Fuzzy to SQL Conversion using Gefred Model with the help of MATLAB

Poonam
Department of Computer Science
AIM &ACT, Banasthali University
Rajasthan

ABSTRACT
For Many Years, achieving unambiguous knowledge has been turned to a serious challenge for human being. The aim of this paper is to emphasize situation when classical {true, false} logic is not adequate for data selection and data classification. Linguistic expression like: high salary, young etc are very often used in life and in statistics. The goal of this paper is brief study of fuzzy logic and sets and how to make it suitable for database queries and classification tasks. Fuzzy approach is introduced with usual relational database model to handle linguistic queries. The purposed fuzzy approach provides flexibility when users cannot unambiguously set hidden boundaries between data. Our work gives the flexibility to query the database in natural language using FRDB, which permits to have a range of answers in order to offer to the user all intermediate variations, which in turn will enhance the expressiveness of human expression, without any effect on searching time and with reduced cost. In this paper we are using Query Builder tool of MATLAB to show the result of query. Fuzzy query interpreter helps to convert fuzzy query into SQL query without need to learn a new query language. In this paper, we extend the work of medina et al. to implement a new architecture of fuzzy DBMS based on the GEFRED model. This architecture is based on the concept of weak coupling with the DBMS SQL Server.

Keywords
Fuzzy SQL, membership function, query builder tool, database

1. INTRODUCTION
Databases are a very important component in computer systems. Because of their increasing number and volume, good and accurate accessibility to a database becomes even more important. Organizations work with very large data collections mainly stored in relational databases. But in every case classical logic{true, false} is not adequate for database queries, and here comes the fuzzy logic in picture which works on "degree of truth "rather than usual "true or false" logic and help in solving the problem of linguistic query processing with relational database. Linguistic expressions are interesting for data extraction, analysis, dissemination and decision making. Fuzzy query being a querying tool, it improves the meaning of the query as well as extracts additional valuable information. [11] Linguistic expressions and degrees of truth are being used to select the wanted scenario for the user. These linguistic expressions have logical meaning for user and define a data selection process in the natural language [8].The research area of fuzziness in Data Base Management Systems (DBMS) has resulted in a number of models aimed at the representation of imperfect information in Databases , or at enabling non-precise queries (often called flexible queries) on conventional database schemas . However, few works have been done from a practical point of view. Our work in this paper further extends that work by using query builder tool of MATLAB. Statistical indicators are often collected with some errors and vagueness and classical techniques may involve some inadequately selected or classified data. This paper after short introduction of fuzzy logic and previous work presents our research in fuzzy database querying area. This fuzzy query idea is explained in a case study.

2. FUZZY IDEA
2.1 Imperfect Information
Fuzzy databases are used basically to handle imperfect information. Imperfect information can be

Inconsistent, imprecise, ambiguous, uncertain or vague. According to Z.M. Ma, imperfect information is [1]

1. Inconsistent when some real world aspect is having more than one value. (Example – the age of a student is stored as 25 and 27.)
2. Imprecise when attributes value choice has to be made from any given interval or range. (Example – the age of a student is the set [23, 24, 25, and 26] or is in interval [50-65].)
3. Ambiguous when some elements of information lead to various possible interpretations.
4. Uncertain when there exists degree of truth in attribute value. It occurs due to lack of information. (Example- the possibility that age of a student is 30 is 80 %.)
5. Vague when attribute value is represented by linguistic variables. (Example – the age of student is „Young".)

2.2 Basic Terms
- Database: A classical database is a structured collection of information (record or data) stored in a computer.
- Fuzzy Database: A fuzzy database is a database which is able to deal with uncertain or incomplete information using fuzzy logic.
- Fuzzy Logic: Fuzzy logic is derived from fuzzy set theory by Zadeh (1965) dealing with reasoning that is approximate rather than precisely deduced from classical predicate logic. It can be thought of as the application side of fuzzy set theory dealing with well thought out real world expert values for a complex problem (Klir 1997).

  A membership function assigns degree of membership to each element of the set in fuzzy logic.

