Journal of Intelligent Information Systems, 2, 311–317 (1993)


# *INTRODUCTION:* The Roles of Fuzzy Logic and Management of Uncertainty in Building Intelligent Information Systems

MARIA ZEMANKOVA mzemanko@mitre.org
*MITRE Corporation 7525 Colshire Dr., McLean, VA 22102*

JANUSZ KACPRZYK kacprzyk@ibspan.waw.pl
*Systems Research Institute Polish Academy of Sciences Ul. Newelska 6 01-447 Warsaw, Poland*

In today's world, virtually all human activities are faced with a highly demanding and competitive environment. As the economy progresses from being driven by the manufacturing technology to being dependent on the information technology, information becomes the most valuable "resource" that determines the success of an individual, a group, or an organization. We often hear the present era being described as the "information age." To satisfy the increasing demand for timely and relevant information, the availability of efficient hardware and software tools is necessary.

We are experiencing falling hardware prices while the processing power is rising, and there is a proliferation of database management systems (DBMSs) that claim to be efficient and user-friendly. Undoubtedly, we are witnessing an unprecedented progress in the availability of means for *data* processing, but at the same time we are also hearing complaints about an "information overload." Nevertheless, even the latest state-of-the-art systems fall short of satisfying the need for providing the critical *information* that enables a particular user to make an *intelligent/informed* decision in matters regarding their particular personal requirements, or those of a company, or a society. Indeed, an abundance of "raw" data is a liability, i.e., "data overload." However, data intelligently transformed into the *right type of information* or *knowledge* is a valuable asset, and one can never have too much of *useful* information or knowledge.

Many users feel that the current information technology is "computer consistent" rather than "human consistent." In general, the systems are designed to take the advantage of the computer architecture in order to make them more efficient, thus restricting their use to highly skilled specialists, with their jargon and trickery, and their way of representing the world in terms of often unnatural constructs such as normalized relations, consistent true-false rules, etc.,—i.e., the world eventually being boiled down to 0's and 1's. However, what an average user needs is a system that can (1) deal with human "imperfections," e.g., a preference for a natural language interface with all its imprecision, vagueness, implied meanings, incomplete sentences, misspelled/mispronounced words, etc., versus a highly structured programming language, e.g., a Boolean logic query language; (2) support the human ability of "approximate" reasoning, e.g., the human acceptance of and reasoning with partial truth; and (3) provide the data/information/knowledge representation paradigms that capture the human view of the world as being "imprecise" or "uncertain", rather than following the closed-word-assumption, that is, allowing only precise data and perfect knowledge. It is easy to see that these "man-machine inconsistencies" can make the acceptance of information technology difficult or its usefulness limited, as most systems do not provide the environment for information representation and retrieval that truly reflects the users' needs. In other words, in order to develop useful man-machine "synergistic" systems, the computers need to become more "human-oriented," while preserving their processing efficiency.

Though most of the users and developers of commercial DBMSs have been aware of the above inconsistency, it has not been until the late 1970s when some (primarily academic) work appeared exploring the use of the then formally established yet not widely accepted–theory of fuzzy sets, fuzzy logic and the related theory of possibility (Zadeh, 1965; Zadeh, 1976a; Zadeh, 1976b; Zadeh, 1978) in making the information modeling and retrieval more "human consistent" (Kunii, 1976; Tahani, 1977). This approach offered a simple formal foundation for "uncertainty management" in conjunction with "data management," and that is what was needed.

