INTELLIGENT SEARCHING TECHNIQUES TO ANSWER QUERIES IN RDBMS

Thesis

Submitted for the award of the Degree of

Doctor of Philosophy
in
Computer Engineering

By

Smita Rajpal

under the supervision of

Prof. M.N. Doja

Department of Computer Engineering
Faculty of Engineering
Jamia Millia Islamia, New Delhi

Abstract

The study of “Searching Techniques” is very common in many areas of computer science; RDBMS is one of the many fields in which search is an important operation.

The existing searching techniques used in RDBMS are techniques which work well on precise data, but not on imprecise data, where data available are not always crisp in real life. Consequently, to deal with such type of data or to deal with the hidden uncertainty in searching, fuzzy logic(or higher order fuzzy logic) ,vague logic or neutrosophic logic could be the appropriate tools. There are a number of generalizations of Zadeh’s fuzzy set theory so far reported in the literature viz., i-v fuzzy theory, two-fold fuzzy theory, vague theory, intuitionistic fuzzy theory, probabilistic fuzzy theory, L-fuzzy theory, Neutrosophic theory etc.. The notion of vague theory recently introduced in IEEE by Gau and Buehrer and neutrosophic theory by samarandech is of interest to us for this present work. For each such generalization, one (or more) extra edge is added with the fuzzy theory with specialized type of aim and objective. Thus, a number of higher order fuzzy sets are now in literatures and are being applied into the corresponding more specialized application domains. While fuzzy sets are applicable to each of such application domains, higher order fuzzy sets cannot, because of their specialization in character by birth. Application of higher order fuzzy sets makes the solution-procedure more complex, but if the complexity on computation-time, computation-volume or memory-space are not the matter of concern then a better results could be achieved. Vague sets defined recently by Gau and Buehrer have also an extra edge over fuzzy sets. Today Databases are Deterministic. An item is either in the database or not, either the query answer is available in the database or not is a very serious matter .An item belongs to the database” is a probabilistic event, or a tuple is an answer to the query” is a probabilistic event, and it Can be extended to all data models; here we discuss probabilistic relational data. Two Types of Probabilistic relational Data are there,

Database is deterministic and Query answers are probabilistic or Database is probabilistic and Query answers are probabilistic.

Probabilistic relational databases have been studied from the late 80's until today. But today Application Need to manage imprecision's in data. Imprecision can be of many types: non-matching data values, imprecise queries, inconsistent data, misaligned schemas, etc.

The quest to manage imprecision is equal to major driving force in the database community is the Ultimate cause for many research areas: data mining, semi structured data, schema matching, nearest neighbor. Processing probabilistic data is fundamentally more complex than other data models. Some previous approaches sidestepped complexity. Now our implementation includes Ranking query answers. Since our Database is deterministic, the query returns a ranked list of tuples but our User interested in top-k answers. Sometimes we get the empty answers for the user queries in the deterministic database.

The present thesis entitled" Intelligent Searching Techniques to Answer Queries in RDBMS" is devoted to find out the Answer of Query if posed in natural language using Vague and Neutrosophic Search. The entire work is divided into twelve chapters. Chapter 1 deals with the Introduction of the current problem. Chapter 2 deals with the preliminaries in respect of vague logic, Neutrosophic logic and Relational Database Management System. In Chapter 3 we propose a new method of intelligent search called vague-search to find the most suitable match for the predicates to answer any imprecise query made by the database users. The method is based on the theory of vague sets introduced by Gau and Buehrer. A corresponding fuzzy method could be generated as a special case of our proposed method. It is also to be mentioned that the vague-search method could be easily incorporated in the existing commercial query languages of DBMS to serve the lay users better. In Chapter 4 we present a generalization of the i-v Vague Relation over Vague Relation. In this chapter our approach is to manipulate imprecise interval valued data through i-v vague relation which cannot be handled by vague relation. In

chapter 5 we propose a new method of intelligent search called interval valued (i-v) vague-search to find the most suitable match for the predicates to answer any imprecise query made by the database users. In Chapter 6 we present a generalization of the Neutrosophic set over Vague Relation. Our approach is capable of manipulating incomplete as well as inconsistent information. In Chapter 7 we propose a new method of intelligent search called Neutrosophic-search to find the most suitable match for the predicates to answer any imprecise query made by the database users. In this Chapter we are suggesting a new method called as α-Neutrosophic-equality Search to answer the queries of Relational database based on ranks. In Chapter 8 we are suggesting a new method called Neutrosophic Relations and their operators to solve the imprecise query based on α-Neutrosophic-equality Search, Neutrosophic Proximity search and combination of above two searches into single neutrosophic search. In chapter 9 we present a generalization of the i-v Neutrosophic relation over Vague Relation. Our approach is capable of manipulating incomplete as well as inconsistent information. In Chapter 10 we propose a new method of intelligent search called i-v Neutrosophic-search to answer any imprecise query made by the database users. In chapter11 we propose cooperative techniques for dealing with some practical problems that could arise in a flexible database querying context. The Two basic common problems which we came across are: the empty answers problem and the overabundant answers problem. A unified approach is proposed to overcome the shortcomings induced by both kinds of answers. It relies on a particular tolerance relation, which can be conveniently modeled in terms of a parameterized proximity relation. Such a relation allows for relaxing/intensifying the fuzzy /vague constraints involved in user's queries in a controlled iterative way.Chapter12 is on Conclusions and Future work.