An Effective Mining Approach Using Query Routing

Ms. J. Laveena Grasy[1], Ms R. Soundharya[2] and Ms. Dhanalakshmi[3]

[1, 2, 3] Dept. Of Software Engineering R.V.S School Of Engineering And Technology Dindigul,Tamil Nadu,India
[1]laveenadgl@gmail.com
[2]soundharyacse.2013@gmail.com
[3]danaapr5@gmail.com

ABSTRACT

Keyword search is an important term while using web because every web page is linked with each other, an appropriate keyword enables the minimization of time consumption and cost consumption. To find a prominent solution, which proposed a novel approach of keyword routing which reduces the task in searching keywords? The proposed system computes top-k routing plan that returns the result based on given query. To make the proposed system complete and strong, we employ keyword-element relationship summary that built a strong relationship between the keywords and data elements. The proposed system computes a multilevel scoring mechanism that makes the routing plan based on the relationship and scores of data elements, keywords, elements sets and sub graphs that connects those elements. Our proposed system is experienced in real time processing on linking web which was computed on single pc. The final result proves greatly about the improved performance on keyword searching by succeeding in achieving best quality result.

Keywords*:* Keyword query, keyword search, graph-structured data, Keyword query routing and RDF

1. INTRODUCTION

The current world is fully based on web, in which web is nothing but links which linked each other with certain relationship. In simple it is nothing but a collection of textual documents those large amounts of data which is transformed to RDF and linked that is known as linked data. Those linked data contains enormous number of RDF triples that are connected by several links. The web user feels difficult in getting the exact data from web by means of query structure based on the database languages like SQL or SPARQL. The various terms shows that keyword searching in a highly sensitive process in achieving the exact results. To enable this there is no need of query language or schema is required.

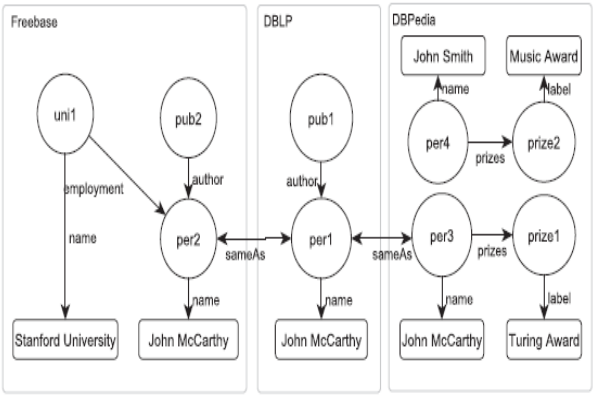

Fig-1: Extract of web data graph

The fig1 shows how the data are extracted from the web, in which sourced are linked in different manner based on certain relationship. It is demonstrated by different methods which are used traditionally. Most common every user who searching on particular data feels to get the expected result in a convenient data. But due to current vast information those expectation is not possible. Without a proper keyword linking makes the

search into lengthy process. To means of achieve this proper routing plan is necessary with exact keyword query. Because this keyword makes possible on better routing plan to get expected result in a better way. While discussing with these keyword search selecting of database is very important because for appropriate result relevant database selection is important. A note should be taken the database must be structured in a similar manner by using keyword relationship. We can see in fig 1 the database is relevant to the keyword relationship model in this case all query were pair wise which connect all of them. In this the problems faced in general more complex information makes the whole process into a tedious one.

2. RELATED WORKS

The keyword search and necessity of low time consumption makes the researchers involvement in developing a prominent approach in dealing with keyword search based on routing plan. Every method deals with the idea describing a solution in overcoming the drawbacks but some terms which is not supported or possible to the real-time scenario making the process in effective. In this database research various decision is carried out for a keyword query for retrieving most relevant data in a structured form [1][2][3][4][5]. It can be also handled in another manner by means of selection most relevant single database [6][7]. Since these are single source location it is failed to applicable on web based linked data. It can't able to succeed against the source selection problems. The major drawbacks in these methods are finding the most relevant source combinations for a better routing plan. In [6][7] they use keyword relationship(KR) method which explains the relationships between the keyword and data elements. It requires a structure for building the collection of linked data but it is not simple and not able to achieve in an effective manner. In the above methods they also introduced IR style ranking but it can't able to build a inter-relationships between elements of different levels.