The 1980s had seen a boom in research efforts exploring the use of fuzzy logic in databases (Bosc et al., 1988; Buckles and Petry, 1982; Dubois and Prade, 1988; Dubois and Prade, 1989; Ichikawa and Hirakawa, 1986; Kacprzyk and Ziolkowski, 1986; Kerre et al., 1986; Prade, 1984; Prade and Testemale, 1984; Prade and Testemale, 1987a; Prade and Testemale, 1989; Raju and Majumdar, 1988; Umano, 1982; Zemankova and Kandel, 1984; Zemankova, 1989; Special Issue on Fuzzy Databases, Information Systems, 14(6), 1989), information retrieval (Bookstein, 1980; Kraft and Buell, 1983; Kerre et al., 1986; Prade and Testemale, 1987b; Radecki, 1983; Yager, 1980) and expert systems (Baldwin, 1983; Prade and Negoita, 1986; Zadeh, 1983; Zadeh, 1989). This wave of interest yielded predominantly theoretical results in a wide range of topics in information systems. Although most approaches advocated "fuzzification" of the widely used models, that is, generalizing or extending them for the added functionality (e.g., allowing natural language terms in a query language and using flexible matches in retrieval), the introduction of these ideas into commercial systems has been slow.

The development of information systems that are capable of dealing with various types of uncertainty (imprecision, vagueness, ambiguity, incompleteness, noise, flexibility, likelihood,...) and possibly employing different approaches for management of uncertainty (fuzzy logic, probability, belief functions,...) can progress along two paths. The first may adopt a standard commercial database as a starting point and construct some "add-on" features to accommodate the extensions needed for handling uncertainty. This approach seems to be more widely used, and quite promising as many relatively inexpensive yet powerful DBMSs exist. The second approach is to construct a new model that allows the representation and manipulation of uncertainty in data, information, or knowledge. Though the latter alternative may lead to a better and more comprehensive solution, the development of a new, reliable and efficient DBMS (that is interoperable with other major commercial systems) is a costly and lengthy procedure, hence may not be feasible as a short-term solution.

Recently, there has been a renewed level of activity in application of fuzzy logic in information systems (Bordogna, 1992; Bosc and Pivert, 1992; Buckles and Petry, 1991; Cross, 1993; DiCesare and Sahnoun, 1990; Dutta, 1990; Kamel et al., 1990; Ng and Abramson, 1990; Rundensteiner and Bic, 1992; Shenoi et al., 1990a; Shenoi and Melton, 1990b; Sudkamp and Cross, 1990; Tripathy and Sexena, 1990; Wang et al., 1990; Wong and Leung, 1990). Evidently, this is to some extent related to the amplified interest in fuzzy logic due to the spectacular commercial successes of fuzzy control in products highly visible to the general public such as home appliances, cameras, automotive parts, etc. Fuzzy logic enabled this development of "user-centered" products (one-button washing machines, self-focusing cameras, trains with a smooth acceleration/deceleration, etc.) that are capable of dealing with uncertainty in an efficient way. To this date there are no commercial information systems that are truly "user centered," that is, permitting the users to represent the (uncertain) world the way they (approximately) perceive it, express their (hard-to-quantify) information requests, and provide (the best possible) information in a format reflecting the user's (incomplete and changing) profile. We view the current situation as a window of opportunity for further research and development of commercial information systems that are designed with the real world and a human user in mind.

The recent increase of interest in fuzzy logic coupled with our belief that management of uncertainty plays a pivotal role in building "intelligent" information systems has motivated the production of this special issue. The contributions presented here discuss the broadly perceived intelligent information systems and concentrate on the application of fuzzy logic and more general paradigms for management of uncertainty. The common theme running through these research efforts is a desire to make the information systems more flexible and natural to use (both for data/information/knowledge representation and information request specification) and provide them with inference mechanisms that can take into account all pertinent information, even if it is imperfect, thus mimicking and/or enhancing human reasoning.

