Contents lists available at Science-Gate

International Journal of Advanced and Applied Sciences

Journal homepage: http://www.science-gate.com/IJAAS.html

A comprehensive framework for the semantic cache systems

Mohammad Ahmed Alghobiri [1], Hikmat Ullah Khan [2, *], Tahir Afzal Malik [3], Saqib Iqbal [4]

[1]Department of Management Information Systems, King Khalid University, Abha, Saudi Arabia
[2]Department of Computer Science, COMSATS Institute of Information Technology, Wah, Pakistan
[3]Department of Management Information Systems, Ibn Rushd College for Management Sciences, Abha, Saudi Arabia
[4]Department of Software Engineering and Computer Science, College of Engineering and Information Technology, Al Ain University of Science and Technology, Al Ain, United Arab Emirates

ARTICLE INFO

Article history:
Received 5 June 2016
Received in revised form
17 October 2016
Accepted 20 October 2016

Keywords:
Semantic cache
Framework
XML

ABSTRACT

Semantic Cache deals in both semantic descriptions as well as the results of previous queries providing improved efficiency. The semantic cache is an active research domain due to its novel ideology and advantages. The proposed framework is formulated based on the extensive study in the domain of semantic cache in general and in the study of framework for semantic cache in particular. About all the previous works contributed to the domain knowledge, but the certain raised issues still needed to be addressed. For the proposed framework, about all factors involving semantic caching systems are explored. The novelty of the proposed framework lies in the deep modular approach. The proposed framework, described at the detailed level, may present guidance for future works on semantic cache systems.

1. Introduction

The paradigm shift of local systems to network and then the evolution of World Wide Web has changed the shape of computing and standard of resource sharing. The use of databases turned to very large databases and then XML became the de facto standard of data representation, storage and exchange. XML has the capacity to represent all sorts of information and can store in XML documents and then can query those documents that can be queried using advanced query languages like XPath and XQuery. Processing of such XML queries can be time consuming as these involves navigating through the hierarchical structure of XML usually known as an XML tree. These query languages target both the value of the data items in the XML documents as well as the structure of it. An XML query can be taken as set of queries progressively targeting the database. It starts from the root node of the XML tree and then filtering up to certain node selecting on the basis of the structure of the tree as well as matching the predicates. As efficiency is always our major issue thus it also remained the main performance requirement when retrieval of XML data is concerned as a result of a query. This requirement resulted in the research of many other fields, including the query caching techniques. Caching is regarded to play an important role in distributed databases, client-server databases, and web-based information systems due to the fact that slow remote servers and high traffic flow can cause long delays in delivering the answers. The query caching techniques, then result into the current form of semantic cache. For a XML database, semantic cache maintains materialized views which are a combination of evaluated queries as well as the relative result node sets (Feng et al., 2007).

There are a few key issues regarding semantic caches. First, the client maintains the semantic description of the data in cache. This information is used by the query processing to determine which part is already available in the cache and which is needed to be fetched from the server. Second, the information is stored in the form of regions such that the set of related tuples is stored in the same region. So such information is retrieved in a collected manner. Lastly, these semantic regions help in avoiding large storage capacities as these are helped by maintaining proper and necessary information within the semantic cache with the help of replacement algorithms (Dar et al., 1996). There has been a lot of work done in the domain of semantic cache since its introduction, various frameworks

* Corresponding Author.
Email Address: Hikmat.Ullah@ciitwah.edu.pk (H. U. Khan)

https://doi.org/10.21833/ijaas.2016.10.012
2313-626X/© 2016 The Authors. Published by IASE.

have been presented with certain advantages and raising some issues as well.

This paper has been presented in the following layout. After introduction, the related works regarding framework of the semantic caches have been mentioned with the brief introduction. Then the proposed comprehensive framework has been proposed with detailed discussion of modules and also the certain services introduced within those modules. Then the proper labeled figure of the proposed framework is given before concluding the paper. Due to shortage of space, many preliminary concepts and terminologies have not been specified in the introductory part but all such terms and concepts can be found including (Balmin et al., 2004; Brunkhorst and Dhraief, 2005) for better understanding of the domain knowledge and problem definition and overall concepts used in the paper.

