

A Peer Reviewed Open Access International Journal

An Effective Keyword Query Routing in Web Applications

K.Sudheer Reddy

M.Tech Student, Department of Computer Science and Engineering, VEMU Institute of Technology, Pakala, Chittoor Dist.

Abstract:

Keyword search is an intuitive paradigm for searching linked data sources on the web. We propose to route keywords only to relevant sources to reduce the high cost of processing keyword search queries over all sources. We propose a novel method for computing top-k routing plans based on their potentials to contain results for a given keyword query. We employ a keyword-element relationship summary that compactly represents relationships between keywords and the data elements mentioning them. A multilevel scoring mechanism is proposed for computing the relevance of routing plans based on scores at the level of keywords, data elements, element sets, and subgraphs that connect these elements. Experiments carried out using 150 publicly available sources on the web howed that valid plan (precision@1 of 0.92) that are highly relevant (mean reciprocal rank of 0.89) can be computed in 1 second on average on a single PC. Further, we show routing greatly helps to improve the performance of keyword search, without compromising its result quality.

1. INTRODUCTION:

In recent years the Web has evolved from a global information space of linked documents to one where both documents and data are linked. Underpinning this evolution is a set of best practices for publishing and connecting structured data on the Web known as Linked Data. The adoption of the Linked Data best practices has lead to the extension of the Web with a global data space connecting data from diverse domains such as people, companies, books, scientific publications, films, music, television and radio programs, genes, proteins, drugs and clinical trials, online communities, statistical and scientific data, and reviews. This Web of Data enables new types of applications.

P.Jagadeesh Babu

Assistant Professor, Department of Computer Science and Engineering, VEMU Institute of Technology, Pakala, Chittoor Dist.

There are generic Linked Data browsers which allow users to start browsing in one data source and then navigate along links into related data sources. There are Linked Data search engines that crawl the Web of Data by following links between data sources and provide expressive query capabilities over aggregated data, similar to how a local database is queried today. The Web of Data also opens up new possibilities for domain-specific applications. Unlike Web 2.0 mashups which work against a fixed set of data sources, Linked Data applications operate on top of an unbound, global data space. This enables them to deliver more complete answers as new data sources appear on the Web.

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 relevant 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 represents 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 setlevel graph essentially captures a part of the Linked Data schema on the web that is represented in RDFS, i.e., relations between classes. Often, a schema might be incomplete or simply does not exist for RDF data on the web. In such a case, a pseudo schema can be obtained by computing a structural summary such as a data guide.

2. EXISTING SYSTEM:

Existing work can be categorized into two main categories:

Volume No: 2 (2015), Issue No: 6 (June) www.ijmetmr.com



A Peer Reviewed Open Access International Journal

1.schema-based approaches

2.Schema-agnostic approaches

There are schema-based approaches implemented on top of off-the-shelf databases. A keyword query is processed by mapping keywords to elements of the database (called keyword elements). Then, using the schema, valid join sequences are derived, which are then employed to join ("connect") the computed keyword elements to form so called candidate networks representing possible results to the keyword query. Schema-agnostic approaches operate directly on the data. Structured results are computed by exploring the underlying data graph. The goal is to find structures in the data called Steiner trees (Steiner graphs in general), which connect keyword elements. 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 and dynamic programming.



Fig.1: Ste-level Web Data Graph

Existing work on keyword search relies on an elementlevel model (i.e., data graphs) to compute keyword query results.

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. » The routing problem, we need to compute results capturing specific elements at the data level.

» Routing keywords return the entire source which may or may not be the relevant sources.

3. PROPOSED SYSTEM:

We propose to route keywords only to relevant sources to reduce the high cost of processing keyword search queries over all sources. We propose a novel method for computing top-k routing plans based on their potentials to contain results for a given keyword query. We employ a keyword-element relationship summary that compactly represents relationships between keywords and the data elements mentioning them. A multilevel scoring mechanism is proposed for computing the relevance of routing plans based on scores at the level of keywords, data elements, element sets, and subgraphs that connect these elements. We propose to investigate the problem of keyword query routing for keyword search over a large number of structured and Linked Data sources.



Fig.2: Proposed System Architecture. ADVANTAGES OF PROPOSED SYSTEM:

» Routing keywords only to relevant sources can reduce the high cost of searching for structured results that span multiple sources.

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

Volume No: 2 (2015), Issue No: 6 (June) www.ijmetmr.com

June 2015 Page 724



A Peer Reviewed Open Access International Journal

4. IMPLEMENTATION:

The system is proposed to have the following modules along with functional requirements.

Keyword Search
Keyword Query Routing
Multilevel Inter-Relationship
Set - Level

1. Keyword Search:

There are schema-based approaches implemented on top of off-the-shelf databases .A keyword query is processed by mapping keywords to elements of the database (called keyword elements). Then, using the schema, valid join sequences are derived, which are then employed to join ("connect") the computed keyword elements to form so-called candidate networks representing possible results to the keyword query. Schema-agnostic approaches operate directly on the data. Structured results are computed by exploring the underlying data graph. The goal is to find structures in the data called Steiner trees (Steiner graphs in general), which connect keyword elements.

2. Keyword Query Routing:

We propose to investigate the problem of keyword query routing for keyword search over large number of structured and Linked Data sources. Routing keywords only to relevant 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 represents the first attempt to address this problem. A solution to keyword query routing can address these problems by pruning unpromising sources and enabling users to select combinations that more likely contain relevant results. For the routing problem, we do not need to compute results capturing specific elements at the data level, but can focus on the more coarse-grained level of sources.

3. Multilevel Inter-Relationship:

The search space of keyword query routing using a multilevel inter-relationship graph. The inter-relationships between elements at different levels are above Fig. 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. We propose a ranking scheme that deals with relevance at many levels.

4. Set Level:

We extract keywords and relationships from the data. Then, based on the elements and sets of elements in which they occur, we create keyword-element relationships. Precomputing relationships (i.e., paths) between data elements are typically performed for keyword search to improve online performance. These relationships are stored in specialized indexes and retrieved at the time of keyword query processing to accelerate the search for Steiner graphs. For database selection, relationships between keywords are also precomputed. This work neither considers relationships between keywords nor relationships between data elements but between keyword-elements that collectively represent the keywords and the data elements in which they occur.

5.CONCLUSION:

We have presented a solution to the novel problem of keyword query routing. Based on modeling the search space as a multilevel inter-relationship graph, we proposed a summary model that groups keyword and element relationships at the level of sets, and developed a multilevel ranking scheme to incorporate relevance at different dimensions. The experiments showed that the summary model compactly preserves relevant information. In combination with the proposed ranking, valid plans (precision@1 ¼ 0:92) that are highly relevant (mean reciprocal rank ¼ 0:86) could be computed in 1 s on average. Further, we show that when routing is applied to an existing keyword search system to prune sources, substantial performance gain can be achieved.

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.



A Peer Reviewed Open Access International Journal

[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 SIG-MOD Conf., pp. 305-316, 2007.

[13] 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.

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

[15] 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.

[16]G. Ladwig and T. Tran, "Index Structures and Top-K Join Algorithms for Native Keyword Search Databases,"Proc. 20th ACM Int'l Conf. Information and Knowledge Management (CIKM), pp. 1505-1514, 2011.