IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 12, December 2015

A Review on Keyword Based Web Search using Multilevel Ranking

Vivek Kale[1], Dr. K. V. Metre[2]

PG Student, Computer Engineering Department, MET'S IOE, Nashik, Maharashtra, India[1]

Asstt. Professor, Computer Engineering Department, MET'S IOE, Nashik, Maharashtra, India[2]

Abstract: Keyword search is a type of search in which matching of related documents containing one or more related words specified by the user. The web not only includes textual record but also include internet of interlinked knowledge. The linked data already contains valuable data in diverse areas, such as e-government, e-commerce, and the biosciences. The growing number of datasets published on the web are consider as linked knowledge which brings possibilities of prime data availability of information. As the information increases challenges for querying also increases. It is very problematic to search in linked data using structured languages. Hence, Keyword Query searching for linked data is used. Different approaches for keyword query routing through which the efficiency of keyword search can be improved greatly. Through routing the key phrase to relevant data source the cost of processing can be reduced. The Multilevel Scoring Mechanism is used to find top-k result for relevant document retrieval. By using this mechanism relevant data can be retrieved effectively and efficiently.

Index Terms: Keyword search, Keyword query routing, Graph-structured data, RDF.

I. INTRODUCTION

The web is collection of textual document and linked data i.e. web of interlinked data source. A tremendous amount of legacy information have been modified to Resource Description Framework (RDF) linked with different sources, and published as Linked knowledge. Linked information includes thousands of sources containing billions of RDF triples, which are connected with the aid of millions of hyperlinks. Whilst special varieties of hyperlinks may also be established, those quite often released are same as links, which denote that two RDF resources symbolize the identical real-world object.

It used a graph-based data model to symbolize person knowledge sources. In that mannequin, it distinguish between an element level knowledge graph representing relationships between person data factors, and a collection level knowledge graph, which captures information about work force of elements. This set level graph captures part of the Linked information schema from the web which is represented in RDFS, i.e., family members between classes. Commonly, a schema possibly incomplete or effectively does not exist for RDF knowledge on the net. In this kind of case, a pseudo schema can also be bought by using computing a structural abstract summery to a data consultant. The net is no longer a group of textual information but also an online of interlinked data sources. A huge amount of structured expertise was once made publicly on hand. Querying that significant quantity of knowledge in an intuitive approach is challenging. Collectively, Linked data incorporate hundreds and hundreds of sources containing billions of RDF triples, which might be linked by using hundreds of thousands of hyperlinks. Even as different forms of links can also be founded, the ones regularly published are same as links, which denote that two RDF resources represent the identical actual-world object. The linked knowledge internet already contains useful knowledge in numerous areas, reminiscent of e-govt, e-commerce, and the biosciences. Additionally, the quantity of available datasets has grown solidly since its inception. In an effort to search such data, it used key phrase search procedures which employ keyword search routing.[1] To cut down the high rate incurred in searching structured outcome that span multiple sources, keyword routing is used on the critical databases. As opposed to the source resolution drawback,[2] which is focusing on computing the most imperative sources, the drawback right here is to compute probably the most relevant combinations of sources. The purpose is to produce routing plans, which can be used to compute outcome from multiple sources. It used a graphs which can be developed established on the relationships between the key phrases gift within the key phrase query. This relationship is viewed at the more than a few stages reminiscent of keyword degree, element stage, set degree etc.

II. LITERATURE REVIEW

Different keyword search techniques have been studied by various authors. To get relevant data various novel method are used in various techniques. For searching on linked data a novel method is used to find Top K result. In BLINKS, it used novel method for searching on graph data. It used bi-level indexing and query processing scheme which reduces index spaces. It provides performance bound and search strategy. In this data graph is divided into blocks. [6]

As rapid growth in web database, new search strategy in information retrieval used ranking strategy for effective keyword search. It author proposed novel IR Ranking

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 12, December 2015

strategy for effective keyword search. This approach can be used at the application level and also incorporated into a RDBMS to support keyword-based search in relational databases. [9]

As the relational database contain more and more text data, so it is necessary to support keyword query over text data in relational database. For effectiveness and efficiency over existing techniques SPARK provide novel method which used new ranking formula by adopting existing information retrieval technique based on natural notation of virtual documents. It used method for minimal database access and provides efficient ranking formula by adopting top k query in relational database system. [7]

Keyword search is used to find information of interest from relational database. Previous work is focus on single database, so for obtaining result from multiple database join is required on multiples tuples. A new technique known as KITE provides solution to problem of keyword search over heterogeneous databases. It combines schema matching and structure discovery technique to find approximate foreign key join across heterogeneous databases. [8]

A novel method known as EASE provide effective keyword search on structured, semi-structured and unstructured data. It provide technique for indexing and querying on heterogeneous database. In this heterogeneous data is summarize and construct index on graph instead of traditional inverted index. It used extended inverted index for keyword based search and used ranking for enhanced effectiveness of search. It provide high efficiency and accuracy.[3]

III. PROPOSED SYSTEM

There are two things to be taken into consideration

a) Relevant source selection
b) Computation of relevant structure result

Cost of keyword processing can be reduced by routing keyword to relevant source. A novel method was used for computing top-k routing plans based on their potentials to contain result for a given keyword query. A multilevel scoring mechanism is used for computing relevance of routing plans based on scores at level of keyword level, set level, element level etc. Over large number of structured and linked data source searching is carried using keyword routing.

This system have more advantages:

1) By the routing to relevant source reduce high cost of searching over linked data
2) Routing plan is used to compute more relevant record over multiple source.

