Malaysian Journal of Computer Science, Vol. 10 No. 1, June 1997, pp. 42-46

DEVELOPMENT OF FUZZY DATABASE SYSTEMS

K. K. Phang
Mashkuri Hj. Yaacob
T. C. Ling
Faculty of Computer Science and
Information Technology
University of Malaya
50603 Kuala Lumpur, Malaysia
email: phang@fsktm.um.edu.my

ABSTRACT

Describes i) necessity of fuzzy DataBase ii) techniques used in the storage and retrieval of fuzzy data in a DataBase or information retrieval system iii) DataBase framework for fuzzy DataBase and iv) the advantages of using object-oriented DataBase framework in fuzzy DataBase. A prototype of fuzzy object-oriented DataBase system, (FOODS), has been developed to demonstrate the feasibility of fuzzy object-oriented DataBase system.

Keywords: ***Object-oriented DataBase management systems, Fuzzy set theory, Fuzzy object-oriented DataBase management system***

1.0 INTRODUCTION

Advances in Computer Science and the recognition of data as a crucial organisation resource have led to the rapid progress and development of information systems [1]. Information system, such as DataBase Management System (DBMS) and Information Storage and Retrieval System (ISRS), entails the gathering, management, utilisation and dissemination of information. In short, it is a means of furnishing information needed by an organisation [2]. The superiority of information systems over the traditional filing cabinets in terms of time and space efficiency has greatly extended the human's ability in data handling.

However, existing DBMS can only handle crisp, precise and non-ambiguous data. In other words, these systems do not cater for vague and ambiguous data which are fuzzy in nature - even though much of the human reasoning is based on fuzzy reasoning. In fact, fuzzy data arise constantly in real life from human thought and cognition processes and we often make decision based on them. It should be noted that information is useful only if it can be obtained in an easy and natural manner [3]. Clearly, as DBMS becomes more important for decision making, the problem of handling fuzzy data, being more compatible with human thought, will become increasingly important. Such a DBMS, if implemented, will improve human-machine interface and broaden the application areas of DBMS.

Attempts to build DBMS's which are able to represent and manipulate fuzzy data have received the attention of researchers recently. As a result, various models and prototypes have been proposed or implemented.

2.0 INFORMATION SYSTEM WITH UNCER-TAINTY HANDLING

In order to identify good practical ideas and evaluate the state-of-the-art of DBMS with uncertainty handling, some existing systems of ISRS and DBMS with uncertainty handling are examined. Our discussion includes ISRS as there are many similarities between ISRS and DBMS.

2.1 ISRS with Uncertainty Handling

An ISRS is a computer system which enables users to acquire information from a collection of documents stored in a DataBase [4]. Classical ISRS's are built within the framework of Boolean set theory. Lately, some approaches based on the framework of fuzzy set theory [5] are reported.

Bookstein [4] proposes a weighted Boolean Retrieval System in which the relative importance of each term in a query can be specified in the range of [0, 1]. Note that, the model will reduce to the classical one if the weights are restricted to {0, 1}.

Nomoto *et al.* [6] develops a fuzzy document retrieval system using fuzzy graph theory. Citations of documents are chosen as the criterion for fuzzification. Citation network, i.e., graph comprises citations and relation between them is created for fuzzy retrieval.

Miyamoto [7], on the other hand, develops a fuzzy document retrieval system based on keywords association. Keywords found in the title of a document are chosen as the criterion for fuzzification, i.e., these

keywords are related with all the keywords found in the ISRS by means of fuzzy binary relation. The usefulness of this model is demonstrated by implementing an efficient algorithm for fuzzy retrieval on a large scale bibliographic DataBase.

Ogawa *et al.* [8] develops a fuzzy document retrieval system using keyword connection matrix. This matrix records similarity between keywords and is used during query processing. Relevance of a document against a query is computed using fuzzy set theory. Their fuzzy approach shows a significant improvement in the measure of recall when compared to the crisp approach.

Lucarella *et al.* [9] discusses a knowledge-based fuzzy information retrieval system. The domain knowledge and the inferencing schema are based upon fuzzy set framework.

Generally speaking, the employment of fuzzy set theory in ISRS demonstrates a more natural query system with efficient algorithm when compared to the conventional ISRS approach.

2.2 DBMS with Uncertainty Handling

The application of fuzzy set theory in DBMS can be classified into two main classes. Class 1 concerns the study of fuzzy query processing in conventional (non-fuzzy) DBMS; Class 2 deals with DBMS which, besides having the ability to store and manipulate fuzzy data directly, also supports fuzzy query.