The approach to performing automated reasoning in probabilistic knowledge bases, developed by Dubois, Godo, Lopez de Mantaras, and Prade, provides a framework for representing and propagating uncertainty in situations when the facts or rules and the associated probabilities are not well known, but can be expressed in linguistic terms. For example, rather than having a strict implication rule, $A => B$, with crisp (precise) predicates $A$ and $B$, or a probabilistic statement "Between 80 and 90% $A$'s are $B$'s", we can have a linguistically expressed knowledge that "Almost all old people are wise," where "almost all" is a linguistic quantifier viewed as an imprecisely known, or fuzzy, probability, and "old" and "wise" can be fuzzy (imprecise) predicates. This form of linguistically represented knowledge is easier for users to understand, both in the stages of building the knowledge base, as well as representing (uncertain) results to (possibly imprecise) queries. It also permits more flexible and powerful reasoning, as instead of propagating strict numbers, qualitative terms are manipulated in accordance with their underlying interpretations. The results obtained through quantified syllogism and probabilistic rule chaining correspond to human common sense reasoning and also provide intuitive results where the knowledge is too complex (in terms of conditional probabilities) for a human to arrive at a conclusion. It is clear that progress in handling of conditional fuzzy probabilities can extend the capabilities of information systems by providing means for representing uncertain knowledge in a natural way and being able to supply answers to queries of the form "How many *A*'s are *B*'s", "How often *A*'s are *B*'s", etc., in a more realistic way than purely logic-oriented systems can do.

Baldwin and Martin describe a system for managing fuzzy and probabilistic uncertainties in a knowledge base composed of a fuzzy database and Fril rules that allow representation of uncertainty in terms of support intervals. The system can perform progressive refinement of queries and applies semantic unification for deriving new facts based on fuzzy data or incomplete knowledge. The incorporation of fuzzy/possibilistic and probabilistic uncertainties introduces a new degree of flexibility and realism into the computer-based handling of information, as without this flexibility, many reasoning paths would be cut short, as is the case when the data does not match exactly the conditions in the rules. In order to improve the user interface, the database and the rules can be represented in a graphical form that enables the user to perceive the structure and the relationships in the knowledge base. This framework is demonstrated on an "intelligent manual," that is, a technical reference document containing structured information about a certain domain. In this case the task described is the performance assessment of a hazardous waste repository.

Until recently, majority of research efforts focused on the enrichment of the conceptual model, that is, moving solely in the direction of the "user-centered" design without paying much attention to the computational complexity of the new features. However, efficiency of such enhanced models needs to be considered, as information is useful only if it is available in a timely manner. Mansfield and Fleishman concentrate on providing high-performance algorithms and hardware architecture for processing of database queries involving fuzzy predicates. Such queries place severe stress on the indexing and I/O subsystems of conventional database systems since they frequently involve the search of a large number of records. The DatacycleTM architecture and a research prototype described here use filtering technology permitting an efficient, exhaustive search of the entire database, thus obviating the need for complex index structures. Furthermore, this technique supports adhoc changes of fuzzy membership functions defining fuzzy predicates used in queries, yet provides deterministic response time largely independent of query complexity and the database size. This approach is combining both "natural language-like" extensions to the query language as well as their efficient processing.

In the contributions above, it was (implicitly) assumed that the data and knowledge represented in the system was either entered by the user or is more or less representing his or her needs. However, it is often the case that we need to use information systems designed and populated by others to satisfy our special information requirements. This is usually the case in information retrieval systems, where documents are preindexed by key words that we might not choose as their best descriptors, if we look at the document having a certain need in mind. Yager and Larsen develop a framework for capturing the concept of importance in the user's information need specification. The underlying process is based on the generalization (or fuzzification) of the requirements and application of special criteria aggregation operators (MOM and MAM). An envelope of potentially relevant items is constructed and presented to the user in a ranked order of their calculated relevance to the user's query.

Tailoring the query output to the user's specific needs is an important factor in providing useful information to the user. As the collections of data grow larger, for example, in digital libraries, it will be most important for the users to be able to represent their preferences, or "information need profiles." It is easy to imagine that such user profiles could act as our "information agents," knowing what is important to us, searching the vast collections of data and bringing to our attention only those items that are critical to or potentially useful in the professional tasks we perform, or relevant to our health, leisure activities, or personal lives.

The Editors hope that this special issue demonstrates the importance of uncertainty management in information systems and the included papers will motivate the readers to further explore these approaches to building intelligent information systems.

