MJRET
Open Access
Multidisciplinary Journal of Research in Engineering and Technology
Journal homepage: www.mjret.in
ISSN:2348 - 6953

DQF FOR LARGE VOLUME DATA

Uzma Kokalgave[1], Prof. Soumitra Das[2]
Department of Computer Engineering
Dr. D .Y .Patil School of Engineering, Lohegaon,
Pune, Maharashtra, India.
uzma.kokalgave@gmail.com[1], soumitra.das.1974@ieee.org[2]

Abstract: *Designing of a query form is one of the easiest way to query databases. Due to this approach users can easily interact with the databases irrespective of the type of database is used. If we are allowing users to dynamically generate a query form then it can be a great help in querying database easily. In this paper I am proposing an approach for dynamically generating a query form so that we can capture user's interests during user interactions so that we can adopt a query form iteratively. While creating dynamic query form we want users to interact with such a data which is unstructured in nature. Such as photo, video, images, any text files of huge data. Today's world is of big data. Big data contains 80 to 90 percent of unstructured data which we capture from various sources like web pages. In this paper we are going to query such data and the interaction of users would be dynamically generating the query form. We will be using F-measure to estimate the perfectness of a query form generated by user. A DQF is an online system where users expect quick responses therefore efficiency is an important aspect to be taken care.*

Keywords*: Dynamic Query Forms, Sentiment Data Analysis, Unstructured Data.*

1. INTRODUCTION

Traditionally query forms are already designed and defined by developers or database administrator [1]. For various types of queries developers have to maintain hundreds of such query forms. And it was very difficult to design such a set of query form to execute variety of ad hoc database queries on very complex and huge databases. Dynamic query form approach will allow users to dynamically create query forms by selecting form components of their choice and until they do not get desired query result by performing certain iterations they

will repeat the process [2]. Due to this approach there will not be any overload on server to maintain huge set of query forms.

We are proposing an architecture where we can read unstructured data in our query forms. Unstructured data is a heterogeneous and variable in nature and comes in many formats including text, document, image, video and more. Unstructured date is growing faster than structured data. Unstructured data is everywhere, for example, satellite images, like Google Earth picture, scientific data such as atmospheric data, photographs and videos such as security, surveillance and traffic videos, Radar or Sonar data, such as meteorological or oceanographic seismic profiles. Some text files such as survey results emails, social media data such as YouTube data, Facebook, twitter data, linked-In and Flicker data, Mobile data such as text messages and location information and website content such as Instagram.

In this paper we are mainly focusing on text sentiment data analysis by using free open source software called Rapid Miner. Sentiment analysis is also known as opinion mining, sentiment analysis is used to determine the attitude of a speaker or a writer with respect to some topic or the overall contextualization of a document, For example. We can mine Twitter, Facebook and other social media conversations for sentiment data about customers view on competition, opinion on usage analytical tool and is an excellent prototyping platform due to flexibility and robustness. It has a comprehensive set of algorithms that allow us to quickly swap out and try different models.

2. EXISTING SYSTEM

In most of the previous work, database administrators and developers have created predefined static query forms for querying database but due to ad hoc query executions it become very difficult to maintain such set of query forms. Most of the technologies support customized query forms such as Microsoft access database etc. but the problem with these systems is that they are created for Professional developers and not at all for end users. After that dynamic query form approach come into the picture due to which it became easy for end users to modify existing query forms at run time and create new as per business requirements. There are other works have been successfully implemented for higher volume databases which are relational in nature.

In previous work a static query form generation has been implemented in which there would be certain set of predefined query forms depending on the query [1]. But in this system there was a difficulty in maintaining such a huge collection of query form and for ad hoc query execution, user will not get desired output. Some technologies allowed user to create

customized query forms such as ColdFusion, Microsoft Access database, but the problem with these technologies was that they are not meant for end users interactions. It means that the user who doesn't have knowledge of their databases will not be able to use those technologies. We have taken an idea of implementing dynamic query form for large volume databases [2]. In this system the DQF will be generated for relational database but there was no consideration for non-relational databases. After that some work has been done which includes non-relational databases in DQF [3]. Further we found certain work has been proposed for DQF using NoSQL for semi structure databases [4]. We found some work done on securing DQF using CAPTCHA in dynamic query form execution [5]. Some other work has been done on query form improvement. Most of the time it has been taken care to give quick responses and better performance of query form as DQF is and online application.

3. PROPOSED SYSTEM

In this paper, the main focus has been put on the type of data is being used for creating dynamic query forms. We will be using some unstructured data for example a text file (.txt) or pdf file with unstructured data. We will analyze that data in such a way that which can be ready to use for end users. For analysis of the data we will be using different analysis tools. Once data will be analyzed it will be ready for use. That data will be in some for structured format. Once it's done we will allow users to generate a dynamic query form and execute the query results until he gets required query result. We will allow user to rank the query Form. F-Measure will be used here for ranking a query form.