2. Related work

A lot of work regarding the concept of Semantic Cache is found. Earlier works were mainly focused on the semantic caching of the relational database systems (Yang et al., 2003; Xu, 2015; Luo and Naughton, 2001; Godfrey and Gryz, 1999). An excellent survey regarding relational databases based semantic cache explores the works with their novelty as well weak points providing guidance for future research guidelines (Lee and Chu, 1999). But with the advent of XML replacing conventional relational databases, the paradigm of research on the domain of semantic Cache has shifted from relational database-backed semantic cache to XML based semantic cache (Amiri et al., 2003; Ahmad et al., 2008; Abiteboul et al., 2006; Balmin et al., 2004).

Let us now briefly introduce the works done regarding the framework of semantic caches. The framework for XML querying by caching frequent query pattern by mining online queries was presented. It introduced four query relationships, i.e., exact query matching, semantic matching, exact containment and semantic containment (Hristidis and Petropolous, 2002). A novel framework for the semantic cache for secure XML query answering was proposed that explored the joint of the semantic caching technique with secure constraints (Feng et al., 2007). Then (Bei et al., 2007) proposed the framework for semantic caching of XML databases. It proposed specialized methods like indexing and broadening of semantic regions for optimal and efficient use of semantic cache. Another framework was introduced for maintaining and using semantic cache of query results which are materialized views used for query processing. It provides a way for efficient cache lookup by using a novel technique of view selection (Hamid and Khan, 2005). Semantic caching architecture for Peer to Peer networks was discussed that mainly emphasized on the size of content for retrieval from the cache, but it does not provide an efficient query matching technique (Mandhani and Suciu, 2005). Another novel framework for the semantic cache system was presented that offered representation system of cached XML data, a new algorithm for semantic query matching, and also to incrementally maintaining the cached XML data (Brunkhorst and Dhraief, 2007). There has been growing concern regarding the efficiency of the overall semantic caching systems and also lesser time for query processing is desirable. A semantic caching architecture favoring faster semantic matching in the overall query processing system was presented (Chen et al., 2002). It enhanced the existing systems and shown the mapping of a hierarchical indexing scheme for the case study to depict the execution of architecture. But the main focus of the proposed architecture was for heterogeneous and distributed multi-databases like data warehouses and grids. Another framework can be found in (Liang and Feng, 2010) proposed system for efficient caching of the broad class of XML queries targeting the XML databases. It organizes the cached data in the form of modification of the incomplete tree, which depicts properties of incremental maintenance, containment decidability and remainder generation in an efficient manner.

We find an effective scheme (Khan and Malik, 2012) for querying metadata that takes into account the target resources in forward, backward as well as in the middle of the RDQL query to answer various types of RDQL queries, in particular, the ones based on the sub-query concepts. Another low-cose and efficient processing algorithm (Bao et al., 2015) is found based on novel features of distinguish-ability and the type of target node to detect and solve the “MisMatch” problem using keyword based search in XML. This portable and lightweight algorithm also generates useful suggestions and explanations to users. Woensel and Casteleyn (2016) proposed a query service that dynamically identifies caches and retrieves the semantic contents relevant to the query. The service supports integrated and dynamic querying of the semantic data, for instance, RDF files or embedded semantics in the annotated web pages. The modern distributed systems for query processing need to balance the query loads and leverage the caches results to maximize the system throughput. The survey (Qureshi et al., 2013) provides valuable insights in learning various types of semantic search modes and the behavior of existing semantic search engines. Another work (Eom et al., 2015) proposed scheduling policies in the context of distributed query processing. The policies are based on the dynamic details of caching infrastructure for distributed systems and apply the statistical prediction techniques into scheduling policies for query processing. Using a single tool to access and retrieve geospatial data on the web would a relief for the researchers. However, the challenges are the diverse data sources and the heterogeneous formats. The authors in (Tian and Huang, 2012) integrated the specifications and standards of Open Geospatial Consortium (OGC) and Universal Description, Discovery and Integration