We would like to thank all of the authors who submitted papers to this special issue. Due to space and time constraints, some of them could not be included in this issue, but will appear in a future issue of the Journal of Intelligent Information Systems.

## References

Baldwin, J.F. (1983). A fuzzy relational inference language for expert systems. Proceedings of 13th IEEE Int. Symp. on Multi-Valued Logic, Kyoto, Japan, 416–423.

Bookstein, A. (1983). Fuzzy requests: an approach to weighted Boolean searches. *J. Amer. Soc. Information Sci., 31,* 240–247.

Bordogna, G., Carrara, P., and Pasi, G. (1992). Extending Boolean information retrieval: A fuzzy model based on linguistic variables. Proceedings of the IEEE International Conference on Fuzzy Systems, San Diego, CA 769–776.

Bosc, P., Galibourg, M., and Hamin, G. (1988). Fuzzy querying with SQL: extensions and implementation aspects. *Fuzzy Sets and Systems, 28*(3), 333–349.

Bosc. P. and Pivert O. (1992). Some approaches for relational databases flexible querying. *Journal of Intelligent Information Systems, 1* (3/4), 323–354.

Buckles, B.P. and Petry, F.E. (1982). A fuzzy representation of data for relational databases. *Fuzzy Sets and Systems, 5*, 213–226.

Buckles, B.P. and Petry, F.E. (1991). Towards a fuzzy object-oriented data model. Proceedings of NAFIPS'91, North American Fuzzy Info. Proc. Society, 73–77.

Cross, V.V. (1993). An analysis of fuzzy set aggregators and compatibility measures. Ph.D. Thesis, Wright State University.

DiCesare, F. and Sahnoun, Z. (1990). Linguistic summarization of fuzzy data. *Information Services, 52*, 141–152.

Dubois, D. and Prade, H. (1988). On incomplete conjunctive information. *Computers and Mathematics with Applications, 15(10)*, 797–810.F676

Dubois, D. and Prade, H. (1989). Processing fuzzy temporal knowledge. *IEEE Transactions of Systems, Man and Cybernetics, 19*(4), 729–744.

Dutta, S. (1990). Approximate reasoning with temporal and spatial concepts. Ph.D. Thesis, University of California, Berkeley.

Ichikawa, T. and Hirakawa, M. (1986). ARES: a relational database with the capability of performing flexible interpretation of queries. *IEEE Trans. Software Engineering, 12*(5), 624–634.

Kacprzyk, J. and Ziolkowski, A. (1986). Database queries with fuzzy linguistic quantifiers, *IEEE Trans. Systems, Man and Cybernetics, 16*, 474–479.

Kamel, M.S., Hadfield, B., and Ismail, M. (1990). Fuzzy query processing using clustering techniques. *Information Processing and Management, 26*(2), 279–293.

Kerre, E.E., Zenner, R.B.R.C. and DeCaluwe, R.M.M. (1986). The use of fuzzy set theory in information retrieval and databases. *J. Amer. Soc. Information Sci., 37*, 341–345.

Ng, K.-C. and Abramson, B. (1990). Uncertainty management in expert systems. *IEEE Expert, 18*(5), 29–48.

Kraft, D.H. and Buell, D.A. (1983). Fuzzy sets and generalized Boolean retrieval systems. *Int. J. Man-Machine Studies, 19*, 45–56.

Kunii, T.L. (1976). Dataplan: an interface generator for database semantics. *Information Sciences, 10*, 279–298.

Prade, H. (1984). Lipski's approach to incomplete information databases restated and generalized in the setting of Zadeh's possibility theory. *Information Sciences, 9*(1) 27–42.

Prade, H. and Testemale, C., (1984). Generalizing database relational algebra for the treatment of incomplete/uncertain information and vague queries. *Information Sciences, 34*, 115–143.

Prade, H. and Negoita, C.V. (eds.) (1986). *Fuzzy Logic in Knowledge Engineering*. Verlag TUV Rheinland, Koln.

