multiple sources, we tend to propose routing of the keywords to the relevant databases.

As hostile the supply choice downside [3], that is that specialize in computing the foremost relevant sources, the matter here is to reason the foremost relevant combos of sources. The goal is to supply routing plans, which may be wont to reason results from multiple sources. For choosing the right routing set up, we tend to use graphs that area unit developed supported the relationships between the keywords gift within the keyword question. This relationship is taken into account at the assorted levels like keyword level, component level, set level etc.. Existing system investigates the downside of keyword question routing for keyword search over an oversized range of structured and joined information sources. Supported modeling the search area as a construction inter-relationship graph, a outline model is employed for grouping keyword and component relationships at the extent of sets. It uses a construction ranking theme to include connectedness at completely different dimensions. This method doesn't reason near uses many mechanisms to prune some answers. It couldn't handle queries with multiple keywords with efficiency.

II. Related work:

Keyword Query Search can be divided into two directions of work. They are: 1) keyword search approaches compute the most relevant structured results and 2) Solutions for source selection compute the most relevant sources.

2.1. Keyword search

In the keyword searching, we mainly follow two approaches. They are *schema-based approaches* and *schema-agnostic approaches*.

Schema-based approaches are implemented on top of off-the-shelf databases. A keyword is processed by mapping keywords to the elements of the databases, called *keyword elements*. Then, using the schema, valid join sequences are derived and are employed to join the computed keyword elements to form the candidate-networks that represent the possible results to the keyword query.

Schema-agnostic approaches operate directly on the data. By exploring the underlying graphs the structured results are computed in these approaches. Keywords and elements which are connected are represented using Steiner trees/graphs. The goal of this approach is to find structures in the Steiner trees.

For the query “Stanley Robert Award” for instance, a Steiner graph is the path between uni1 and prize1 in Fig. 1. Various kinds of algorithms have been proposed for the efficient exploration of keyword search results over data graphs, which might be very large. Examples are bidirectional search [3] and dynamic programming [4].

Recently, a system called Kite extends schema-based techniques to find candidate networks in the multi source setting [5]. It employs schema matching techniques to discover links between sources and uses structure discovery techniques to find foreign-key joins across sources. Also based on pre computed links, Hermes [6] translates keywords to structured queries.

2.2 Database Selection

In order to get the efficient results for keyword search, the selection of the relevant data sources plays a major role. The main idea is based on modeling databases using keyword relationships. A keyword relationship is a pair of keywords that can be connected via a sequence of join operations. For instance, (Stanley, Award) is a keyword relationship as there is a path between uni1 and prize1 in Fig. 1. A database is considered relevant if its keyword relationship model covers all pairs of query keywords.

M-KS considers only binary relationships between keywords. It incurs a large number of false positives for queries with more than two keywords.

This is the case when all query keywords are pair wise related but there is no combined join sequence which connects all of them.

G-KS [7] addresses this problem by considering more complex relationships between keywords using a Keyword Relationship Graph (KRG). Each node in the graph corresponds to a keyword. Each edge between two nodes corresponding to the keywords (ki, kj) indicates that there exists at least two connected tuples ti ↔ tj that match ki and kj. Moreover, the distance between ti and tj are marked on the edges.

III. THE PROPOSED SYSTEM

To route keywords only to relevant sources to reduce the high cost of processing keyword search queries over all sources. A novel method was used 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 was 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. Also to investigate the problem of keyword query routing for keyword search over a large number of structured and Linked Data sources. This system was having more advantages: 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.

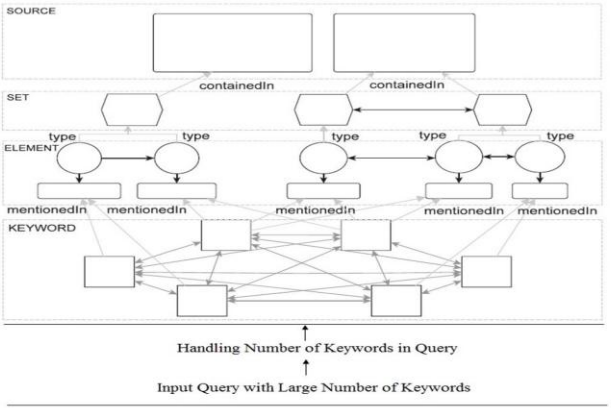

Figure 2: Inter Relationship between Elements

However, queries with more keywords could not be handled efficiently. For instance, queries with more than two keywords needed several seconds up to one minute. Thus, while this setting produced results of highest quality, it is not really affordable in a typical web scenario demanding high responsiveness. To produce results in minimum time, while not compromising too much on quality. The results suggest that keyword search without routing is especially problematic when the number of keywords is large. Thus the proposed system uses routing keyword search for the queries having large number of keywords.