(UDDI) respectively to tackle these challenges. Semantic search has proved to be the next generation of search paradigm. One of the scalable, open-source frameworks "Mímir" proposed. (Tablan et al., 2015) supports complex queries for semantic search and knowledge inference from formal text documents, linguistic annotations and document structures; concept-based search algorithm (Sah and Wade, 2016) for data web where query results relating to a concept are categorized. The results are personalized based on when user explores a particular concept. On the other hand, the authors (Ristoski and Paulheim, 2016) investigated various approaches based on integrating semantic web with knowledge discovery and the data mining processes to show the importance of linked open data for the recommender systems. A method based on many knowledge resources (KRs) for semantic search and annotation in target corpus (Berlanga et al., 2015). The method depends on a statistical framework where corpus documents and the KR concepts are uniformly represented using models of statistical language. Xu et al. (2015) proposed a semantic method to design XML views and the relational data using conceptual models. The method transforms the data of one model into another via transformation rules derived from the conceptual models. Peng et al. (2016) proposed an algorithm for the processing of SPARQL queries in distributed systems for large-scale RDF graphs. The algorithm has two major functions; (1) partial evaluation in the form of local partial and (2) centralized as well as distributed assembly of the partial answers. The allegorical nature and unique style of the Holy Quran makes it difficult to search using keywords. The authors in (Khan et al., 2013) applied SPARQL Queries and the ontology based concepts of th semantic web to analyze the role and importance of ontology in the context of semantic web. MIIB (Khan et al., 2015) is an effective framework to identify the top influential bloggers using the properties of bloggers and the blogging networks. MIIB is based on blogger popularity and productivity in a blogging community as well as the importance of the blogging community.

3. Proposed framework

The proposed model has been designed at a sub-modular level, that is the proposed framework has been designed one level into specific modules, then the modules have been decomposed into independent services to accomplish dedicated tasks.

3.1. Query management module

The importance of the query composition in modern techniques of information retrieval and search engines is well known. The semantic search engines have adopted the concept of advanced options as well as use of wizard for complex query formulation (Bashir et al., 2007; Hristidis and Petropoulos, 2002); semantic caching of XML databases. Three different ways for query input from user are proposed. For basic users, simple text input is easy and straight forward option, while for advanced level user, there may list of options and preferences for complex query formulation. These options may include certain conditions. For expert level user, the option at that level can even the choice of selecting of Node, Element, attribute and predicate can also is taken as input. This will help him to be precise and thus specific result extraction from the semantic cache can be ensured.

3.2. Query

This is module to better manage the poor form of semantic cache query. As the level of query id diverse thus semantic query optimizer can optimize the query with certain alterations for preprocessing options. Semantic query merging service can merge the different options of complex query into one query and also different options of expert level query checks the linkage of various options of an element, attribute, predicate etc. with one another. The semantic Query parsing service helps in managing the query overall.

3.3. Query decomposition module

As the xml based structure consists of a variety of ingredients, thus decomposition of the query into sub parts level so that matching and searching can be attained in a better way. Semantic schema pattern evaluator matches the query to the xml data found in the semantic cache, thus this helps in defining the generation of probe query and remainder query, two separate parts generated from a client query. Probe query is the one which extracts the related part of the result set already present in the local cache while remainder Query is the one that fetches the missing part or in other words non-cached part (Majid et al., 2013). It is proposed to introduce the concept of finding overlapping evaluator modules that can find the portion of xml tree that looks for the part of xml tree that looks for the part of xml tree that has already been searched and thus that portion can help us making short to the probe query and thus already existing data in Xml semantic cache can be fetched directly.