Prade, H. and Testemale, C. (1987a). Fuzzy relational databases: representational issues and reduction using similarity measures. *J. Amer. Soc. Information Sci., 38*, 118–126.

Prade, H. and Testemale, C. (1987b). Application of possibility and necessity measures to documentary information retrieval. In: Bouchon, B. and Yager, R.R. (Eds.), Uncertainty in Knowledge-Based Systems, Springer Verlag, 265–274.

Prade, H. and Testemale, C. (1989). The possible approach to handling of imprecision in database systems. *IEEE Database Engineering Bulletin—Special Issue on Imprecision in Databases, 12*(2), 4–10.

Radecki, T. (1983). Generalized Boolean methods for information retrieval. *Int. J. Man-Machine Studies, 19*, 45–56.

Raju, K.V.S. and Majumdar, A.K. (1988). Fuzzy relational dependencies and lossless join decomposition of fuzzy relational database systems. *ACM Trans. on Database Systems, 13*(2), 129–166.

Rundensteiner, E.A. and Bic, L. (1992). Evaluating aggregates in possibilistic relational databases. *Data and Knowledge Engineering, 7*, 239–267.

Shenoi, S., Melton, A., and Fan, L.T., (1990a). An equivalence classes model of fuzzy relational databases. *Fuzzy Sets and Systems, 38*(2), 153–170.
Shenoi, S. and Melton, A. (1990b). An extended version of the fuzzy relational database model. *Information Sciences, 52*(1), 35–52.
Sudkamp, T.A. and Cross, V.V. (1990). Support generation in fuzzy relational databases. Proceedings of NAFIPS'91, North American Fuzzy Info. Proc. Society, 265–268.
Tahani, V. (1977). A conceptual framework for fuzzy query processing—a step toward very intelligent database systems. *Information Processing and Management, 13*, 289–303.
Tripathy, R.C. and Sexena, P.C. (1990). Multivalued dependencies in fuzzy relational databases, *Fuzzy Sets and Systems, 38*(3), 267–279.
Umano, M. (1982). FREEDOM-0: a fuzzy database system. In Gupta, M.M. and Sanchez, E., (Eds.), *Fuzzy Information and Decision Processes*. North Holland, Amsterdam, 339–349.
Wang, F., Hall, G.B. and Subaryono (1990). Fuzzy information representation and processing in conventional GIS software: database design and application. *Int. J. of Geographical Information Systems, 4*(3), 261–283.
Wong, M.H. and Leung, K.S. (1990). A fuzzy database-query language. *Information Systems, 15*(5), 583–590.
Yager, R.R. (1980). A logical on-line bibliographic searcher: an application of fuzzy sets. *IEEE Trans. Systems, Man and Cybernetics, 10*(1), 51–53.
Zadeh, L.A. (1965). Fuzzy sets. *Information and Control, 3*, 177–200.
Zadeh, L.A. (1976a). PRUF: A meaningful representation language for natural language. *Int. J. Man-Machine Studies, 10*, 395–460.
Zadeh, L.A. (1976b). The concept of a linguistic variable and its application to approximate reasoning. *Information Sciences, 8*, Parts I–II: 199–249 and 301–357, *9*, Part III: 43–80.
Zadeh, L.A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems, 1*, 3–78.
Zadeh, L.A. (1983). The role or fuzzy logic in the management of uncertainty in expert systems. *Fuzzy Sets and Systems, 11*, 199–227.
Zadeh, L. (1989). Knowledge representation in fuzzy logic. *IEEE Transactions on Knowledge and Data Engineering, 1*(1), 89–100.
Zemankova, M. and Kandel, A. (1984). *Fuzzy Relational Data Bases—a Key to Expert Systems*. Verlag TUV Rheinland, Koln. (Japanese translation by M. Mukaidono, Keigaku Publishing Co., Tokyo, 1987.)
Zemankova, M. (1989). FILIP: An intelligent information system with learning capabilities. *Information Systems, 14*(6), 473–486.