The most of the tradition ways use the term of keyword elements [8][9][10] it generally uses a join sequences for querying the keywords. In [11][12] they had mentioned about the bidirectional search based on dynamic programming. In this, according kite it extends the schema based techniques in analyzing the candidate network from the multisource links. In which the process is large and the time taken and cost consumption is not reliable to carry the process in reliable manner. Moreover it implies a schema matching technique but for discovering the exact source links to be linked by means of using foreign keys but the process is not that much succeeded to the real time process. According to Hermes he translates the keyword into query structured but based on ex periments is applicable for the minimum number of sources only. In general the existing is described into two main categories such as schema-based approaches and Schema-agnostic approaches. In the first one it follows mapping keywords to the data elements but in next method it's operated directly to it. They are also grouped the element by means of compact summary called the (KERG) set-level keyword-element relationship graph. In this method the relationship summary is very important for addressing the linked data in a web scenario. In [7] they deals by means of G-KS method but it faces the problem complex relationship since no combined join sequence relationships is established between the keywords using a keyword relationship graph. Above all the traditional methods it clears the search space drastically increases along with the number of sources and links between them. The major problem in these of all is the achieved results are not necessary to the user that is it is not the relevant result expected by the user. More than these the problem of routing plan can't be solved by the existing approach.

3. OVERVIEW

The proposed system is designed based on multilevel inter-relationship graph which builds the relationship based on certain keyword relationship. The interrelationships are analyzed in different levels which can be characterized as entities, set element in this the entities are associated with the set level elements. It reduced the cost of searching by investing the problem of keyword query routing over a large structured and linked data sources. Our proposed system implements ranking system which deals with many levels.

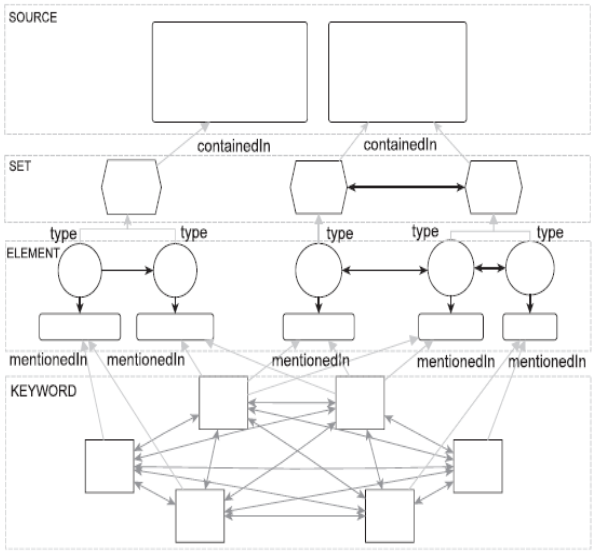

Fig-2: Multilevel interrelationship graph

In the Multilevel interrelationship graph clearly explained that keywords are based on some entities and entities are associated with elements. In which the set- level elements which are present in the source and the keyword relationships must be stored together with the elements they contained. Moreover the information must be relying at the element level in the database. We also support element-level keyword-element relationship graph for solving the keyword routing problem, to do this all sources are collected and constructed as a model but the relationship is large . To overcome the complexity the search space is introduced. While searching the keyword the importance should be taken in path also unwanted path will increased the computational cost. So the paths were retrieved and explored which self-possessed of several edges that must be clear and more expensive than retrieving relationships between keywords.

4. SYSTEM DESIGN

To implement the proposed system in order to achieve the exact result the system is designed with main categories like keyword routing, routing plan and ranking score.

4.1 Keyword Routing: In proposed system, implement keyword level model, element level model, set level model and query expansions. The table 1 sample database is used for analyzing the efficiency of the proposed system. In the keyword level the system is designed by means of connection keyword and keyword query. Here is the filtering stage by using KRG model the keywords contains tuple in the database but it does not confirms that all tuples are connected. So this information's are recorded in the KRG which can derive directly during the process. By means it the keyword search on the database finds the answer by means of tuple in the database using primary or foreign key.

4.2 Routing Plan: The routing plan is designed in such a manner that the information stated which enables the user to assess whether it is relevant or not. The routing plan is designed by using the data contains a set of data sources. It helps to build relationships between the data that corresponds to the intended information required, which helpful in retrieving the reliable path which can be directly repossess and engaged for routing. In addition to it can structured to process in three stages such as computation of routing graphs, aggregation of routing graphs, and ranking query routing plans.