3.4. The semantic cache

The semantic cache is the core module that contains the xml based semantic cache, as well the xml cache query engine and the essential semantic cache management part. The XML based semantic cache contains the cached data in XML format. XML cache query engine interacts with semantic cache management module within the semantic cache. In addition, it has the role of interacting with the query decomposer regarding probe query and also it deals with the cache-memory transfer management service as explained later. The semantic cache management module is a collection to three more modules named as the Replacement manager, Index/Semantic Index manager and Semantic Region Manager. The Replacement manager has the role of replacement of xml data from semantic cache to memory. Various replacement strategies are used, including least recently used, most recently used, etc. In conventional systems, the content is usually based on the concepts of temporal locality or spatial locality. Temporal locality is the characteristic that items of the content have recently been referenced and there is a probability that these items will be used in near future queries. So, as evident from the concept that LRU uses temporal locality. Spatial locality is the characteristic that is based on tuples clustering concepts where the related tuples to pages are clustered in static form. So if an item has been referenced then it is supposed that the next near future queries will be from the items of the tuples of the same cluster; related from the items of the cluster. It is proposed that such an algorithm should be based on the Semantic Locality. Semantic locality is modified form of spatial locality as it is dynamic in nature as it adapts itself with respect to pattern of queries. Indeed, the use of dynamic in nature of semantic locality is more efficient than static spatial locality or traditional temporal locality. The use of partial replacement instead of total replacement has also proven to be important phenomenon. In this technique, un-important XML fragments are replaced while retaining useful XML fragments within the semantic cache (Nakatoh et al., 2003). It is proposed that Index for xml based semantic can be maintained not in simple tree, but also using numeric based tree computation or suffix tree as the work (Chidlovskii and M. Borghoff, 2000). Semantic Regions are like pages which are associated with set of tuples. Semantic Region provides a means for the cache manager to collect information about multiple tuples. It is proposed here that aggregated information should be maintained in the semantic region. The Semantic region may consist of two parts, semantic region metadata and content part. Semantic region metadata may consist of the size of the semantic region as it may be dynamic in size, count of the use of the content of the semantic region to predict the probability that it may be used in near future queries and also maintaining the time of content access as it will be helpful in determining the least recent usage and most recent usage. The semantic region metadata is used to detect the regions relevant to the query while semantic region content part contains the content that will be accessed to retrieve answer tuples. The Semantic Distance should be maintained using ranking wise. It is also recommended that concept of dynamic region be initiated, i.e., the region size can be adjusted according to the size collection of tuples resulted from user queries. The Dynamic Region Size Manager should be specified for this important task. The result of the probe query in the semantic cache, in raw format, is sent to the result refinement module.

3.5. Cache-memory management module

The cache-memory management service consists of three modules, schema extraction service, consistency maintenance service and cache-memory mapping service. Schema extraction service deals in the extraction of xml tree schema and then cache-memory mapping service is used to map them so that the consistency maintenance manager maintains the consistency between the xml data found in the semantic cache with respect to memory. To accomplish such tasks, the module has to be in a consistent link to the both the semantic cache as well as memory as shown in Fig. 1.

3.6. XML data store

The xml data store is present in the primary memory. The xml data store is managed by the xml data query engine. The result of the remainder query, produced by the query decomposer, is sent to the result refinement module in the raw format.

3.7. Results refinement module

This is the final module that deals in the refinement of result in a proper manner. Result integration service deals in the integration of the results from the remainder and probe queries. This is important since user should get the complete result. Result ranking service can alter the sequence of the results so that the result may be arranged in required a proper sequence.

It is further proposed that the result may be arranged in required sequence as that of query wizard.

4. Conclusion

In this paper, we have presented the comprehensive framework for semantic cache. The framework has been formulated in such a way to depict how the query is processed in the semantic cache based systems. The framework has been divided into several necessary modules, which have further been divided into services.

It is notable that the proposed framework is a novel work; all the services within each module must be followed because in doing so, we will gain the level of effectiveness, but that system will be taking a lot of time in query processing thus causing lesser efficiency.

The proposed framework is not compared with existing framework for semantic cache systems, because it is based on the exploratory study of the domain of semantic cache based systems. The proposed framework will be considered as a guideline for future work in the domain of semantic cache.

