RESEARCH ARTICLE   OPEN ACCESS

Retrieval of Information Using Fuzzy Queries

Priyanka Sharma[1]
1(CSE, Advanced Institute of Technology and Management, Palwal)

Abstract:

The classical or traditional information system provides answer after a user submits a complete query. It is even noticed that presently, almost all the relational database systems rely on the query which has syntax and semantics defined completely to access data. But often it is the case that we are willing to use vague terms in our query. The main objective of database management system is to provide an environment that is both convenient and efficient for people to use in storing and retrieving information. A recent trend of supporting auto complete is a first step to cope up with this problem. We can have design of both classical and fuzzy database and can use effectively fuzzy queries on these databases. Fuzzy databases are developed to manipulate the incomplete, unclear and vague data such as low, fast, very high, about etc. The primary focus of fuzzy logic is on the natural language. This Paper provides the users the flexibility or freedom to query database using natural language. Here this paper implements "interactive fuzzy search". This framework for interactive fuzzy search permits the user to explore the data as they type even in the presence of some minor errors. This paper applies fuzzy queries on relational database so that it is possible to have the precise result as well as the output for the uncertain terms we generally use based on some membership function

Keywords — **vague, fuzzy logic, interactive search, membership function, fuzzy variables**

I. INTRODUCTION

So far we have been implementing the classical or simple queries on classical database but here we are actually implementing the fuzzy queries on the classical because sometimes we are likely to use vague terms in our query. So our implementation in this paper is imparting the users an efficiency to make use of data using natural language. Classical SQL queries have outstanding capabilities in terms of data retrieval and gathering of answer from information stored at completely dispersed databases.

Human queries are rarely precise which poses challenges in efficient answer formation and data retrieval. These are based on human perception which is most of the time unclear and imprecise depending on world knowledge. The primary concentration of fuzzy logic is always on human language where reasoning with imprecise or inexact proposition is much more difficult, not user friendly. As it is known that a database management system includes in itself a catalogue of data which is related to one another in some form and a collection of programs for the purpose of accessing that

Interrelated data. And searching in a database is an important operation among other operations in database management system. As the size of database increases, the searching time also increases. So numbers of algorithms have been developed to improve the performance but these are only for the classical database [4]. This research aims to integrate query language with fuzzy logic in order to improve the data retrieval. Information retrieval is seen as the very active or progressive area while using fuzzy set theory. The fuzzy information retrieval makes use of the fuzzy sets to depict the documents and the level of membership to show the aptness of terms and vague compatibility measures with the aim of having status value for the retrieved document. This system gives references to documents on the basis of two main components and these two components are the document representation and the use query. Here the document representation is applied to those terms which are considered as the atomic part of the documents. And user query is explained using any language for query which is depending on these

terms and also permits the concatenation of the customer needs with the AND, OR, and NOT logical operator.

In the Boolean information retrieval model, a term which is appearing in a document is either relevant or it is irrelevant. Information retrieval is shown by inexactness in this model [3].

The main disadvantage of this model is that it cannot consider a semi degree of relevance. And the method that also combines the elements depending on their contents, it is assumed that document depiction is imprecise. So the users who are having any knowledge regarding data would like to use their knowledge of importance and the relevance of terms in any particular document. The value of membership function for any object in a non fuzzy set is considered to be 0 or 1 where 0 value specifies that the object is not the part of set and the value 1 indicates that the object is fully belonging to the set. A fuzzy set is considered as the generalization of the traditional set having the range of [0, 1].

It is the set to which an object can belong to some degree of membership. And it is assumed that if more objects are part of this set then the degree of membership will be higher. A fuzzy set F is taken as the subset of the union of discourse $T\{t1, t2,....................tn\}$. It is specified by a membership function μ. And range always lies between [0, 1]. Where μF(ti) is showing the membership degree of universe of discourse belonging to the fuzzy set F.

II.   LITERATURE SURVEY

A far-reaching literature survey has been done to arrive at this approach and this literature survey has been very beneficial from the all the aspects considered. This extensive literature survey belong to the below mentioned research papers.