- Fuzzy Degrees: Fuzzy attributes, whose domain is in the interval [0, 1], although other values over this unit interval (such as possibility distributions) are also possible, which may be related to specific linguistic labels (like "a lot," "normal," etc.). So to keep it simple, usually only degrees in the interval [0, 1] are used. The important meanings of degrees used are: fulfillment degree, uncertainty degree, Possibility degree and Importance degree.
- Membership Function: Fuzzy membership function is being used to express fuzziness in the query. Zadeh proposed a series of membership functions that could be classified into two groups: those made up of straight lines, or "linear," and Gaussian forms, or "curved." [6]

2.3 Basic Preliminaries on Fuzzy Set Theory

Let X = {x1, x2,..., xn} be universe of discourse.

2.3.1. Fuzzy Sets

Fuzzy sets are extension and generalization of the basic concepts of crisp sets. It allows partial membership. A fuzzy set F over a universe of discourse X can be defined as set of ordered pairs [2].

$F = \{\mu F(x) / x: x \in X, \mu F(x) \in [0,1]\}$

Here, $\mu F(x)$ is membership function of the element x to the fuzzy set F. So, *a fuzzy set is thus defined by a function that maps objects in a domain of concern to their membership value in a set*. Such a function is known as membership functions (μ) whose value lies in range [0,1] i.e.

$\mu F(x): X \rightarrow [0,1]$

$\mu F(x) = 0$ indicates that x does not belong to the fuzzy set F.

$\mu F(x) = 1$ indicates that x completely belongs to the fuzzy set F.

The universe of discourse X can be classified as

1 .Finite or discreet universe of discourse X = {x1, x2,..., xn}, where a fuzzy set F can be represented by:

$F = \mu 1 / x1 + \mu 2 / x2 + ... + \mu n / xn$

Here, μi with i = 1, 2,..., n represents the membership degree of the element xi..

2.3.2 Different Forms of Fuzzy Sets to Calculate Membership function

We are using fuzzy membership function to express fuzziness in the query. Zadeh proposed a series of membership functions that could be classified into two groups: those made up of straight lines, or "linear," and Gaussian forms, or "curved." In our work we are using linear Trapezoid function. For implementing fuzzy conditions following membership functions have been used in our thesis work which describes *small values, about values and big values* respectively.

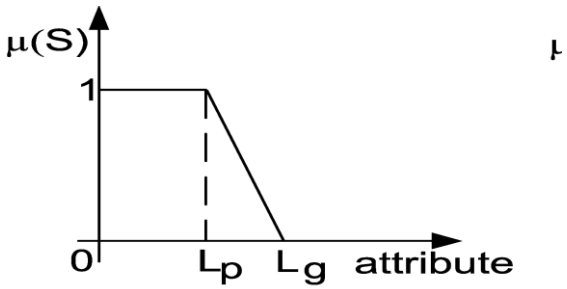

Figure 1(a) For Small Values

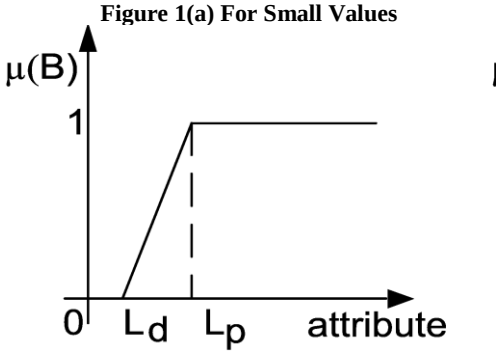

Figure 1(b) For Big Values

2.3.3 Linguistic Expression

In case when it is required to find records which have a big value of the attribute, the query has the form As in Figure1(b),and written as-

select attribute_1,…,attribute_n
from T
where attribute_p >=Ld

In the case when it is required to find records, which have a small value of attribute (Figure 1a), the query has the following structure:

select attribute_1,…,attribute_n
from T
where attribute_p <=Lg

In the case when it is required to find records that have attribute value about the exact value (Figure1 (c)), the query has the following structure:

select attribute_1,…,attribute_n
from T
where attribute_p >=Ld and attribute_p <=Lg

2.4 Classical Query

The SQL is used to obtain data from relational databases. Its advantages among others are the optimized work with RDBMS and understandable interpretation for users. The simple SQL is as follows:

select attribute_1,….., attribute_n

from T

where w attribute_p>P and attribute_r<R.