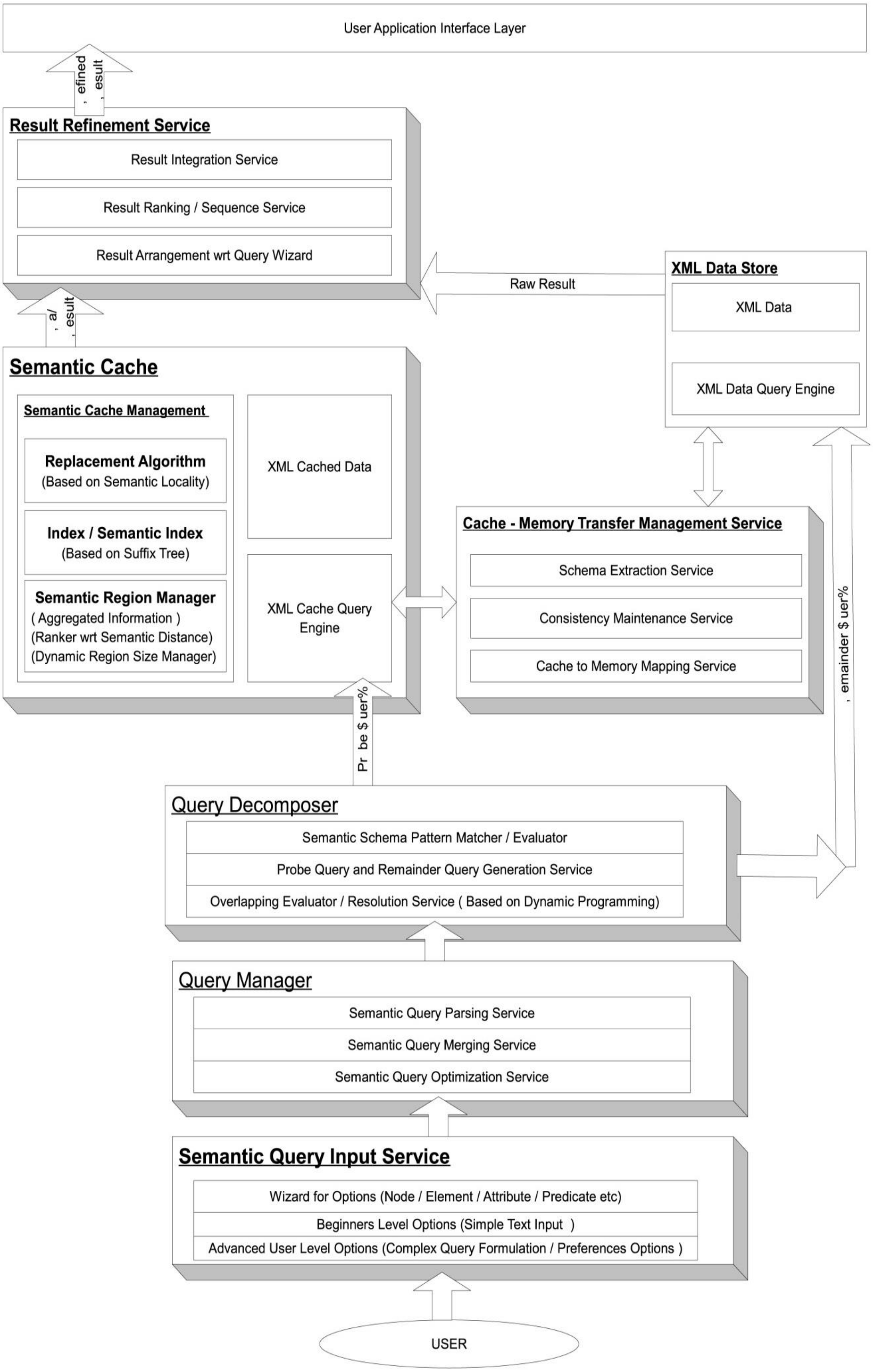
User Application Interface Layer
, efined , esult
Result Refinement Service
Result Integration Service
Result Ranking / Sequence Service
Result Arrangement wrt Query Wizard
Raw Result
XML Data Store
XML Data
XML Data Query Engine
, al/ , esult
Semantic Cache
Semantic Cache Management
Replacement Algorithm
(Based on Semantic Locality)
Index / Semantic Index
(Based on Suffix Tree)
Semantic Region Manager
(Aggregated Information)
(Ranker wrt Semantic Distance)
(Dynamic Region Size Manager)
XML Cached Data
XML Cache Query Engine
Cache - Memory Transfer Management Service
Schema Extraction Service
Consistency Maintenance Service
Cache to Memory Mapping Service
, emainder $ uer%
Pr be $ uer%
Query Decomposer
Semantic Schema Pattern Matcher / Evaluator
Probe Query and Remainder Query Generation Service
Overlapping Evaluator / Resolution Service (Based on Dynamic Programming)
Query Manager
Semantic Query Parsing Service
Semantic Query Merging Service
Semantic Query Optimization Service
Semantic Query Input Service
Wizard for Options (Node / Element / Attribute / Predicate etc)
Beginners Level Options (Simple Text Input)
Advanced User Level Options (Complex Query Formulation / Preferences Options)
USER

