Survey on Semantic Caching and Query Processing in Databases.

P.Mohan Kumar, T.K.Das, DR.J.Vaideeswaran.

School of Information Technology and Engineering.

Vellore Institute of Technology, Vellore Tamilnadu.

pmohankumar@vit.ac.in , jvaideeswaran@vit.ac.in

Abstract.- **Data caching mechanism for query processing in database turned the researchers due to its performance such as minimum access of latency time, reduced network traffic ,pre fetched data etc, As extension with semantic data caching further made interest among not only the database researchers but also the design and developers even at application level ,due to the existence of data along with its description, pre-processed result which drastically improved the processing speed and user convenience, irrespective of the area such as semantic cache based query processing over peer to peer network, web environment, parallel distributed database,artigficial intelligence,fuzzysets ,mining and cloud. In this paper with this significance a survey is made based on efficiency of query processing as well as pros and cons of existing caching techniques and presented as amending assistance for the database people for enhancement of real world application.**

Keywords: **semantic caching, query processing**

I.   Introduction.

Semantic caching has become familiar and popular due its efficient utilization of all resources that are involved in process of query and the improved response time. Query processing cost depends on the network latencies which can be reduced by efficient semantic cache. The utmost efficiency relies on the management of cache. (i.e: how the query related semantic can be stored and provide space for current query) and query content matching(ie query processing –user query exactly matches with the cache content, if not how to decompose query into partial match ie remainder and probe and process) In this paper the survey focus on the cache management and query processing,

II. Survey of existing techniques.

Qun Ren and Margret Dhunam[1] papers discuss the complete vision of semantic caching .This paper compares the data caching of client server model with semantic caching and explains about the through details of semantic cache based query processing, further paper gives a formal definition for a semantic caching and a detailed view of query processing techniques and experimented with sample and a measure is given how the semantic caching based query process is flexible, low overhead and efficient with respect to normal data caching mechanisms. But this paper limits to the query operations with select, project and equijoins, not discussed with nested queries or large database or network level.

Sangwon kang,Jongwon kim[2] discuss the cache strategies for semantic pre-fetching by maintaining the query results in the cache before a user request a query and an algorithm preference priority replacement is designed and tested. Location dependent and nearest neighbor data concept is tested with range queries. This paper focused on the semantic pre fetching mechanism with cache improves data access efficiency and cache validity in location aware computing. Various cache replacement algorithms were discussed and a model define to process the location dependent queries with respect to semantic pre fetching approach and analyzed the data access efficiency and cache availability.

B. Zheng, D. L. Lee,[3] have proposed the proactive semantic caching algorithm for query processing. The algorithm does not incorporate efficient processing over semantic caches the paper provides the analysis of proposed technique with the help of cost model. This

model is also incomplete due to the fact that the paper does not consider faster semantic matching as a part of processing. That is why this model does not include the cost of query matching process over semantic caching.

Parke Godfrey [4] discusses the general logical frame work for semantic caches for relational operators based queries. This paper focus on the analysis of semantic independence, semantic overlap, semaintic partial and semantic remainder with respect to user queries and evaluation estimated .No model for semantic cache is defined and no novelty is presented.

Jerome Euzeanat[5] report gives a through detail about the research challenges and perspectives of semantic web here the issues focused on both the present and future technologies for user processing through semantics. Various researchers and experts shared their ideas and a way is provided for research issues on semantic cache based query processing in database applications.

S.Prabha, A.Kannan[6] A database optimization framework that supports data and computation reuse, query scheduling and caching mechanism to speed up the evaluation of multi query workload is described. Performance analysis is made with XML cache and database. Semantic approach not discussed.

Sumalatha A.Vaidhei [7] An analysis is made how the semantic technology can help to improve the efficiency of XML query processing in web environment. Semantic caching systems exploit the idea of reusing cached query results to answer new queries based on the query containment and rewriting techniques is discussed in detail.

Min wang, Haixun wang[8] A query-by-example approach that enables to support semantic queries in relational databases with ease are outlined. The inherent fuzziness in the semantics of the query is discussed. The difficulty in expressing queries against graph structured ontology is discussed.