2.2.1 Conventional DBMS with Fuzzy Queries

Early DBMS's with uncertainty handling are developed within the framework of non-fuzzy DBMS. Generally, these systems deal with the construction and evaluation of fuzzy query against a crisp DataBase, and ignore the problem of direct representation of fuzzy data in DBMS. This section gives an overview of the direction and problems that researchers in this area attempt to address.

Chang *et al.* [10] explore the use of fuzzy query and propose the DataBase Skeleton concept which allows user to specify the contents and meaning of a collection of data. DataBase Skeleton is later used as a semantic base which supports fuzzy query. The methodology proposed is able to process fuzzy query such as the following:

Query: get supplierName ; goods equal 'tv'.

This query is fuzzy because it does not provide enough information to the DBMS - no access path is specified. The translation of the fuzzy query into a complete query is formulated as a problem of converting a partially-specified query graph into a completely specified one. A full query for the above query takes the form

Query: get supply.supplierName ; stock.goods equal 'tv'.

In Chang's paper, the term "fuzzy" refers to incompletely-specified information in the query, such as the lack of access path. It should be noted that fuzzy set theory is not used explicitly in formulating the methodology. Consequently, this model is less powerful, as it is less suitable in handling fuzzy query involving fuzzy data which are vague and ambiguous.

Wong [11] proposes a framework to handle incomplete information in non-fuzzy DBMS. Sometimes, due to incomplete or imprecise information in the DBMS, a DataBase cannot provide answers to some queries. In situation where the imprecision is due to recording error, obsolete data, incompatible scaling or measurement error, Wong suggests the use of statistical approach to produce approximate but meaningful answers. It is to be noted that Wong's model is based on the theory of probability and statistics rather than fuzzy set theory. As fuzzy data are, in the main, possibilistic rather than probabilistic in nature, this model like Chang's, fails to handle properly fuzzy query involving fuzzy data.

Tahani [12] develops a high-level conceptual framework for processing fuzzy query in a conventional non-fuzzy DataBase environment. The proposed framework, Fuzzy Retrieval System, uses the approach of associative retrieval scheme. Under this scheme, a fuzzy query is replaced by its associated meaning. Later, matching operation is applied to compare the fuzzy sets to the precise data to obtain an answer.

Kacprzyk *et al.* [13, 14] present a fuzzy query system called Fquery III. Through Fquery III, Dbase III plus (a commercial non-fuzzy micro computer-based RDBS) data can be operated on using fuzzy query. Fquery III is based on the framework of fuzzy set theory.

Wong *et al.* [15] develop a fuzzy query language for VAX Rdb/VMS (a conventional non-fuzzy mini computer-based DBMS). This fuzzy query system which is formulated under the framework of fuzzy set theory, also supports multi-criteria decision making.

Bosc *et al.* [16] discuss the extension of the SQL language to handle fuzzy query based on the framework of fuzzy set theory. The extended SQL has the following format:

Select n/t <attribute> from <relation> where <fuzzy condition>
where n and t are output regulating parameters. Here, the threshold t allows user to specify the desired membership

threshold. The presence of n serves to prevent a set of solution emerging by returning only those n-tuples relevant to the user. As a result, fuzzy query processing tends to become more efficient.

In general, under the conventional non-fuzzy DBMS environment, fuzzy query processing based on the theory of fuzzy set is more powerful when compared to those using ad hoc approach or probability theory.

2.2.2 Fuzzy DataBases with Fuzzy Quarries

Recent DBMS's with uncertainty handling are more advanced when compared to the earlier ones. They address the problem of direct representation of fuzzy data in the DBMS as well as the construction and evaluations of fuzzy query.

Buckles *et al*. [17, 18, 19, 20] propose one of the earliest version of Fuzzy Relational DataBase System (FRDBS) by merging the theory of fuzzy set and Relational DataBase System (RDBS). They formulate a robust theoretical framework of similarity-based FRDBS which has the following properties:

1. It allows none-atomic tuple components;
2. It requires similarity relation for each domain set of data in order to preserve important properties of classical RDBS;
3. It accepts user-defined threshold of acceptance during query evaluation;
4. It only supports a specific class of fuzzy number but does not support Possibility Distribution data type. (for comparison, see the model provided by Umano below)

Shenoi *et al*. [21, 22, 23] generalise the similarity-based model. They observe that the preservation of the properties of classical RDBS above can also be done by restricting the components of fuzzy tuples to be non-empty subsets of equivalence classes from domain partitions. Since the notion of equivalence classes is more general than the notion of similarity relation, an equivalence model of FRDBS, which is a generalisation of similarity based model, has been proposed.