Fig. 1: Proposed framework for semantic cache systems

Acknowledgments

Authors are thankful to Prof. Dr. Abdul Qadir, for his comments. Dr. Abdul Qadir, an eminent professor and expert in the field of semantic web and semantic cache, is Head of the Centre for Distributed and Semantic Computing and Dean, Faculty of Computing, Capital University of Science and Technology, Islamabad.

References

Abiteboul S, Segoufin L and Vianu V (2006). Representing and querying XML with incomplete information. ACM Transactions on Database Systems (TODS), 31(1): 208-254.

Ahmad M, Qadir MA and Sanaullah M (2008). Query processing over relational databases with semantic cache: A survey. In the IEEE International Multitopic Conference (INMIC), Karachi, Pakistan: 558-564. https://doi.org/ 10.1109/INMIC.2008.4777801

Amiri K, Park S, Tewari R and Padmanabhan S (2003). Scalable template-based query containment checking for web semantic caches. In the Proceedings of 19th IEEE International conference on Data Engineering: 493-504. https://doi.org/ 10.1109/ICDE.2003.1260816

Balmin A, Özcan F, Beyer KS, Cochrane RJ and Pirahesh H (2004). A framework for using materialized XPath views in XML query processing. In the Proceedings of the Thirtieth international conference on Very large data bases (VLDB '04), Toronto, Canada, 30: 60-71.

Bao Z, Zeng Y, Ling TW, Zhang D, Li G and Jagadish HV (2015). A general framework to resolve the MisMatch problem in XML keyword search. The VLDB Journal, 24(4): 493-518.

Bashir MF, Zaheer RA, Shams ZM and Qadir MA (2007). SCAM: Semantic caching architecture for efficient content matching over data grid. In Advances in Intelligent Web Mastering, 43(1): 41-46

Bei Y, Chen G, Hu T and Dong J (2007). A caching system for XML queries using frequent query patterns. In the 11th IEEE Conference on Computer Supported Cooperative Work in Design (CSCWD), Melbourne, Australia: 47-52. https://doi.org/ 10.1109/CSCWD.2007.4281408

Berlanga R, Nebot V and Pérez M (2015). Tailored semantic annotation for semantic search. Web Semantics: Science, Services and Agents on the World Wide Web, 30: 69-81.

Brunkhorst I and Dhraief H (2007). Semantic caching in schema-based p2p-networks. In the Proceedings of International Conference on Databases, information Systems, and Peer-to-Peer Computing, Berlin, Germany: 179-186.

Chen L, Wang S, Cash E, Ryder B, Hobbs I and Rundensteiner EA (2002). A fine-grained replacement strategy for XML query cache. In the Proceedings of the 4th International Workshop on Web Information and Data Management (WIDM '02), McLean, USA: 76-83. https://doi.org/10.1145/584931.584947

Chidlovskii B and Borghoff UM (2000). Semantic caching of Web queries. The International Journal on Very Large Data Bases, 9(1): 2-17.

Dar S, Franklin MJ, Jonsson BT, Srivastava D and Tan M (1996). Semantic data caching and replacement. In the Proceedings of the 22th International Conference on Very Large Data Bases (VLDB '96), San Francisco, USA: 330-341.

Eom Y, Hwang D, Lee J, Moon J, Shin M and Nam B (2015). EM-KDE: A locality-aware job scheduling policy with distributed semantic caches. Journal of Parallel and Distributed Computing, 83: 119-132.

Feng J, Ta N, Li G, Liu Y and Lv D (2007). A framework of semantic cache for secure XML query answering: an interesting joint and novel perspective. In the Proceedings of 2nd International Conference on Scalable Information Systems (InfoScale '07), Suzhou, China.

