From: Coward, Jeffery

Sent: 10/31/2013 12:34:49 PM

To: TTAB EFiling

CC:

Subject: U.S. TRADEMARK APPLICATION NO. 85561168 - DEEP WEB INTELLIGENCE - 4335.14US01 - Request for Reconsideration Denied - Return to TTAB - Message 5 of 5

****

Attachment Information:

Count: 7

Files: RDDW2-16.jpg, RDDW2-17.jpg, RDDW3.jpg, RDDW4-1.jpg, RDDW4-2.jpg, RDDW4-3.jpg, RDDW4-4.jpg

http://webcache.googleusercontent.com/search?q=cache:WRDbUx-VNV4J:ftp://ftp.cse.ohio-state.edu/pub/tech-report/2008/TR10.pdf+&cd=7&hl=en&ct=clnk&gl=us 10/29/2013 01:27:41 PM

of related topics, including query planning, keyword search on relational databases, database ranking, deep web mining, and semantic search.

Query Planning: Query planning has been extensively studied in databases. Raschid and co-workers have developed a navigational-based query planning strategy for mining biological data sources [3, 19].They build a source graph representing integrated biological databases and an object graph representing biological objects in each database. The key difference in our work is we focus on deep web data source dependencies, not the physical links between the database objects

Much work on query planning is based on the well known Bucket Algorithm [9, 8, 16, 24, 21, 20, 25]. In this work, it is assumed that the user query specifies the databases or relations that need to be queried. The task of the work is to find a query order among the specified relations or databases. Based on user specified relations or sub-goals, a bucket is built containing all the databases which can answer the corresponding sub-goal. In comparison, in our work, the user query only contains keyword and does not specify any databases or relations of interest. Our system takes advantage of domain ontology and data redundancy to select the best data sources automatically. In other words, our system figures out sub-goals by itself. At the same time, query planning is also performed by the system.

In [27], a query planning algorithm is presented which minimizes the query's total running time by optimally exploits parallelism among web services. The main differences between our work and theirs are: first, they assume that one attribute can only be provided by one data source, and second, they assume all available databases should be used in the query plan, and their focus is ordering them to minimize the query plan running time. We use a domain ontology to rank databases and select the most relevant data sources to mation for each data source. We do not know the actual content of the data sources. The edges in our graph are inserted based on multi-source inter-dependencies, which is a directed relationship. This is different from the foreign key relationship, which is single-source and undirected. The multi sources dependence relationship determines that our graph is a hyper-graph with multiple hyper-nodes. As a result, our search algorithm is adapted to this special feature of the graph. Finally, data redundancy is an important consideration. We use ontology to rank data sources according to their relevance to user query.

Database Ranking: There has been some efforts on database ranking. In [28], the authors rank relational databases' relevance with respect to a set of keywords. The number of joins needs to be made to connect two keywords within one database is used as the measurment. In [15], a search result ranking strategy in the presence of overlapping data sources is addressed. A surprisingness and confidence score are computed for each search result. The results are then ranking by these two scores. In our work, we are considering data redundancy in deep web data sources, not in relational databases. We use domain ontology and similarity measure to select the most relevant databases according to user query.

Deep Web Mining: Lately, there has also been a lot of work on mining useful information from the deep web [11, 5, 4, 12]. However, the above work is generally driven by e-commerce domain applications, focusing on database integration and schema matching. Dependencies between data sources have not been considered by these.

Semantic Search: The semantic issues in keyword search and web services have motivated much research [26, 18, 29]. Two main issues have been addressed by the existing research. The first one is using ontology to determine the type

or scope of a keyword, and the second is using ontology relations to obtain the intention of a query. In our work, we are also considering the semantics issue. A domain ontology is used to decide the type and relation between keywords. But in our case, the keywords involved are technical terms from a specific domain, which are different from some random keywords in a relational text database. Therefore, it is relatively easy to decide the intention underlying a given query.

8. CONCLUSIONS

In this paper, we have considered keyword search over online deep web data sources. Our algorithm is able to support two type of queries, which are keyword-attributes query and

[7] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. Introduction to Algorithms. The MIT Press, second edition, 2001.

[8] AnHai Doan and Alon Halevy. Efficiently ordering query plans for data integration. In Proceedings of the 18th International Conference on Data Engineering, page 393, 2002.

[9] Daniela Florescu, Alon Levy, and Ioana Manolescu. Query optimization in the presence of limited access patterns. In Proceedings of the 1999 ACM SIGMOD international conferenc on Management of Data, pages 311–322, 1999.

