ISSN: 2321-7782 (Online)
Volume 2, Issue 12, December 2014
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

A Survey: Various Approaches for Search Engine Query Optimization

Vinay Bhardwaj[1]
M.Tech Scholar, CSE Dept.
Chhatrapati Shivaji Institute of Technology
Durg (Chhattishgarh), India

Khomlal Sinha[2]
Sr. Assistant Professor, CSE Dept.
Chhatrapati Shivaji Institute of Technology
Durg (Chhattishgarh), India

Abstract: ***Number of users is increases every minute in internet environment that has little knowledge about services and feature of internet. In same way many user are new in their field, work and study .They have to improve their knowledge and information on that sector. In that situation the best resource of this type user is internet application like search engine. Recently there are many people know about search engine in internet. Search engine provide information to user based on query submit in engine. Search query is free from the structure query language .We can't apply this query in straightforward manner in database. The main challenges related to query processing is that to find out relation schemes and conversion of search query in structured way .So our space of research on query evaluation and optimization .Since the focus of this paper is on a survey: search query optimization and evaluation on centralized databases. We try to find out some techniques which are useful for search query optimization and evaluation.***

Keywords: search engine; query optimization ;query evaluation ;query representation ;query translation

I. INTRODUCTION

In real time information doesn't easy to get data from the data sources. When you go for search for any information in web that time you thought about the how the exact information will you get in single submit query. Basic problem to extract data that Client regardless to structure of relational database, they concern about their information and data form database. Now days user want quickly and easily data from resources without too much effort. User put queries multiple time and multiple pattern for exact data.

This type of queries is persistent queries that allow users to receive new results when they become available. Number of user fire query continuously on systems that can transform a passive web into an active environment, they need to be able to support millions of queries due to the scale of the Internet. No existing systems have achieved this level of scalability [2].

The safe and efficient implementation of the data resources .Computerized data have become a central resource of most organizations. Every user queries required central resource account by guaranteeing the safety of the data in the cases of concurrent access. One major criticism of many early data resources have been their lack of efficiency in handling the powerful operations they offer, particularly the content-based access to data by queries. User operates on relational databases and facilitates to apply query on them by allowing its user to issue keyword queries without any knowledge of the database schema or of SQL. On sets of tuples that are associated because they join on their primary and foreign keys and collectively contain all the keywords of the query [2].

Query optimization tries to solve this problem by integrating a large number of techniques and strategies, ranging from logical transformations of queries to the optimization of access paths and the storage of data on the file system level.

In the interest of space, the focus of the paper is primarily on the problem of optimizing queries in the user search query; we will concentrate on the problem of optimizing the execution of a set of retrieve only commands (queries).

II. PROBLEM IDENTIFICATION

Continuous query processing in web system for search engine, it is most heavily task by the server due to query evaluation procedures . There are some more factor which are effect to performance of search engine. First number of user correlated with the same information and query so it requires recursion. Second lack of statistical information about the database .Third query evaluation algorithms must rely heavily on heuristics. Queries are using several setting. The most obvious search engine system is that of direct requests by end users who need information about the structure or content of the database. If the requests are limited to a set of standard queries, they can be optimized manually by programming the associated search procedures and restricting the user's input to a menu format [2].

In dynamic or transaction information of database is updated frequently then query procedure get problem to evaluation and data consistency does not maintained. We have to maintain integrity constraints, and synchronize concurrent accesses correctly.

In optimization procedures either attempt to maximize the output for a given number of resources or to minimize the resource usage for a given output. Query optimization tries to minimize the response time for a given query language and mix of query types in a given system environment. This general goal allows a number of different operational objective functions. We find out some issue in query evaluation in search engine.

Communication effort: We have to transmits user query to site where they computation are performed and result presented. This effort can be minimized in presence of good speed of internet services and transmitting speed. If both are absent then response time will be increase and performance will decrease.

Storage access effort: Time for loading data pages from main memory. This is influenced by the number of data to be retrieved, the size of the available buffer space, and the speed of the devices used.

Computation effort: In search engine user apply non structure query for get information from database. We can't apply where clause , less than , greater than, AND , OR and NOT in user query . If user apply these expression then it treated as single query, not as conditional query. We apply effort to convert to non structure query into structure query then perform query evaluation and result representation.

In long-range distributed database with relatively slow communication lines, communication delay dominates performances, whereas the other factors are relevant only for local sub optimization. Storage accesses effort although the CPU process may be quite high for complex queries. Since the focus of this paper is on centralized databases, communication effort are not considered because in such systems communication requirements are independent of the evaluation strategy.

III. PROPOSED METHODOLOGY

A query is a language expression that describes data to be retrieved from a database. In the context of query optimization, it is often assumed that queries are expressed in a content-based manner, giving the optimizer sufficient choices among alternative evaluation. There are some techniques which are useful for search query optimization.

A. Query Representation

B. Query Transformation

C. Query Evaluation

D. Non standard Query optimizer

A. Query Representation

Queries can be represented in a number of forms for submit query. In the search engine content of query optimization, a query representation form must fulfill the following requirements: It should be powerful enough to express a large class of queries, and it should provide a well-defined basis for query transformation.

In DBMS the (tuple) relational calculus as introduced by Codd is a notation for defining the result of a query through the description of its properties. The representation of a query in relational calculus consists of two parts: the target list and the selection expression. The selection expression specifies the contents of the relation resulting from the query by means of a first-order predicate (i.e., a generalized Boolean expression possibly containing existential and/or universal quantifiers). The target list defines the free variables occurring in the predicate and specifies the structure of the resulting relation[2]. We can find relation among table from search query of engine and try perform query representation for user query which can be useful in query transformation.