The result of the query is shown in graphical mode in Figure 2. Values P and R delimit the space of interesting data. Small squares on the graph show records that satisfy and not satisfy the query criteria. In the graph is obviously shown that two records are very close to satisfy query criteria whereas other records either satisfy the query or are far for satisfying it.

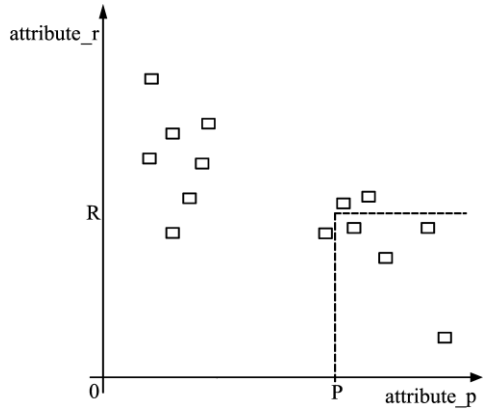

Figure 1(c):The result Of classical Query

2.4.1Drawbacks Of Classical Query System

The main drawback of classical query is that even the nearby values are present in database for a given query but the search show no result found. To get the result user have to expand the query again and again until he get the exact match in database. In short, classical query system makes a brittle selection. In this way more data from database are selected ,but user has lost the accuracy of his query. These two problems show the instead of changing the boundary conditions in the WHERE clause, it is necessary to change the way in which WHERE clause is evaluated.

The aim of this work is to present the query improvement with the fuzzy SQL approach. This development enables supporting queries based on linguistic expressions from user's point of view and also enables accessing classical relational databases in the unchanged structure**.**

2.5 THE GEFRED MODEL

The GEFRED (Generalized Fuzzy Relational Database) model was published in 1994 by Medina-Pons-Vila[11]. It is developed in possibilistic framework , so fuzzy domains are considered. It also includes thecae where the underlying domain is not fuzzy i.e. numeric. Various data types given in GEFRED are:

1. A single scalar (e.g., Age = Young, represented by the possibility distribution 1/Young).
2. A single number (e.g., Height = 160, represented by the possibility distribution 1/160).
3. A set of mutually exclusive possible scalars (e.g., Age = {Young, Old}, represented by {1/Young, 1/Old}).
4. A set of mutually exclusive possible numbers (e.g., Age = {14, 50}, represented by {1/14, 1/50}).
5. A possibility distribution in a scalar domain (e.g., Age = {0.6/Young, 1.0/Middle}).
6. A possibility distribution in a numeric domain (e.g., Age = {0.5/23, 1.0/26, 0.8/24}, fuzzy numbers or linguistic labels). It includes Umano-Fukami models' data types UNKNOWN ,UNDEFINED and NULL also.
7. An Unknown value with possibility distribution:
Unknown = {1/x : x ∈ D}
8. An Undefined value with possibility distribution:
Undefined = {0/x : x ∈ D}
9. A NULL value given by:
NULL = {1/Unknown,1/Undefined}

As GEFRED suggests, Generalized fuzzy domain (D) and generalized fuzzy relations ( )) concepts are being included. Generalized fuzzy domain (D) includes classic domain, with the possibility distributions defined for this domain and the NULL value also. If X is the universe of discourse and P (X) is set of all possibility distributions including unknown, undefined types and Null type. Then, generalized fuzzy domain (D) can be represented as D P(X) NULL.
Generalized fuzzy relations are relations whose attributes are having generalized fuzzy domain. Fuzzy attributes may have associated with compatibility attribute where compatible degree can be stored. Generalized fuzzy relations are given by two sets: Head H and Body B. The head includes the name of each one of the attributes, their domains, and their compatibility attributes (which are optional). The body includes the values of the m tuples:

$$R=\begin{cases} H=\{(A1:D1[,C1], ..., An:Dn[,Cn])\} \\ B=\{(A1:di1[,ci1], ..., An:din[,cin])\}\ i=1,..., m \end{cases}$$

GEFRED model also defines fuzzy comparators which are based on the classic comparators (<, >, =etc). When these fuzzy comparators are used on non-fuzzy values, meanings of comparators does not change i.e. classical output should come during crisp values.