Bryan Genet and Annika Henet[9] discussed about the open issues in optimization of semantic queries and proposed a semantic equivalences .Identified impediments to the implementation of structured query language optimizer in RDBMS,

Significant changes in computer hardware were identified and current SQL optimizers performance over semantic reasoning is analyzed ,schema constraints over RDBMS and constraints query at a time execution to be used were defined and implemented but not on cache mechanism either replacement or storage performance.

Muhammad Azeem Abbas, Muhammad Abdul Qadir[10] here the efficiency of semantic query processing is highlighted by analyzing the ignorance and inconvenience that exists in maximum data retrieval to incoming query and propose an approach of implicit knowledge identification and representation (Cross Attribute Knowledge) present in semantics of the cache. So that maximum possible or even complete answer can be found.

K.U Sattler[11] The algorithm presented in this architecture is naive in sense that is only emphasis is on the amount of data to be retrieved from cache. The searching of query over caching is not handled in the algorithm. Here the scheme of semantic caching for heterogeneous web sources is concentrated more.

Z. Erlangung, Des Akademischen[12] Ontology-based Semantic Query Processing in Database Systems a set of semantic rules for transforming queries using terms derived from the ontology. Different mappings that relate concepts of ontology with those of an underlying database and develop a set of algorithms that allow us to find these mappings in a semi-automatic way. Basically concentrated on how the ontology gets semantically mapped with real world data related to the incoming user query.

Jekkin D.Shah[13] discussed his research on semantic caching and query processing in mobile environments. This work is contributed as system prototype- proof of concept capturing essential elements of semantic caching in mobile environments, cache replacement algorithms, techniques and strategies for semantic caching and query processing. Answering query from cache, addressing the query containment problems was analyzed and algorithms were designed and implemented.

Sha Guo,Wei sun,Xauin,HeSun [14] discussed the satisfiability and implication issues for checking semantic cache to find whether the answer to query is available in cache. Further

UACEE International Journal of Advances in Computer Science and its Applications – IJCSIA
Volume 3 : Issue 2   **[ISSN 2250 – 3765]**

Publication Date : 05 June 2013

analyzed the inequalities among the query predicates used for matching.

Xiaolei Qian [15] Discussed about query rewriting i.e.: determining if and how the user query can be answered from the given resources. Polynomial time algorithm is designed to test the availability of full or partial rewriting of acyclic conjunctive queries

Sheldon Finkelstein [16] here a query graph model is designed in order to optimize the process. Based on this model an inter-query optimization is achieved. it doesn't satisfy the complete solution.

Ren and Dhunam [17] here a dynamic replacement d-LRU is used. Semantic region grow or shrink dynamically, each of them is tie stamped according to its most recent access or modification. The replacement strategy works by removing the older time stamps since each new coming region is assigned a time stamp.

Shaul Dar [18] Suggest a replacement based on the semantic locality. Here the distance is calculated from the cache semantic region to most recently accessed region. The region with largest distance is replaced for new data .Every time replacement arise the distance must be calculated .But this methodology is not much efficient.

Muhamamd Farhan Bashir, Muhammad Abdul Qadir [19] This paper provides the precise definition of efficiency of semantic cache and introduces the neglected area in semantic cache. Survey of existing techniques creates solid grounds for us to conclude that the faster query matching is being neglected is presented. Further mapping of existing techniques against the benchmark is represented.

M.R. Sumalathal, V. Vaidehi[20] given a new strategy, dynamic hash mapping technique which gives fast information retrieval semantically with the cache. A new mapping technique is proposed, which increases the efficiency in information retrieval in the semantic caching system. An analytical solution is given gives enhanced performance by reduced network traffic and searches the exact information from a large database, which is crucial, in a range of applications, especially in network-constrained environments.

H.Kitawaga [22] here the work concentrated on semantic caching technique with linear constraints to implement faster query matching.

The authors concluded this algorithm can be improved by using indexing techniques but has not proposed any indexing technique.

Chen [23] has presented semantic caching architecture for XML queries. Here an overview of semantic cache components. A running prototype is provided for the proposed architecture .Further the query containment mapping technique is presented. But the efficiency is not analyzed.

H. Hopfner, K. U.Sattler[24] presented the caching architecture for mobile DBS. The architecture this paper presents lacks the faster query processing over semantic cache. This paper concentrates on the retrieval of large amount of data for cache. No optimization or query tuning or cache replacement algorithm is discussed.