These research papers were highly beneficial and we have taken many approaches into account and derived this idea. The literature review is presented as follows:

In the paper of Juan Carlos Santamarina and Jjean-lou Chameau, it is presented that the fuzziness which is present in comprehension is most of the time is given or shown in the outputs; the terms like as complexity, inexactness , and the structure of membership functions are purely related to each other in one or other way. And also the changes or variations between crisp extremes through the vagueness have characteristic manifestations. He has presented the emphasis on the practical implications of these ideas. Most of the applications are including developing standard and type h fuzzy sets, which are helpful in increasing the exactness and consistency of presented data, and also having the support for preparation of knowledge-based systems which are using fuzzy sets for their proper functioning.

In the paper of Abhijeet R Raipurkar, G.R.Bamnote on fuzzy logic based query optimization in distributed database, this paper it is given that the query optimization is a complicated task in a distributed environment having both client/server environment because the location of the data becomes a major factor. So by integrating query processing subsystem with a distributed database management system and also having fuzzy logic with it is used for the estimation of response time of the query, this paper concludes that the by using the fuzzy logic based method for query optimization in distributed database have very significant effect on the efficiency of query processing of distributed database.

In the paper of Miroslav Hudec on "approach to fuzzy database querying, analysis and realisation "it is presented that querying fuzzy variables is not only a simple tool for query; it also enhances the main idea of a query and also retrieves the extra priceless information along with it. Fuzzy method has also few shortcomings in itself that are clearly seen while querying a process. This paper is basically presenting these shortcomings of the fuzzy method and also the concepts or ideas which can be used for resolving those limitations.

In the paper of Kevin Woods, Diane Cook, Lawrence Hall on "learning membership functions in a function-based object recognition system they have presented the learning component for GRUFF system called OMLET that has the ability to learn the membership by itself from the number of objects which are also named with their category measures.

In the paper of Sujit Kumar Mondal, Joyassree Sen, md. Rabiul Islam, md. Shamim Hossian on "performance comparison of fuzzy queries on fuzzy database and classical database" it is given that the users have the efficiency and flexibility to interact with the database by using human language. The time and cost of both the classical and fuzzy query over traditional and fuzzy database (db) is also calculated in the paper and is showing the decrease in time cost while fuzzy querying on fuzzy database as compared to classical database.

III.   BENEFITS OF FUZZY QUERIES

While putting one's thought as query in terms of natural languages into database, a lot of problems are experienced due to inefficiency of relational database to handle such query. As it is not possible for computers to use natural language, there is number of ways which are having their own advantages and disadvantages. A fuzzy querying interface as meant here makes it possible for a user to employ his or her own dictionary of linguistic terms to be used in queries and these linguistic terms provide for a direct representation of presumed vagueness and imprecision of query. It provides the better representation for the user requirement by expressing imprecise condition through linguistic terms. The terms in fuzzy queries include certain vagueness or uncertainties that information system operated on two valued logic usually do not understand and this is the reason they cannot use these terms. For example: A customer of real estate agency looking for a house would rather express his or her own criteria using imprecise description as cheap, large garden only. And this possibility to use fuzzy linguistic term should be attractive not only for a non expert casual user. Similarly the query "give me the name of young people who are suffering from severe cancer" cannot be executed or processed directly by SQL since it contains the lot vagueness as young people and severe cancer. The purpose of using linguistic variable is to provide a means of approximate characterization of phenomenon that is not properly defined. Fuzzy logic also reduces the complication of mathematical analysis of difficulties in comparison to traditional queries or precise query. Whenever classical queries on classical database are applied then a kind of lack of expressiveness is seen because the real world data is not always very clear but simply has imperfections or vagueness. But on the other side the implementation of fuzzy queries on classical database leads to the more interactivity of human expression without having any effect on searching time. But on further proceedings it is seen that we move further we see that when we are implementing the fuzzy queries on fuzzy database then the cost here is less than the fuzzy queries on classical database because here fuzzy value is calculated during the building of fuzzy database so extra time to compute fuzzy value is not required here. Next this implementation gives the user the flexibility to query database using natural language. The lack of expressiveness can be entirely eliminated and searching time can be reduced by fuzzy database management system and using fuzzy queries on it. This implementation provides an easy to use functionality for the purpose of extracting or retrieving data similar to conventional structured query.