3. SQL WITH FUZZY COMPONENT

3.1 Concept Of GLC(Generalized logical condition)

The generalized logical condition (GLC) for the WHERE part of the SQL based on linguistic expressions is created. This GLC enables matching fuzzy and classical constraints in the same WHERE clause and to select only records that have the query satisfaction greater than zero. These records are transferred to the client side where t-norm and t-conorm functions, which can be easy aggregated to n variable case, are used to calculate query satisfaction index for each of these records. The query compatibility index (QCI) indicates how the selected record satisfies a query request. If the record fully satisfies query, the QCI value is 1 and if record partially satisfies query conditions, QCI value is in (0,1) interval and represents the distance to the full query satisfaction. The QCI value 0 means that the record does not satisfy a query. It is also possible to use additional filtering functions to choose appropriate number of records or to set the threshold value of the QCI.

3.2 Concept Of Weak Coupling

Weak coupling approach with DBMS SQL Server is being worked upon. The concept of weak coupling is shown in Figure 6. The FRDBMS proposed respects the GEFERD model. The language of description and manipulation of the data is therefore FSQL. Seen that the FSQL language is an extension of the SQL language, a FRDBMS can model a RDB (described in SQL language) or a FRDB (described in FSQL language).

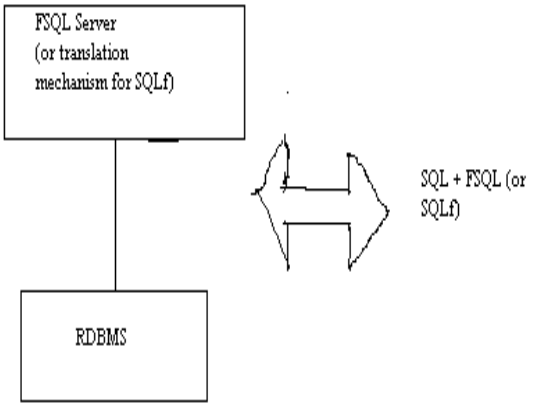

Figure 1(d) Concept of Weak Coupling

The principle of this coupling is the definition of a software layer that allows the transformation of the command written by the user in FSQL language in their equivalent written in SQL

3.3 FUZZY SQL INTERFACE (PROPOSAL)

An FRDBMS is a database which provides many functions such as description of data (DDL-Data Description Language), manipulation of data (DML-Data Manipulation Language), [5] integrity maintenance of FRDB using integrity rules, confidentiality by verifying access rights, management of competition of access, security in case of breakdown and user help. FRDBMS is capable of representing fuzzy information in many shapes and offers an adequate setting for storing and representing the information. [1] In our implementation we are going to show how the software layer in this architecture works, means how conversion of fuzzy query into SQL query is done. We are using Query Builder tool of Database of MATLAB to show the final report of our query

- ARCHITECTURE:

The new architecture of FRDBMS is based on *GEFRED* model [7] which was first proposed by *Bosc* and then by *Medina et al.* FSQL is being used as language of description and manipulation of data. FRDBMS can model RDB as well as FRDB. A software layer transforms the command written by user in FSQL language in their equivalent written in SQL. [3]The new architecture of FRDBMS is shown infigure1(f).

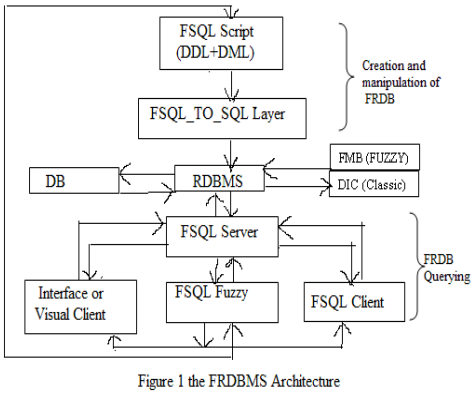

Figure 1 the FRDBMS Architecture

Figure 2. FRDBMS architecture

3.3.1 Case Study

This fuzzy query approach is under development for statistical information systems. Statistics have become one of the potential applications of data mining techniques, due to the great volume of available data from various databases. Statistics also works on data dissemination for various kinds of statistical data users. Classical tools also enable useful data dissemination but the fuzzy approach satisfies the demand on data in a human understandable form and supports the human-computer interaction in area of obtaining data using linguistic expressions.