[10] Bin He, Mitesh Patel, Zhen Zhang, and Kevin Chen-Chuan Chang. Accessing the deep web: A

http://webcache.googleusercontent.com/search?q=cache:WRDbUx-VNV4J:ftp://ftp.cse.ohio-state.edu/pub/tech-report/2008/TR10.pdf+&cd=7&hl=en&ct=clnk&gl=us 10/29/2013 01:27:41 PM

two type of queries, which are keyword-attributes query and keyword-keyword relation query, and derive interesting relationships between the keywords from the deep web. We use a hyper-graph model to capture the multi-sources inter-dependencies between data sources. Since there is data redundancy across deep web data sources, we have also designed a domain ontology based data source ranking and selection strategy. We have developed a bidirectional query planning algorithm which is based on bidirectional searching algorithm for traditional database keyword search, but addresses several new challenges.

Our query planning algorithm is capable of finding an query answering plan for a keyword query with a high relevance score and a low execution time. We compared our algorithm with the naive algorithm, the exhaustive algorithm, and a backward planning algorithm. We find that our algorithm can generate optimal results for most queries, with much smaller query planning time, and also runs much faster than the backward algorithm.

9. REFERENCES

[1] Sanjay Agrawal, Surajit Chaudhuri, and autam Das. Dbxplore: A system for keyword-based search over relational databases. In Proceedings of the 18th International Conference on Data Engineering, page 5, 2002.

[2] B.Aditya, Gaurav Bhalotia, Soumen Chakrabarti, Arvind Hulgeri, Charuta Nakhe, Parag Parag, and S.Sudarshan. Banks: Browsing and keyword searching in relational databases. In Proceedings of the 28th International Conference on Very Large Data Bases, volume 28, pages 1083–1086, 2002.

[3] Jens Bleiholder, Samir Khuller, Felix Naumann, Louiqa Raschid, and Yao Wu. Query planning in the presence of overlapping sources. In Proceedings of the 10th International Conference on Extending Database Technology, pages 811–828, 2006.

[4] K. Chang, B. He, and Z. Zhang. Toward large scale integration: Building a metaquerier over databases on the web, 2005.

[5] Kevin Chen-Chuan Chang and Junghoo Cho. Accessing the web: From search to integration. In Proceedings of the 2006 ACM SIGMOD international conference on Management of Data, pages 804–805, 2006.

[6] Philipp Cimiano, Gunter Ladwig, and Steffen Staab. Gimme' the context: Context-driven automatic semantic annotation with c-pankow. In Proceedings of the 14th international conference on World Wide Web, pages 332–341, 2005.

Chen-Chuan Chang. Accessing the deep web: A survey. Communications of ACM, 50:94–101, 2007.

[11] Bin He, Zhen Zhang, and Kevin Chen-Chuan Chang. Knocking the door to the deep web: Integrating web query interfaces. In Proceedings of the 2004 ACM SIGMOD international conference on Management of Data, pages 913–914, 2004.

[12] Hai He, Weiyi Meng, Clement Yu, and Zonghuan Wu. Automatic integration of web search interfaces with wise integrator. The international Journal on Very Large Data Bases, 12:256–273, 2004.

[13] Vagelis Hristidis, Luis Gravano, and Yannis Papakonstantinou. Efficient ir-style keyword search over reltional databases. In Proceedings of the 29th international conference on Very Large Data Bases, 2003.

[14] Vagelis Hristidis and Yannis Papakonstantinou. Discover: Keyword search in relational databases. In Proceedings of the 28th international conference on Very Large Data Bases, pages 67–681, 2002.

[15] Philipp Hussels, Silke Tribl, and Ulf Feser. What's new? what's certain? - scoring search results in the presence of overlapping data sources. Data Integration in the Life Sciences, 4544:231–246, 2007.

[16] Azchary G. Ives, Daniela Florescu, Marc Friedman, and Alon Levy. An adaptive query execution system for data integration. In Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data, pages 299–310, 1999.

[17] Varum Kacholia, Shashank Pandit, Soumen Chakrabarti, S.Sudarshan, Rushi Desai, and Hrishikesh Karambelkar. Bidirectional expansion for keyword search on graph databases. In Proceedings of the 31st international conference on Very Large Data Bases, pages 505–516, 2005.