Godfrey P and Gryz J (1999). Answering queries by semantic caches. In the Proceedings of 10th International Conference on Database and Expert Systems Applications, London, United Kingdom: 485-498. https://doi.org/10.1007/3-540-48309-8_45

Hamid A and Khan S (2005). Optimization of semantic caching for XML database. In the IEEE First International Conference on Information and Communication Technologies (ICICT '05), Karachi, Pakistan: 201-205. https://doi.org/10.1109/ICICT.2005.1598584

Hristidis V and Petropoulos M (2002). Semantic caching of XML databases. In the 5th International Workshop on the Web and Databases (WebDB), Madison, USA: 25-30.

Khan HU and Malik TA (2012). Finding resources from middle of RDF graph and at Sub-Query level in suffix array based RDF indexing using RDQL queries. International Journal of Computer Theory and Engineering, 4(3): 369-372.

Khan HU, Daud A and Malik TA (2015). MIIB: A Metric to identify top influential bloggers in a community. PloS one, 10(9): 1-15.

Khan HU, Saqlain SM, Shoaib M and Sher M (2013). Ontology based semantic search in Holy Quran. International Journal of Future Computer and Communication, 2(6): 570-575.

Lee D and Chu WW (1999). Semantic caching via query matching for web sources. In the Proceedings of the eighth international conference on Information and knowledge management (CIKM '99), Kansas City, Missouri, USA: 77-85. https://doi.org/10.1145/319950.319960.

Liang G and Feng JH (2010). An effective semantic cache for exploiting XPath Query/View answerability. Journal of Computer Science and Technology, 25(2): 347-361.

Luo Q and Naughton JF (2001). Form-based proxy caching for database-backed web sites. In the Proceedings of the 27th International Conference on Very Large Data Bases Journal (VLDB '01), San Fransisco, CA, USA: 191-200.

Mandhani B and Suciu D (2005). Query caching and view selection for XML databases. In the Proceedings on the 31st International Conference on Very Large Data Bases (VLDB '05), Trondheim, Norway: 469-480.

Nakatoh T, Ohmori K, Yamada Y and Hirokawa S (2003). Complex query and metadata. In the Proceedings of International symposium on Information Science and electrical engineering (ISEE2003), Fukuoka, Japan: 291-294.

Peng P, Zou L, Özsu MT, Chen L and Zhao D (2016). Processing SPARQL queries over distributed RDF graphs. The International Journal on Very Large Data Bases, 25(2): 243-268.

Qureshi MM, Asma B and Khan HU (2013). Comparative analysis of semantic search engines based on requirement space pyramid. International Journal of Future Computer and Communication, 2(6): 562-566.

Ristoski P and Paulheim H (2016). Semantic Web in data mining and knowledge discovery: A comprehensive survey. Web Semantics: Science, Services and Agents on the World Wide Web, 36: 1-22.

Sah M and Wade V (2016). Personalized concept-based search on the linked open data. Web Semantics: Science, Services and Agents on the World Wide Web, 36: 32-57.

Tablan V, Bontcheva K, Roberts I and Cunningham H (2015). Mímir: An open-source semantic search framework for interactive information seeking and discovery. Web Semantics: Science, Services and Agents on the World Wide Web, 30: 52-68.

Tian Y and Huang M (2012). Enhance discovery and retrieval of geospatial data using SOA and Semantic Web technologies. Expert Systems with Applications, 39(16): 12522-12535.

Woensel WV and Casteleyn S (2016). A mobile query service for integrated access to large numbers of online semantic web data sources. Journal of Web Semantics: Science, Services and Agents on the World Wide Web, 36: 58-76.

Xu F, Li Y and Gu J (2015). Semantic cache replacement strategy for XML algebra-based query optimization. Wuhan University Journal of Natural Sciences, 20(2): 165-172.

Yang LH, Lee ML and Hsu W (2003). Efficient mining of XML query patterns for caching. In the Proceedings of the 29th international conference on Very large data bases (VLDB '03), Berlin, Germany, 29: 69-80.