This system is tested on data from a school. We are using database of a school to implement the concept of flexible querying. We take three tables -class, student and teacher. Class and student table have fuzzy as well non fuzzy attributes. In this case study student age is categorize as "young" and "toddler" and salary of teacher is as "high","low".This can be analyzed with a example as-

select Age

from Student

where Age is toddler

We have used the DBMS SQL Server in implementation. Figure 3 shows the result of one of the query. In the snapshot we can see the resulting rows of the table .After selecting required table, field and fuzzy value, we click "Generate SQL from Fuzzy query" button, which will give resultant table with desired fields and also the calculated attribute "Satisfying Degree" which in turn shows the satisfaction degree of correspondence row as per fuzzy value selected.

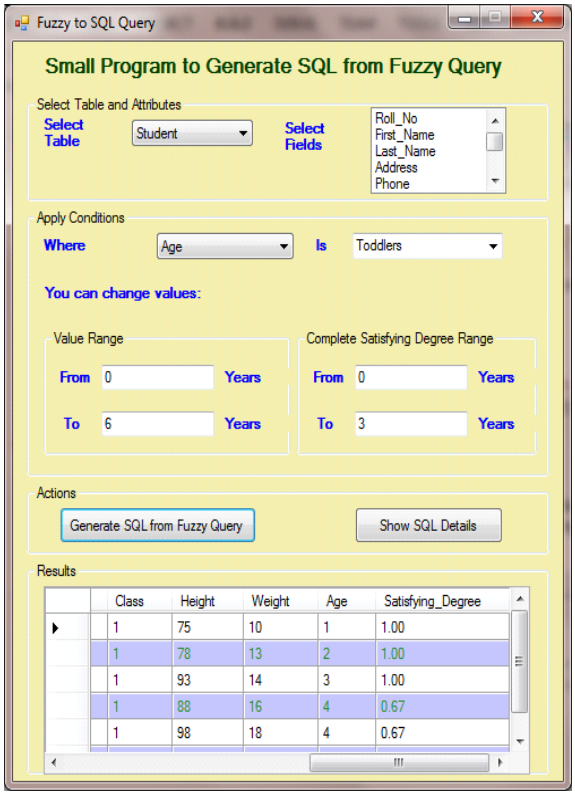

Figure 3: Result of selecting Table-Student, Condition Field =Age, Fuzzy Value =Toddlers

If we click "Show SQL Details "button, we get the query (which converts the vague data into precise form) in the form of SQL as shown in figure4.

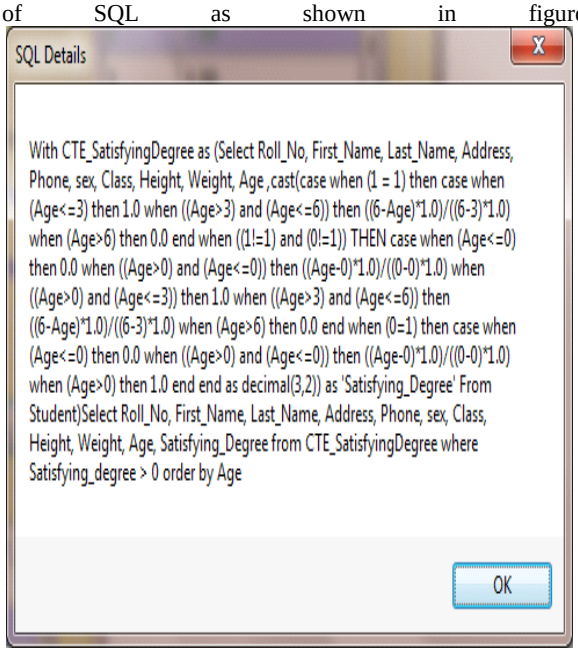

Figure 4: SQL corresponding to Fuzzy Query

Now this SQL is being implemented in MATLAB using MATLAB database tool. The tool used is Query Builder tool. After execution of this query in MATLAB we get how much memory is used, what is size of the variable created, which is shown in figure 5.