[18] Eser Kandogan, Rajasekar Krishnamurthy, Sriram Raghavan, Shivakumar Vaithyanathan, and Huaiyu Zhu. Avatar semantic search: A database approach to information retrieval. In Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data, pages 790–792, 2006.

[19] Zoe Lacroix, Louiqa Raschid, and Maria-Esther Vidal. Efficient techniques to explore and rank paths in life science data sources. In Proceedings of the 1st International Workshop on Data Integration in the Life Sciences, pages 187–202, 2004.

[20] Ulf Leser and Felix Naumann. Query planning with information quality bounds. In Proceedings of the 4th International Conference on Flexible Query Answering, pages 85–94, 2000.

http://www.academia.edu/253659/From_Deep_Web_up_to_the_Surface_Web 10/29/2013 01:30:25 PM

Academia.edu
share research
Search People, Research Interests and Universities
Home Log In Sig

From Deep Web up to the Surface Web

by Muhammad Maaz Rehan

""Deep Web" is the hidden but distributed repository of data lying in the databases and other documents which is not accessible through search engines. This information is mostly accessible through query interfaces and only the authorized users can query it after login. As a result data access is restricted to the search engines and web crawlers as well. In this paper a new algorithm DWToSW (Deep Web To Surface Web) is proposed that aims to bring Relational Database over the surface web via semi structured XML thus attempting to locate and quantify material on the Web that is hidden from typical search techniques."

More Info: Published in 1st NC-SCC, pp.78-83, May 23-28 2008, KUST Pakistan

16

Download (.pdf)

From_Deep_Web_...
58.4 KB

Loading Preview

Job Board About Press Blog Stories We're hiring engineers! FAQ Journals Terms Privacy Copyright **Send us Feedback**

Academia © 2013

http://webcache.googleusercontent.com/search?q=cache:frJXZFuE6OoJ:iswc2013.semanticweb.org/content/relational-database-rdf+&cd=19&hl=en&ct=clnk&gl=us 10/29/2013 01:32:39 PM

This is Google's cache of http://iswc2013.semanticweb.org/content/relational-database-rdf. It is a snapshot of the page as it appeared on Oct 16, 2013 13:37:11 GMT. The current page could have changed in the meantime. Learn more
Tip: To quickly find your search term on this page, press **Ctrl+F** or **⌘-F** (Mac) and use the find bar.

The 12th International Semantic Web Conference

21-25 October 2013, Sydney, Australia

Relational Database to RDF

Relational Database to RDF (RDB2RDF) Tutorial: The new R2RML and Direct Mapping Standards, from Semantics to Practice

Tutorial webpage

http://rdb2rdf.org/

Presenters

- *Juan F. Sequeda (Department of Computer Science, The University of Texas at Austin)*
- *Barry Norton*
- *Daniel Miranker (Department of Computer Science, The University of Texas at Austin)*

Abstract

http://webcache.googleusercontent.com/search?q=cache:frJXZFuE6OoJ:iswc2013.semanticweb.org/content/relational-database-rdf+&cd=19&hl=en&ct=clnk&gl=us 10/29/2013 01:32:39 PM

Abstract

In one study it was reported that there is 500 times as much data in the hidden or deep web, then there is in crawlable, indexable web pages. Most of this data is stored in relational databases. Hence, the Semantic Web's promise of Web-scale data integration will only live up to its true promise with the inclusion of legacy relational database management systems (RDBMS). Two new, interrelated and complementary W3C standards, "Direct Mapping of Relational Data to RDF" and "R2RML: RDB to RDF Mapping Language", are poised to enable this data into the web of Linked Data. This tutorial will introduce, contrast and explain the preferred solution organizations embodies by these two standards. The tutorial will also include a hands-on session.

Sponsors » Become a Sponsor

Semantic Web Challenge Sponsor

Invited Speakers Sponsor

Platinum

http://webcache.googleusercontent.com/search?q=cache:frJXZFuE6OoJ:iswc2013.semanticweb.org/content/relational-database-rdf+&cd=19&hl=en&ct=clnk&gl=us 10/29/2013 01:32:39 PM

Gold

Silver

Doctoral Consortium Sponsor

http://webcache.googleusercontent.com/search?q=cache:frJXZFuE6OoJ:iswc2013.semanticweb.org/content/relational-database-rdf+&cd=19&hl=en&ct=clnk&gl=us 10/29/2013 01:32:39 PM

Student Travel Award Sponsor

Organiser

Follow @ISWC2013

Tweets by @iswc2013

© 2013 The 12th International Semantic Web Conference (ISWC2013)