In the traditional information systems user is usually unaware about the data specially whenever there is the case that they are having less knowledge about the underlying data and is forced to make use of the traditional try and see approach to extract the information so we can say that this paper is presenting an interactive fuzzy search. The vast expansion of internet gives rise to a significant growth of a large number of databases. This framework has made the user to explore the data as they type even in the presence of some small errors. Frequently most of the relational database systems use query which are having all syntax and semantics which have been defined previously to retrieve data but sometimes we are likely to use uncertain terms in our query. Secondly if we use distributed database it permits much faster execution of the local queries and also make it possible to have the reduction in the traffic. But there is the main issue of maintaining data integrity. This method of fuzzy set theory with its membership function is widely used and is popular to form realistic explanations of evaluation. So this implementation provides the method to capture the uncertainty situation of the real world data.

IV. PROPOSED WORK

This paper is implementing the fuzzy query on relational database to have information that is the most relevant on some fuzzy basis. So for this purpose first of all we have to create a database and that can be created in any language either in SQL or in oracle. Then we have to create a search interface and again that can be in any language. And this interface must be created in such a way that it should connect with the database we have made in SQL or in oracle and when we fire any query on this interface, it should retrieve the result from the database made in SQL or oracle.
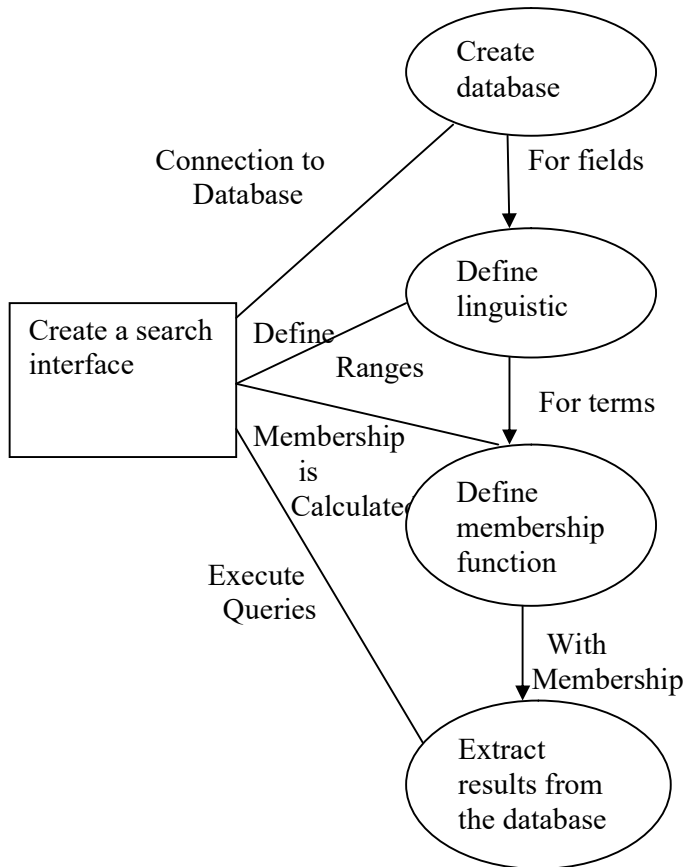

Figure1: Architecture of information retrieval using fuzzy query

So here it is started with the database that has been created in SQL language and after that a search interface is created in java that is all coding will be done in java. Then this all coding will be used in eclipse software which has some inbuilt functions. This software makes work easier.

Then the coding will connect to our database of SQL and fired query will retrieve the results from that database in SQL.

A. CALCULATION OF MEMBERSHIP OF LINGUISTIC VARIABLES

The membership of the linguistic variables is evaluated by a function known as "Membership Function". The membership function is denoted by π.

If there are three linguistic variables Low, Moderate and High and their range is between 10,000 to 25,000, 20,000 to 100000 and 70,000 to 130,000 respectively.

The membership for variable Low is given as:
μ (Low) =

$$\begin{cases} 0 \ when \ salary < 10{,}000 \ or \ salary > 25{,}000 \\ 1 \ when \ salary \geq 10{,}000 \ and \ salary \leq 20{,}000 \\ \frac{25{,}000 - salary}{25{,}000 - 20{,}000} when \ salary > 20{,}000 \ and \ salary < 25{,}000 \end{cases}$$

It means that the membership of the employee for salary low will be 0 when the salary is less than 10,000 or it is greater than 25,000. The membership of the employee for salary low will be 1 when the salary is greater than or equals to 10,000 and it is less than or equals to 20,000.
And membership of the employee for salary low will be (25,000 – salary)/ (25,000 – 20,000).