Another different approach in the representation and manipulation of fuzzy data is advanced by Umano [24] who develops Freedom-0, a FRDBS. Unlike the model by Buckle *et al*. and Shenoi *et al*. which limits the fuzzy data to specific fuzzy number, Freedom-0 allows for both possibility distribution and Fuzzy number. However, even though Freedom-0 is more powerful in terms of its fuzzy data structure, it lacks the formal DataBase framework which is found in the models of Buckle *et al*. Freedom-0 uses an embedded programming language in Fortran for fuzzy data manipulation.

Zemankova [3] also develops a FRDBS which can handle both fuzzy set and possibility distribution data. RIM, a conventional non-fuzzy RDBS, is chosen as a host in implementing the FRDBS. Vector attribute type supported by RIM is used to represent fuzzy data. Extension is made to the RIM RDBS data structure so that tuple component is not restricted to atomic values. The FRDBS developed can handle fuzzy query.

Generally speaking, the FRDBS's mentioned above demonstrate the following major advantages over the conventional RDBS model [3].

1. It allows a more natural way of handling data because fuzzy data are more compatible with human thoughts and cognitions;
2. The use of fuzzy set and possibility distribution theory provides a formal mathematics foundation for the systematic representation and manipulation of crisp and fuzzy data;
3. It provides a DataBase environment to handle both crisp and fuzzy data.

However, in this model, there is no doubt that RDBS is the key DataBase framework in fuzzy data handling. However, like the older generations of DataBase technology, RDBS was developed for conventional data-processing involving crisp and atomic data structure, and is thus not suitable for applications requiring intensive computation and complex data-structure [25] like fuzzy set data.

In view of this, we. develop one version of Fuzzy Object-Oriented DataBase System (FOODS) by merging the theory of fuzzy set and Object-Oriented DataBase System (OODBS) [26, 27, 28]. FOODS can handle both fuzzy set and possibility distribution data. FOODS is implemented using an object-oriented programming language - Smalltalk/V. New object classes were created to represent and manipulate fuzzy data. Extension is made to the FOODS data structure so that DataBase object is not restricted to atomic values. FOODS can also handle fuzzy query, i.e., Fuzzy Object SQL which is an extension of SQL.

Generally speaking, FOODS demonstrates the following major advantages over the conventional fuzzy relational DataBase model.

Compatible data structure. Conventional Relational model supports only atomic data type, whereas the data structure of fuzzy data are often non-atomic. As a result, it is often difficult to represent fuzzy data in the conventional relational model. However, object-oriented model can handle non-atomic data with ease.

Better house-keeping. The querying process in fuzzy

DataBase requires intensive computation, fuzzification, defuzzification and operations on fuzzy data are difficult to handle in the framework of RDBS. Application programmer must explicitly manage these computations; this is not the case with fuzzy object-oriented DataBase model. Here computations are invoked automatically using message passing which is a powerful feature of object-oriented system.

Better modelling power. Relational model does not support semantic concepts like aggregation and generalisation relationships. Application programmer must explicitly handle these semantic concepts in their programs. In contrast to its relational counterpart, Object-Oriented model allows complex data structure and also provides greater modelling power [25].

3.0 CONCLUSION

This paper reviews various models of fuzzy DataBase systems in terms of fuzzy data representation and DataBase framework. It seems that among these, Fuzzy Object-Oriented DataBase model exhibits many advantages. We believe that this approach to the design of a fuzzy DataBase is feasible, versatile and have many advantages. In order to demonstrate the feasibility of the proposal, a prototype fuzzy object oriented DataBase, FOODS, has been created.

REFERENCES

[1] Howard L. Resnikoff et al. "Information Science" in *Identifying Research Areas in the Computing Industry to 1995*, Robert F. Cotellessa. Noyes Publications, 1984.

[2] Peter Gray, K. G. Kulkarni and N. W. Paton. *Object-Oriented DataBases - A Semantic Data Model Approach*. Prentice-Hall, 1992.

[3] Maria Zemankova-Leech and A. Kandel. Fuzzy *Relational DataBases - a Key to Expert Systems*. Verlag TUV Rheinland 1984.

[4] Abraham Bookstein. "Fuzzy Requests: An Approach to weighted Boolean Searches", *Journal of the American Society for Information Science*, July 1980.

[5] L. A. Zadeh. "Fuzzy Set As a Basis for a Theory of Possibility", *Fuzzy Set and System*, Vol. 1, 1978, pp. 3-28.

[6] K. Nomoto et al. "A Document Retrieval System Based on Citations Using Fuzzy Graphs", *Fuzzy Sets and Systems*, Vol. 38, 1990, pp. 207-222.