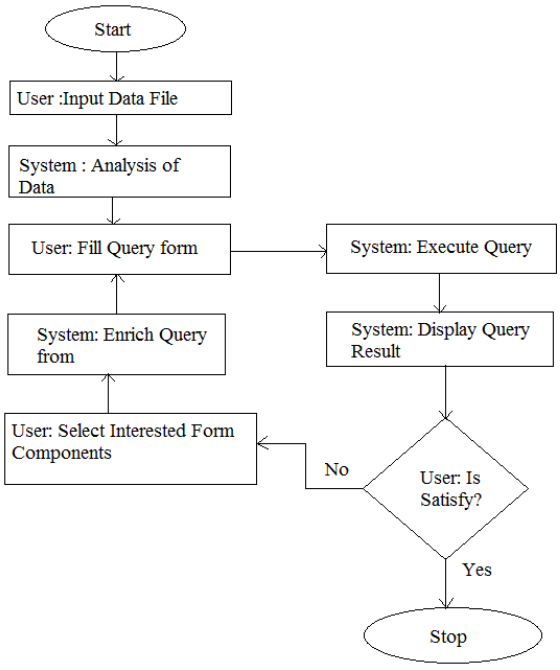

Fig.1 Basic Flow Chart of DQF

DQF will actually work in two phases one is Query Execution and Query enrichment [2]. In Query execution user will fill the query form with desired form components, query form will be executed by the system and result will be displayed. If user gets desired output of that query form then he will rank the query form for future use. Depending on the ranking, query form will be restored for further use. And then he can exit from the systems else he will enriched the query form by selecting other form components of his desire to get expected query result. After that he will execute the query form and the process will repeat until user will get desired output.

3.1 Ranking score estimation and ranking metrics:

For evaluating query results there are two measures available such as precision and recall. Based on different inputs provided to query forms can give different output in query results. To achieve expected query result we will be using expected recall and expected precision. Expected precision is the proportion of the query results which is interested by current user and expected recall is the proportion of users expected interested data instances which are returned by current query form.

For ranking score estimation two components will be ranked one is projection and another is selection [2]. In projection components ranking, entities (Tables) and their respective attributes (Columns) will be ranked. Here attribute with maximum F-score will be selected. And in raking selection form components first important attribute will be selected and ranked, here the F-score would be computed incrementally on desired attributes.

4. TAXONOMY CHART

The taxonomy chart shown below is a comparison with the existing work related to Dynamic Query Form Generation based on which we can get a clear idea about what type of work we would like to extend to the existing work.

Paper Reviewed	Parameters				
		Database Used			
	Dynamic Query form	Relational Data	Non relational Data	Unstructured Data	Ranking of Query Results
Dynamic Query Form for Database Queries	Yes	Yes	No	No	Yes
Dynamic Query form for Non relational Database	Yes	No	Yes	No	Yes
Randomized query formulation for Database Queries	Yes	No	Yes	No	Yes
DQF for Large volume Data	Yes	No	No	Yes	Yes

Table 1: Taxonomy for previous work and proposed work

5. CONCLUSION

In proposed system dynamic query form will be generated for unstructured data for querying on data such as image, video, or any sentiment data or survey result stored into any text file or pdf file to get desired output to the end users. In that system users are allowed to rank the query form based on which it can be enriched further. Thus we conclude our proposed system will help users to get desired query results using dynamic query form for large volume unstructured data.

ACKNOWLEDGEMENT

I would like to thank to our guide Prof. Soumitra Das, Head of Computer Engineering Department who has always been source of motivation for me. We would also like to thank to all other faculty and staff members of our Department, who have put their great effort and guidance at appropriate times without which it would have been very difficult on our Paper work.

REFERENCES

[1] M. Jayapandian and H. V. Jagadish. "Automated creation of a forms-based database query interface". In Proceedings of the VLDB Endowment, pages 695–709, August 2008.

[2] Liang Tang, Tao Li, Yexi Jiang, Zhiyuan Chen, "Dynamic Query Forms for Database Queries," IEEE Transactions on Knowledge and Data Engineering, 19 April 2013.

[3] S.BhaskaraNaik, B.VijayaBhaskar Reddy, "Dynamic Query Forms for Non-Relational Database Queries", International Journal of Engineering And Computer Science ISSN: 2319-7242 Volume - 3 Issue -8 August, 2014 Page No. 7415-7419

[4] VISHNU R , SWAPNA HARI , "DYNAMIC QUERY FORMS WITH NoSQL" International Journal of Research in Engineering & Technology (IMPACT: IJRET) ISSN(E): 2321 8843; ISSN(P): 2347-4599 Vol. 2, Issue 7, Jul 2014, 157-162

[5] Gopi Krishna Lakkasani, Balakrishna Nayudori "Random Query Formulation for Database Queries" International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 8, August 2014, ISSN: 2277 128X