Similarly the membership for variable Moderate is calculated based on the below method.
μ (Moderate) =

$$\begin{cases} 0 \ when \ salary < 20{,}000 \ or \ salary > 100000 \\ 1 \ when \ salary \leq 25{,}000 \ and \ salary \geq 70K \\ \frac{salary - 20{,}000}{25{,}000 - 20{,}000} when \ salary > 20K \ and \ salary < 25K \end{cases}$$

$$\begin{cases} \frac{100000 - salary}{100000 - 70{,}000} when \ salary > 70K \ and \ salary < 1L \end{cases}$$

It means that the membership of the employee for salary moderate will be 0 when the salary is less than 20,000 or it is greater than 100000. The membership of the employee for salary moderate will be 1 when the salary is greater than or equals to 25,000 and it is less than or equals to 70,000.And

membership of the employee for salary moderate will be (salary – 20,000)/(25,000 – 20,000).

And membership of the employee for salary moderate will be (100000-salary) / (100000 – 70,000).

Now the membership for variable High is calculated as given below. The membership for variable high is as shown:

μ (High) =

$$\begin{cases} 0 \text{ when salary} < 70K \text{ or salary} > 130K \\ 1 \text{ when salary} \geq 1L \text{ and salary} \leq 130K \\ \frac{salary-70,000}{100000-70,000} \text{ when salary} > 70K \text{ and salary} < 1L \end{cases}$$

It means that the membership of the employee for salary high will be 0 when the salary is less than 70,000 or it is greater than 130000. The membership of the employee for salary high will be 1 when the salary is greater than or equals to 100000 and it is less than or equals to 130,000.And membership of the employee for salary moderate will be

(salary – 70,000) / (100000 – 70,000).

V. CONCLUSION

This Paper has provided an overview of retrieval of relevant information using fuzzy queries and then discussed some approaches of applying fuzzy queries to the database. By trying different techniques and algorithms or by comparing different techniques or by combining the algorithms, data analyst gives the best result. One approach is to apply fuzzy queries on fuzzy database and other one is to apply fuzzy queries on classical database. Here we applied fuzzy queries on relational database and we find some meaningful results after applying queries. After applying these fuzzy queries on the database, it is concluded that it is extremely helpful in reducing the lack of expressiveness in the queries and in handling the real world problems related with the inexact or vague data. It has provided us with a flexible system to deal with the most common human queries.

VI. FUTURE ENHANCEMENT

Handling of vague or inexact data is very difficult but it is an important thing to do and this retrieving information using fuzzy query has helped a lot to solve this problem. In future, fuzzy querying techniques need to be integrated with traditional techniques with the help of disciplines like machine learning, operation research, statistics and simulation. The main goal is to hide the complexity of real world data from end user and visualize the data which is easily read and understand by any novice user.

REFERENCES

1. George, J. Klir, Bo Yuan, "Fuzzy Sets and Fuzzy Logic Theory and Applications"Prentice Hall of India, 1997.

2. C. T. Yu, K. C. Guh, W. Zhang, M. Templeton, D. Brill, and A. L. P. Chen, "An Integrated Algorithm for Distributed Query Processing," in IFIP Conference on Distributed processingAmsterdam, 1987.

3. D. Chiang, L.R. Chow and N. Hsien, "Fuzzy information in extended fuzzy relational databases", Fuzzy Sets and Systems 92, pp.1-10. (1997).

4. Sujit kumar Mondal,Joyassree Sen, Md. Rabiul Islam, Md. Shamim Hossian," performance comparison of fuzzy queries on fuzzy database and classical database" in International Journal of Computer Application Issue 3, Volume 1 (February 2013)

5. Abhijeet Raipurkar, G.R. Bamnote, "Fuzzy Logic Based Query Optimization in Distributed Database" in International Journal of Innovative Research in Computer and Communication Engineering Vol. 1, Issue 2, April 2013.

6. J. Kacprzyk and S. Zadrozny, "On summarization of large datasets via a fuzzy-logic- based querying add-on to Microsoft Access. In: Intelligent Information Systems VII, Malbork, IPI PAN, Warsaw 1998, pp. 249-258.