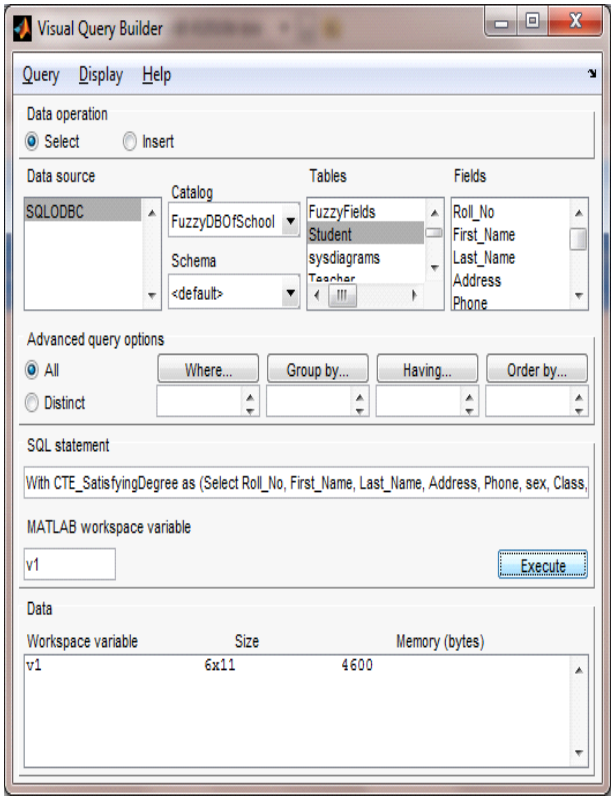

Figure 5: Execution of SQL query in MATLAB Query Builder tool

Now Query Builder displays the final result of query in the form of 3-D chart, which is shown in Figure6.

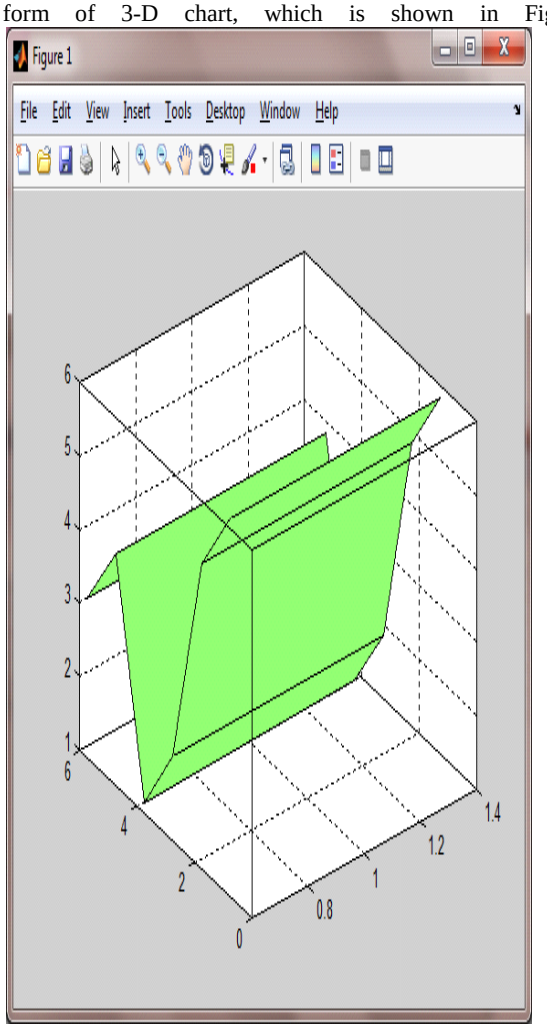

Figure 6: Query Builder tool displaying the chart of the result of query

All the fuzzy attributes of class students with their membership function are shown in Figure8.