B. Query Transformation

The transformation of a given expression into an equivalent one by means of well-defined rules is the subject of this section .The goals of query transformation are three objective : (1) the construction of a standardized starting point for query optimization (standardization), (2) the elimination of redundancy (simplification), and (3) the construction of expressions that are improved with respect to evaluation performance (amelioration)[2]. It is most useful approaches for query optimization due to performances of search engine. As shown in figure 1 query transformation is involved in step of cost based optimizer [10].

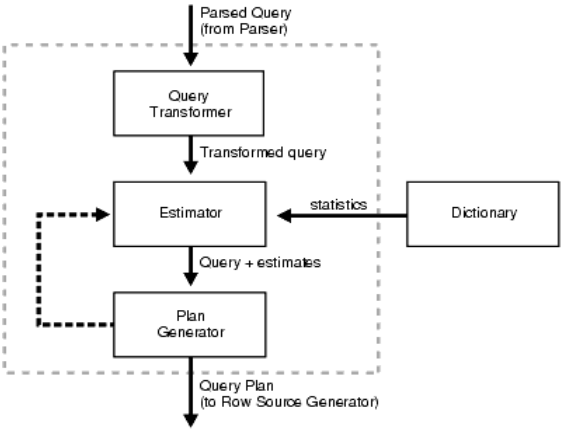

Figure 1 Cost Based Optimizer Component

C. Query Evaluation

The evaluation of query components of varying complexity such as one -value expressions, two-variable expressions, and multivariable expressions. The individual approaches can be viewed as the building blocks of a general query evaluation system. Based on search queries their associated effort of the query optimization process, which generates the optimal access plan .One-variable expressions describe conditions for the selection of elements from a single relation. It provides data result on single relation [2] .we can perform many time one variable expression in search engine because users are preferences single query frequently. Two-variable expressions describe combinational condition between two relation[2]. This type of search query apply user is less than average. Multiple variable expressions describe more than two combinational conditions between relations. This type of query apply by expert and analysts who know about query processing.

D. Non standard Query optimizer

Search engine system consist of different type database and different structure . This approach covers much of the work done in the area, some query-processing problems exceed the framework either because of a query complexity that goes beyond relational completeness or as a result of the structure of the underlying physical database. We can work on the complexity of

query and provide simplest way. One approach for non standard query optimization we have to required efficient translation of relational queries and view processing .

IV. Conclusion

An overview of query transformation techniques and query evaluation methods for database queries was given, using on search engine query processing. It was shown that a large body of knowledge has been developed to solve the problem of processing queries efficiently in web based database where user are applying document query or non structure query .We discussed query optimization research of queries on databases with computational capabilities, and the simultaneous optimization of multiple queries and update transactions. Other interesting areas only briefly addressed in this survey are query optimization in database systems that utilize more techniques for search query string for central and distributed database.

Acknowledgement

The heading of the Acknowledgment section and the References section must not be numbered.

Causal Productions wishes to acknowledge Michael Shell and other contributors for developing and maintaining the IEEE LaTeX style files which have been used in the preparation of this template. To see the list of contributors, please refer to the top of file IEEETran.cls in the IEEE LaTeX distribution.

References

1. Panagiotis G. Ipeirotis, Eugene Agichtein, Pranay Jain, Luis Gravano ," To Search or to Crawl? Towards a Query Optimizer for TextCentric Task ," SIGMOD 2006, June 27–29.
2. MATTHIAS JARKE ,JURGEN KOCH," Query Optimization in Database Systems". Computing Surveys, Vol. 16
3. Jianjun Chen David J. DeWitt Feng Tian Yuan Wang," An NiagaraCQ: A Scalable Continuous Query System for Internet", ACM 2000
4. Thanh Vu, Dawei Song, Alistair Willis, Son N. Tran, and Jingfei Li," Improving Search Personalisation with Dynamic Group Formation" , The Open University's repository of research publications 2014
5. Yun Zhou and W. Bruce Croft ,"Query Performance Prediction in Web Search Environments ",University of Massachusetts, Amherst.2005
6. Vagelis Hristidis,Yannis Papakonstantinou," DISCOVER: Keyword Search in Relational Databases", VLDB Conference, Hong Kong, China, 2002
7. Sanjay Agrawal, Surajit Chaudhuri, and Gautam Das," DBXplorer: A System For Keyword-Based Search Over Relational Databases", *ICDE*, 2002.
8. Sumit Ganguly,TVaqar Hasan,Ravi Krishnamurthy," Query Optimization for Parallel Execution", ACM 1992
9. Timos K.Sellies," Multiple query optimization" University of California , National Sciences foundation under the the grant CDR-85-00108
10. https://docs.oracle.com/cd/B10501_01/server.920/a96533/optimops.htm

Author(s) Profile

Vinay Bhardwaj, received B.E. (Computer Sciences & Engg) a in year 2007 and in pursuit for M.Tech. (Computer Sc.) from Chhatrapati Shivaji Institute of Technology (CSIT), Durg, Chhattisgarh, India. His interests are Data warehouse and data mining, Web technology and development and Advances computer architecture. Also he is having Life Membership of Indian Society of Technical Education, India (ISTE) and Membership No- LM 99074.

Sr. Assistant Professor Khomlal Sinha, received B.E. (Information Technology) in year 2005 and M.Tech. (Computer Sc.) in the years 2010 and 2012 respectively. Currently working as Sr.Assistant Professor in the Department of Computer Science & Engineering at Chhatrapati Shivaji Institute of Technology (CSIT), Durg, Chhattisgarh, India, His interests are Digital Image Processing, Operating Systems and Data Mining. Also he is having Life Membership of Indian Society of Technical Education, India (ISTE), Membership No- LM 74674 and Institutional Member of Computer Society of India (CSI).