Heiner Stuckenschmidt [25] in this paper the foundations of an approach for semantic caching of RDF queries is discussed. Here the actual information is taken into account. In particular the same queries with different similarity based on the underlying information thereby ensuring that an optimal result related to the cost model.

Laurent d Orazio [26] works a caching approach for improving query evaluation and propose solution for semantic oriented approach and combines query and object caching. Efficiency is analyzed over grid environment. The dual cache is based on the cooperation of a query cache and an object cache is deployed.

N.Ashish, C.A knblock, C.Shahabi[27] Described a semantic caching for optimizing the performance of information mediators. The relaxed merge (RM) algorithm is designed which used to determine what class of information are useful to cache as auxiliary data source. Here an enumerated data table is used to determine the cached data which seem to efficient comparatively further it reduces the query processing time considerably.

III. Discussions.

The goal of this survey is to the query processing via semantic caching mechanism in order make understand a user convenience. In this paper an analysis is made from semantic definition to real world applications, First of why semantic based approach, how it can be defined,

UACEE International Journal of Advances in Computer Science and its Applications – IJCSIA
Volume 3 : Issue 2   **[ISSN 2250 – 3765]**

Publication Date : 05 June 2013

mapped and processed. Relatively various algorithms in order to improve the query performance proof with respect to various experimental analyses .The survey brings us the detail information of how the efficient mapping of semantic cache content is applicable to various domains such as constraints database ,web environments ,caching in grid level, peer to peer ,distributed and mobile databases .Further this also the gives various approaches on faster retrieval approach and maximum data retrieval details as well how they are given significance and ignored .The survey exposes the detail of cache issues with respect it storage such as replacement policies ,and dual cache for larger data and faster processing and its management ,that is how the collaboration arise among the dual cache. This also gives details about how cache data is organized on behalf of indexing and dynamic hashing mechanism during runtime. The details of how the previous processed query results are stored and reused for new query processing .The survey gives details of how the user query is semantically processed in case if its not complete a algorithmic details is discussed. The details of ontology based semantic processing also discussed and the comments on the challenges and research issues of various researches were mentioned. The use of semantic cache in terms of different domains is the main focus of the researches. Here as, there is still a need to improve algorithms presented for the semantic caching paradigm to reduce the time and cost of result retrieval from caches. Hence this survey reveals that efficient query matching is one of the neglected areas in semantic cache. Yet there is lot of issue such as fast retrieval, exact matching, and amending approach to full fill incomplete queries filtering exact from similarity as the underlying research issues related to semantic query processing in database.

Conclusion.

This paper gives the precise definition of efficiency of Data caching and how it extended as semantic cache and introduces formal level to the real world application level. Survey of existing techniques creates basic grounds to conclude that the quicker query matching method is yet to be improved, not only that the survey states that various issues such as exact matching ,amending approach to full fill incomplete queries filtering exact from similarity as the underlying research issues related to

semantic query processing in data base. In the above point of survey the view is most of research is being done only to use the semantic caching model to improve the deficiencies of other domain, whereas the improvement not in semantic caching model Improvement of response time of semantic cache is not considered by the researchers. If the issues over come most of the applications can achieve an optimistic solution. With hope the conclusion is, this paper gives an idea and suggestions for most of the semantic cache related database researches and in future the good model may be designed and developed.

References.

[1]. Q.Ren and Margaret H. Dhunam Semantic caching and query processing Technical report 1998.

[2]. Sang-Won Kang, Jongwan Kim, SeokJin Im, HaRim Jung, and Chong-Sun Hwang "Cache Strategies for Semantic Pre-fetching Data "Proceedings of the Seventh International Conference on Web-Age Information Management Workshops (WAIMW'06) IEEE 2006.

[3]. B. Zheng, J. Xu, D. L. Lee, "Cache Invalidation and Replacement Strategies for Location Dependent Data in Mobile Environments", in proceedings of IEEE Transactions on Computers, Vol. 21, No. 10,October 2002.

[4]. Parke Godfery and Jarek Gryz. "Answering queries by semantic caches" In Database and Expert system applications, Page 485-498, 1999.

[5]. Jérôme Euzenat , Jean-Eric Pin and Remi Ronchaud "Research challenges and respectives of the Semantic Web "Report of the EU-NSF strategic workshop held at Sophia-Antipolis, France, October 3rd-5th, 2001.