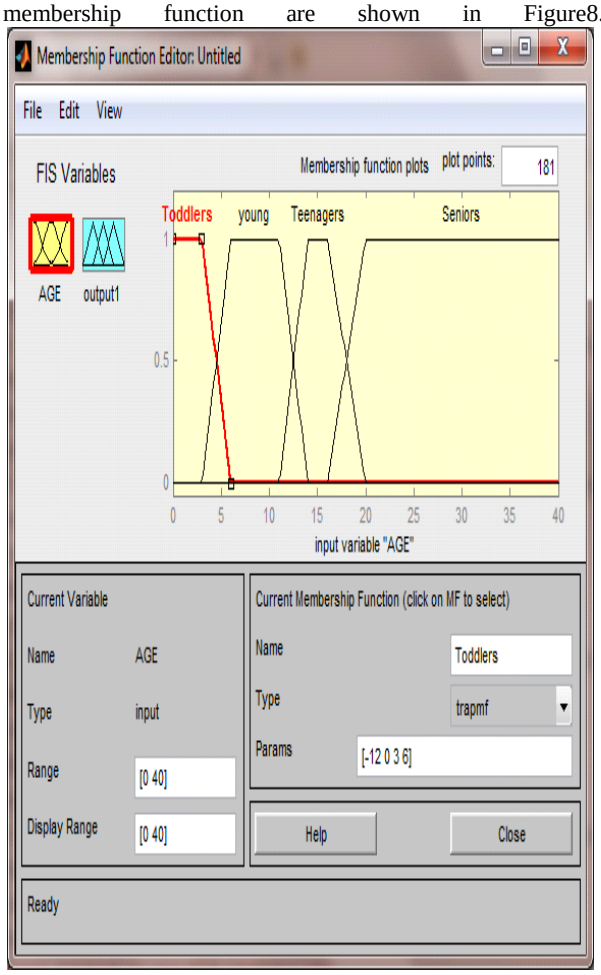

Figure 7: Fuzzy attributes of condition field AGE of table Student

So this implementation gives us a tool to convert fuzzy query into SQL query, which helps to give the exact value with satisfying degree.

4. CONCLUSION

For users a query is best described in the terms of a natural language without any exactness means with ambiguities and uncertainties. So when users work with a software tools they have to change the many valued logical thinking into two-valued computer logic i.e. SQL. The SQL requires crisp specification of a query. So there is a need of converting the fuzzy query into SQL query. This paper outlined the brief introduction of fuzzy relational models. This paper also discusses imperfect information. Fuzzy set theory basics are also being reviewed in brief. Use of fuzzy Logic provides the solution that allows the creation of Queries based on linguistic expression from the users point of view and does not change the structure and concept of obtaining data from relational database, enables an improved usage of SQL.The SQL and fuzzy approach creates a simple and easy way to use data mining tool.

- Further search could be directed towards improving the fuzzy SQL by developing the web application with a fuzzy module for data dissemination and the automatic mapping of existing relational DB to FRDB may be implemented.

5. REFERENCES

[1] Amel Grissa T. and Mohamed Hassine, "New Architecture of Fuzzy Database Management Systems," The International Arab Journal of Information Technology, Vol. 6, No. 3, July 2009 213

[2] Ben Hassine A., Grissa A.,Galindo J. and Ounelli H., " IGI Global: How to achieve Fuzzy Relational Databases Managing Fuzzy Data and Metadata" 2008

[3] "FRDBMS for Fuzzy Querying based on GEFRED Model" IJCA Proceedings on National Workshop-Cum-Conference on (RTMC2011), 2012.

[4] Galindo J., Urrutia A., and Piattini M., Fuzzy Databases: Modeling, Design and Implementation, Idea Group Publishing, Hershey, 2005.

[5] Grissa A., Ben Hassine A., and Ounelli H., "Extended_FSQL_Server: A Server for the Description and the Manipulation of FRDB," in Proceedings of the 4th International Multi Conference on Computer Science and Information Technology (CSIT), Jordan, pp. 454-464, 2006.

[6] Jose Galindo, "Handbook of research on fuzzy information processing in databases" IGI Global, 2008.

[7] Medina M., Pons O., and Vila A., "GEFRED: A Generalized Model of Fuzzy Relational Data Bases", Computer Journal of Information Sciences, vol. 76, no. 1, pp. 87-109, 1994.

[8] Miroslav Hudec, "An Approach to Fuzzy Database Querying, Analysis and Realisation", UDC 004.4'2, DOI: 10.2298/csis0902127H, 2009.

[9] P Gupta, "FUZZY QUERYING IN TRADITIONAL DATABASE," International Journal of Artificial Intelligence and Knowledge Discovery, Vol. 1, No.4, 2011.

[10] S.K.Mondal, J.sen, Md.R.Islam and Md.S.Hossian, "Performance comparison of fuzzy queries on fuzzy database and classical database," International journal of Computer Application Issue 3, Volume1 (February 2013)

[11] Dr. Rahul Rishi and Punkaj Gupta,"Database Design for Storage of Fuzzy Information inTraditional Datbase,"IJCA Journal,2011

[12] Handbook of Research on Fuzzy Information Processing in Databases by José Galindo,Volume 1