The system consists of following component

1. Keyword search
2. Keyword Query Routing
3. Multilevel Inter-relationship
4. Set level search

1. Keyword search

Keyword search can be classified in two categories

a) Schema Based approach which implemented on top off the self-database. Mapping of keyword to elements of the database is called as keyword element. [5],[7],[10]
b) Schema-agonistic approaches which is operate on directly on data. Structured results are computed by exploring underlying data graph. [11]

2. Keyword query routing

Routing keywords relevant data source can decrease the high cost of searching over structured results that span multiple sources. Keyword query routing provide solution which prune unpromising sources and enabling users to select combination of results which contain more relevant results.

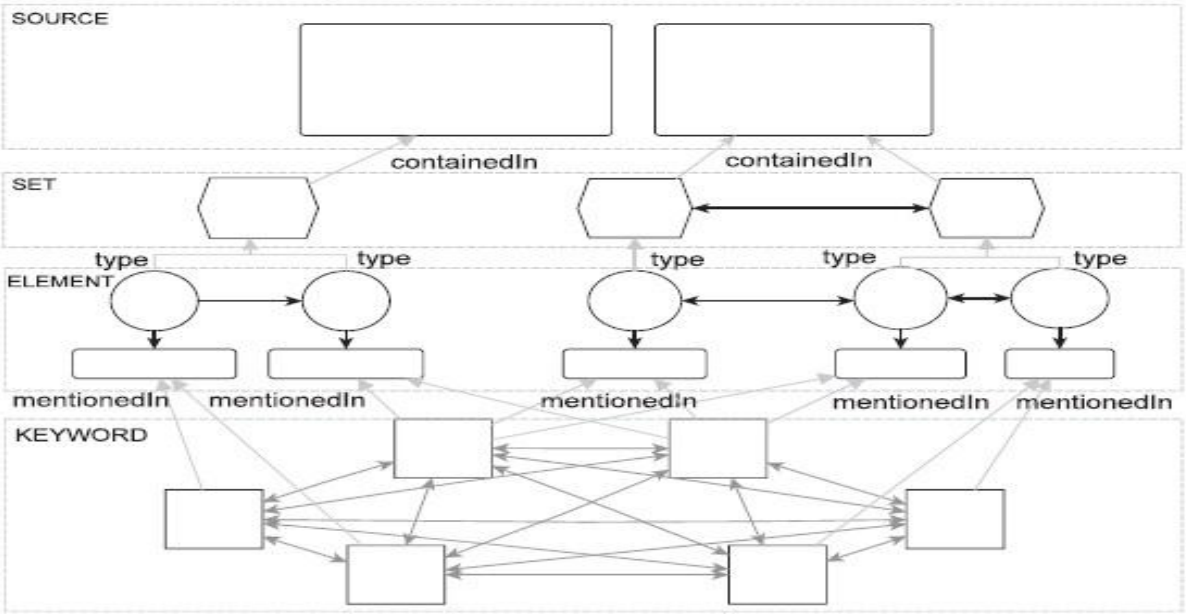

Fig.1: Multilevel Interrelationship Graph

DOI 10.17148/IJARCCE.2015.41206

3. Multilevel Inter-relationship

The search space of keyword query routing using multilevel inter-relationship graph. The different levels of interrelationship are shown in fig. 1. A keyword is described in some entity description at element level. Entities at element level are associated with a set level elements via type. Source contained set-level elements in it. Edge is present between two elements iff two elements are connected by path between them.

4. Set level search

Set level search extract keyword and relationship from data. Keyword-element relationship can be derived based on elements and set level of elements in which they occurs. These relationships are stored in specialized indexes and retrieved at the time of keyword query processing to accelerate the search for Steiner graphs.

Computing Routing Plans

Routing plans are computed by searching for Steiner graphs in the summery contain data source. It contains information that enables used to access relevant data results. Edges in the summery denotes path between elements & subgraph of summery capturing Steiner graphs.

Routing plan can be computed in to three stages. 1) Computation of routing graphs, 2) aggregation of routing graphs, and 3) ranking query routing plan.

IV. CONCLUSION

Web consists of linked data which brings opportunities for high data availability of data. As data increases challenges for searching over data is also increases. Keyword search can be categorized into schema based and schema agnostic approaches. Keyword search approach computes the most relevant structured result. Keyword query search is used for retrieving linked data over web. The Multilevel interrelationship graph is used to find most related result from the linked sources. It searches keyword to relevant source which reduced time required to get results.

REFERENCES

[1] Thanh Tran and Lei Zhang, "Keyword Query Routing", *IEEE Transactions On Knowledge And Data Engineering,VOL. 26, NO. 2, FEBRUARY* 2014

[2] T.Berners-Lee, "Linked Data Design Issue", 2009; www.w3.org/DesignIssue/LinkedData.htm

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

[4] 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

[5] B. Yu, G. Li, K.R. Sollins, and A.K.H. Tung, "Effective Keyword-Based Selection of Relational Database",*Proc. ACM SIGMOD Conf.,* pp. 139-150, 2007.

[6] H. He, H. Wang, J. Yang, and P.S. Yu, "Blinks: Ranked Keyword Searches on Graphs", *Proc. ACM SIGMOD Conf.,* pp. 305-316, 2007.

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

[9] B. Yu, G. Li, K.R. Sollins, and A.K.H. Tung, "Effective Keyword-Based Selection of Relational Databases",*Proc. ACM SIGMODConf.,* pp. 139-150, 2007.

[10] F.Li, C.T. Yu, W. Meng, and A. Chowdhury, "Efficient Keyword search in Relationship Database", *Proc. ACM SIGMOD Conf.,* pp. 563-574,2006.

[11] V. Hristidis and Y. Papakonstantinou, "Discover: Keyword search in Relational Databases", *Proc. 28th Int'l Conf. Very large Data Bases (VLDB),* pp. 695-706, 2009.

DOI 10.17148/IJARCCE.2015.41206