[6]. .S.Prabha, A.Kannan P.Anandakumar "An optimizing query processing with an efficient caching mechanism for Distributed databases". The International Arab Journal of International Technology,Vol 3 No3 July 2006.

[7]. M.R.Sumalatha, V.Vaidehi A.Kannan "Xml Query Processing – Semantic

UACEE International Journal of Advances in Computer Science and its Applications – IJCSIA
Volume 3 : Issue 2   **[ISSN 2250 – 3765]**

Publication Date : 05 June 2013

Cache System" IJCSNS , VOL.7 No.4, April 2007.

[8]. Lipyeow Lim Haixun Wang Min Wang IBM T. J. Watson Semantic queries in Databases: Problems and Challenges *CIKM'09,* November 2–6, 2009, Hong Kong, China. Copyright 2009 ACM.

[9]. Bryan Genet, Annika Hinze Open Issues in Semantic Query Optimization in relational DBMS Technical report 2001.

[10].   Muhammad Azeem Abbas, Muhammad Abdul Qadir "Cross Attribute Knowledge: A Missing Concept in Semantic Cache Query Processing" IEEE proceedings 2009 .

[11].   K. U. Sattler, I. Geist, E. Schallehn, "Concept based Querying in Mediator Systems", in proceedings of The VLDB Journal 14, 1 (Mar.2005), 97- 111.

[12].   Zur   Erlangung   des akademischen   Ontology-based Semantic Query Processing inDatabase Systems DISSERTATION A technical Report Humboldt-University zu Berlin 2006 .

[13]. Jekkin D.Shah University of Maryland Semantic   query caching in mobile environments, [2005].

[14]. Sha Guo,Wei sun,Xauin,HeSun and Mark A.Weiss Solving satisfiability and implication problems in database systems. In ACM transaction Database system 21(2)270-93-1996.

[15]. Xiaolei Qian Query folding.In Stanley Y.SU   editor   12th   International Conference on Data engineering pages 45-55 New Orleans Louisiana1996.

[16]   Sheldon   Finkelstein.   Common expression   analysis   in   database applications In ACM-Sigmoid, Pages 235-245, 1982.

[17]. Q.Ren and Margaret H. Dhunam Using   clustering   for   effective management of a semantic cache in mobile environment. In MobiDE Pages 94-101,1999.

[18]. Shaul Dar, ,Michel J. Franklin,Bjorn T.Jonsson,Michel Tan semantic data

caching and replacement pages 330 - 341 1996.

[19].   Muhamamd   Farhan   Bashir, Muhammad Abdul Qadir Survey on Efficient Semantic Content Matching over Semantic Cache 2006 IEEE.

[20] M.R. Sumalatha V. Vaidehi, A. Kannan,   M.   Rajasekar   and   M. Karthigaiselvan   Hash   Mapping Strategy   for   Improving   Retrieval Effectiveness   in   Semantic   Cache System IEEE - ICSCN 2007, Feb. 22-24,,Pages .233-237, 2007.

[21]. Qun Ren, Margaret H. Dunham, and Vijay Kumar, Semantic Caching and Query Processing IEEE Transactions on knowledge and data engineering. Vol 15,No 1, January- 2003

[22]   Y. Ishikawa, H. Kitagawa, "A Semantic Caching Method Based on Linear Constraints", in proceedings of International Symposium on Database Applications   in   Non-Traditional Environments, IEEE, 1999.

[23]. L.Chen, E. A. Rundersteiner, & S. Wang, "XCache - A Semantic Caching System for XML Queries", ACM SIGMOD, Madison Wisconsin, June 4-6, 2002.

[24]   H. Hopfner, K. U. Sattler, "Towards Trie-Based Query Caching in Mobile DBS", In proceedings of the Workshop Scalability,   Persistence,Transactions-Database Mechanisms for Mobile Applications, Lecture Notes in Infor-matics (LNI), pages 106-121,2003.

[25]   Schmidt   Similarity-Based   Query Caching Vrije University   Amsterdam Lecture Notes.

[26]. Laurent d Orazio Claudia Roncancio Cyril Labbé   Fabrice Jouanot " Semantic   caching   in   large   scale querying   systems".   Institute   of National Polytechnique de Grenoble .

[27] N.Ashish,C.A.Knoblock   ,C.Shahabi Intelligent caching for information mediators –A KR based approach proceedings of 5th KRDB workshop Seattle WA 31st May 1998.