The search space of keyword query routing using a multilevel inter-relationship graph. At the lowest level, it models relationships between keywords. In the upper most levels, there are W (N, ε) and the source-level web graph, which contains sources as nodes. The inter-relationships between elements at different levels are illustrated in Figure 2. 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. Fig.2 represents a holistic view of the search space. Based on this view, we propose a ranking scheme that deals with relevance at many levels. Further, Fig. provides different perspectives on the search space. Based on this representation of the search space, existing work on keyword search and database selection can be extended to solve the problem of keyword query routing.

For selecting the correct routing plan, we use graphs that are developed based on the relationships between the keywords present in the keyword query. This relationship is considered at the various levels such as keyword level, element level, set level e.t.c. The goal is to produce routing plans, which can be used to compute results from multiple sources. However, queries with more keywords could not be handled efficiently. For instance, queries with more than two keywords needed several seconds up to one minute. Thus proposed system tries to handle such queries with number of keywords and tries to minimize the computing time.

IV. 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}
$G^* = \{g1(N1^*, \varepsilon 1^*), g2(N2^*, \varepsilon 2^*),\ldots\ldots\ldots, gn(Nn^*, \varepsilon n^*)\}$
N^* = {set of all nodes }
$N^* = \{Unl = 1N^*L\}$

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

$\varepsilon i^* = \{\cup nl=1\ \varepsilon^* il \cup \varepsilon^* el\}$

εe^* = {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 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

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 μ : K → 2G 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 = {(ni ↔ nj) for all ni , nj ε Ns} Note : Path between ni and nj for all ni , nj ε Ns.

V. CONCLUSION

This paper helps to enhance the performance of keyword search, while not compromising its result quality. Investigate the matter of keyword question routing for keyword search over an outsized variety of structured and joined information sources. Routing keywords solely to relevant sources will cut back the high value of checking out structured results that span multiple sources. We have a tendency to use a graph-based information model to characterize individual data sources. For choosing the proper routing arrange, we have a tendency to use graphs that area unit developed supported the relationships between the keywords gift within the keyword question. This relationship is taken into account at the varied levels like keyword level, part level, set level e.t.c. within the existing system, Routing keywords come back all the supply which can or might not be the relevant sources. However, queries with a lot of keywords couldn't be handled expeditiously. For example, queries with quite 2 keywords required many seconds up to at least one minute. Thus, whereas this setting created results of highest quality, it's not very cheap in a very typical net state of affairs hard-to-please high responsiveness. To provide leads to minimum time, whereas not compromising an excessive amount of on quality. The results counsel that keyword search while not routing is particularly problematic once the amount of keywords is massive. so the planned system uses routing keyword rummage around for the queries having sizable amount of keywords.

IV.References

[1] T. Berners-Lee, "Linked Data Design Issues," 2009;www.w3.org/DesignIssues/LinkedData.html

[2] 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.

[3] V. Kacholia, S. Pandit, S. Chakrabarti, S. Sudarshan, R. Desai, and H. Karambelkar, "Bidirectional Expansion for Keyword Search on Graph Databases," Proc. 31st Int'l Conf. Very Large Data Bases (VLDB), pp. 505-516, 2005.

[4] B. Ding, J.X. Yu, S. Wang, L. Qin, X. Zhang, and X. Lin, "Finding Top-K Min-Cost Connected Trees in Databases," Proc. IEEE 23rd Int'l Conf. Data Eng. (ICDE), pp. 836-845, 2007.

[5] 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.

[6] T. Tran, H. Wang, and P. Haase, "Hermes: Data Web Search on a Pay-as-You-Go Integration Infrastructure," J. Web Semantics, vol. 7, no. 3, pp. 189-203, 2009.

[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] Jianhua Feng, Guoliang Li and Jianyong Wang, "Finding Top-k answers in keyword search over relational databases using tuple units" IEEE transactions, VOL. 23 NO. 12, December 2011.

[9] G. Li, B.C. Ooi, J. Feng, J. Wang, and L. Zhou, "Ease: An Effective 3-in-1 Keyword Search Method for Unstructured, Semi-Structured and Structured Data," Proc. ACM SIGMOD Conf., pp. 903-914, 2008.

[10] R. Goldman and J. Widom, "DataGuides: Enabling Query Formulation and Optimization in Semistructured Databases," Proc. 23rd Int'l Conf. Very Large Data Bases (VLDB), pp. 436-445, 1997.

[11] Thanh Tran and Lei Zhang, "Keyword Query Routing" IEEE Transactions, VOL.26, NO.2, February 2014.

[12] K. Collins- Thompson, Reducing the risk of query expansion via robust constrained optimization. In *CIKM*. ACM, 2009.