[7] S. Miyamoto. "Information Retrieval Based on Fuzzy Citations", *Fuzzy Sets and Systems*, Vol. 38, 1990, pp. 191-205.

[8] Yahushi Ogawa, T. Morita and K. Kobayashi. "A Fuzzy Document Retrieval System using the keyword connection matrix and a learning method", *Fuzzy Sets and Systems*, Vol. 39, 1991, pp. 163-179.

[9] D. Lucarella and R. Morara. "FIRST: Fuzzy Information Retrieval System", *Journal of Information Science*, Vol. 17, 1991, pp. 81-91.

[10] Shi-Kuo Chang and Wu-Haung Cheng. "DataBase Skeleton and Its Application to Logical DataBase Synthesis", *IEEE Transaction on Software Engineering*, Vol. SE-4, No. 1, January 1978.

[11] Eugene Wong. "A Statistical Approach to Incomplete Information in DataBase Systems", *ACM Transactions on DataBase Systems*, Vol. 7, No. 3, September 1982, pp. 470-488.

[12] Valiollah Tahani. "A Conceptual Framework for Fuzzy Query Processing - A Step toward very Intelligent DataBase Systems", *Information Processing & Management*, Vol. 13, 1977.

[13] Janusz Kacprzyk and A. Ziolkowski. "DataBase Queries with Fuzzy Linguistic Quantifiers", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-16, No. 3, May/June 1986.

[14] Janusz Kacprzyk et al. "FQUERY lll+: A 'Human-Consistent' DataBase Querying System Based on Fuzzy Logic with Linguistic Quantifiers", *Information Systems*, Vol. 14, No. 6, 1989.

[15] M. H. Wong and K. S. Leung. "A Fuzzy DataBase-Query Language", *Information Systems*, Vol. 13, No. 5, 1990.

[16] P. Bosc et al. "Fuzzy Querying with SQL: Extensions and Implementation Aspect", *Fuzzy Sets and Systems*, Vol. 28, 1988, pp. 333-349.

[17] Billy P. Buckles. "Fuzzy DataBases and Their Application" in [29], pp. 361-371.

[18] Billy P. Buckles and F. E. Petry. "A Fuzzy Representation of Data for Relational DataBases", *Fuzzy Sets and Systems*, Vol. 7, 1982, pp. 213-226.

[19] Billy P. Buckles and F. E. Petry. "Information-Theoretical Characterisation of Fuzzy Relational DataBases", *IEEE Transactions on Systems, Man and Cybernetics*, Vol. SMC-13, No. 1, January/February 1983.

[20] Billy P. Buckles and F. E. Petry. "Uncertainty Models in Information and DataBase Systems", *Journal of Information Science*, Vol. 11, 1985 pp. 77-87.

[21] Sujeet Shenoi and Austin Melton. "Proximity Relations in the Fuzzy Relational DataBase Model", *Fuzzy Sets and Systems*, Vol. 31, 1989, pp. 285-296.

[22] Sujeet Shenoi and Austin Melton. "An Equivalence Classes Model of Fuzzy Relational DataBases", *Fuzzy Sets and Systems*, Vol. 38, 1990, pp. 153-170.

[23] Sujeet Shenoi and Austin Melton. "Functional Dependencies and Normal Forms in the Fuzzy Relational DataBase Model", *Information Science*, Vol. 60, 1992, pp. 1-28.

[24] Motohide Umano. "Freedom-0: A Fuzzy DataBase System" in [29], pp. 339-347.

[25] Won Kim. *Introduction to Object-Oriented DataBases*, MIT, 1990.

[26] K. K. Phang and P. A. Lee. "The Application of Fuzzy Mathematics in DataBase Systems", *International Seminar on Mathematical Science, Kuching, Sarawak*, 1993.

[27] K. K. Phang and Mashkuri Hj. Yaacob. "Fuzzy Object-Oriented DataBase System", *IASTAD, Washington D.C., June 1996.*

[28] K. K. Phang. *FOODS - a Prototype Fuzzy Object-Oriented DataBase System*, M. Comp. Sc. Thesis, University of Malaya, 1994

[29] M. M. Gupta and E. Sanchez. *Fuzzy Information and Decision Process*, North-Holland 1982.

BIOGRAPHY

K. K. Phang is a lecturer in the Faculty of Computer Science and Information Technology, University of Malaya, Malaysia.

Professor Mashkuri Hj. Yaacob is the Dean of the Faculty of Computer Science and Information Technology, University of Malaya, Malaysia.

T. C. Ling is a lecturer in the Faculty of Computer Science and Information Technology, University of Malaya, Malaysia.