4.3 Ranking Score: In this section, we use multilevel IR-style ranking scheme for achieving the rank of a keyword routing plan. Here the significance score of component keyword-elements and keyword element relationships stored in the ranking of a routing graph. The best ranked routing graphs in RP are extracted that enables us to control the quality. It is not to associate a high rank to those plans, which can be derived from a large number but less relevant graphs.

4.4 Expected Output: The fig3 describes the final result in which the search is given and the sourced links are collected and stored in a database. After this it undergoes data extraction and preparation in that;

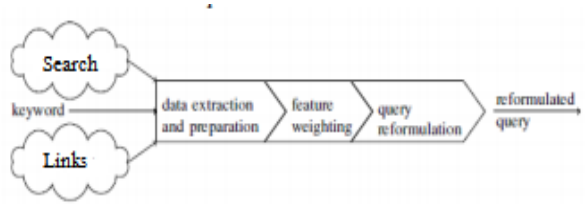

Fig-3: Expected result overflow

4.5 Tokenization: The data are extraction into of individual words and removing the punctuations.

4.6 Stop Word Removal: In stop word the removal of common words which including articles and prepositions should is done.

4.7 Word Lemmatization: This section determines the lemma of the word

After doing this the feature selection and feature weighing is done in thins the routing path along with ranking score is implemented and the resultant query is reformulated and used for search. Thus the sequence of information attained is often used to make decisions which of the features are the most relevant. In which the nodes along with its relationship scores are computed for the entire collection of sources. As a result the validity and measure the relevance of the generated keyword routing plan we use sample data in table 3.

#	Keywords
Q1	Capital Washington
Q2	Barack Obama University
Q3	Film Titanic
Q4	ISWC Ontology Mapping 2004
Q5	ESWC2006 Semantic Search Publication
Q6	Ivan Herman W3C
Q7	Pascal Sebastian OWL Reasoning
Q8	Rudi AIFB ISWC2008
Q9	Studer AIFB Semantic Web
Q10	Semantic Web Publication
Q11	Andreas Harth Semantic Search Engine
Q12	Markus Denny Semantic Wikis
Q13	Information Retrieval Database 2006
Q14	Knowledge Management Conference
Q15	Beijing Conference Database 2007
Q16	Town River America
Q17	Paris Hotel
Q18	Karlsruhe Palace
Q19	Software Project 2003
Q20	Project AIFB 2005

Table-1: Sample keywords

The underlying query is the reformulated queries which generate the efficient result by producing the relevant information thus achieving the computation cost and reduce the complexity as faced by the traditional approaches. It also minimized the time consumption and the searching process more reliable and simple.

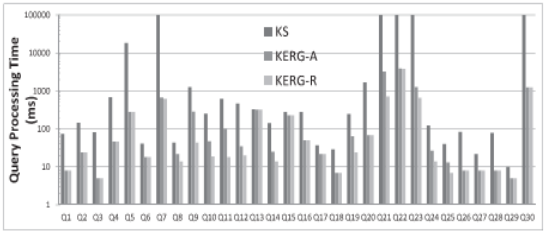

Fig. 11. Performance for KS with/without routing.

Fig-4: Performance of Keyword and Without Keyword Routing

5. CONCLUSION

The process of searching based on keywords is the most effective approach in attaining the exact result by makes use of retrieved data links. Our approach proves the efficiency of doing search using the keyword queries with the help of search space as a multilevel inter-relationship graph by the combination of groups' keyword along with element relationships at the level of sets. To reduce the maximum searching cost our proposed system redirects the keywords to the relevant data sources. We implement keyword routing to redirect the keywords using graph based methods for computing routing plans. The final stage proves the system compactly preserves relevant information in a better way that the existing system has done.

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] 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] 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.

[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] V. Hristidis and Y. Papakonstantinou, "Discover: Keyword Search in Relational Databases," Proc. 28th Int'l Conf. Very Large Data Bases (VLDB), pp. 670-681, 2002.

[9] L. Qin, J.X. Yu, and L. Chang, "Keyword Search in Databases: The Power of RDBMS," Proc. ACM SIGMOD Conf., pp. 681-694, 2009.

[10] G. Li, S. Ji, C. Li, and J. Feng, "Efficient Type-Ahead Search on Relational Data: A Tastier Approach," Proc. ACM SIGMOD Conf., pp. 695-706, 2009.

[11] 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.

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

