Foreword

Few scientific communities within computer sciences have thought over both their present and future like the one devoted to databases. During more than 15 years, a very comprehensive and well-known group of researchers of recognized prestige in this field meets regularly to fix the expected main challenges and problems within its scope and to propose what research lines are the most promising and necessary. From the first meetings in Laguna Beach and Palo Alto (1988, 1990), one can follow the proposals and results in the databases field through a series of reports, which in some way constitute a guideline for the development of research in this area (Abiteboul et al., 2005; Bernstein et al., 1998; Silberschatz, Stonebraker, & Ullman, 1991, 1996; Silberschatz, Zdonik, et al., 1996). Because of the experience and quality of the people involved in these meetings, we can assess the opportunity and novelty of any work in databases in light of the recommendations and research hints proposed in these reports. One of the lines that appear with more continuity and insistence is the treatment of imprecise and uncertain information in databases.

In 1991 the first report contained a section on new concepts in data models, which remarked on the management of uncertainty as a need for inclusion in new data models. At that time, data never being entirely precise, such as those in satellite photographs, justified the need. Later in 1996, the second report also posed new problems associated with vague queries concerning images, but in this case the imprecision was supposed to arise from two sources: first, imprecise features such as color, texture, and so forth, and second, imprecise valuations by the user of time and/or space, such as statements made that images are close to something or, for example, made in the morning. Also in this second report, data mining arose as a new trend in the treatment of information, and it was conceived as a new way for imprecise querying.

In the next report, the same ideas appeared again but included the interpretation and management of the imprecise results as one of the key research subjects to be studied.

Nevertheless, it was in the fourth and fifth reports where the need for including imprecision and uncertainty as natural elements in databases was more clearly reflected. Specifically, in the fifth report, the most recent one, with reference to approximate data, it is said, “When one leaves business data processing, essentially all data is uncertain or imprecise. Scientific measurements have standard errors. Location data for moving objects involves uncertainty in current position. Sequence, image, and text similarity are approximate metrics.” With reference to the imprecise queries, it is said, “users should also be able to ask imprecise queries and have the processing engine include this further source of uncertainty. Of course, with imprecise answers comes a duty for the system to characterize the accuracy offered, so users can understand whether the approximation is good enough for their needs.”

Data mining, conceived as a new form of accessing databases, is considered again as a top-priority research line in the fifth report, where it is also emphasized that data mining contains in its own essence the task of answering some kind of imprecise query. Quoting this fifth report, “users invariably point

out they have a single data mining query: Tell me something interesting," which is clearly an imprecise query.

These are the reasons why this book containing a collection of chapters devoted to research on fuzzy information processing in databases appears to be quite adequate and timely. Its key subject is directly focused on the resolution of problems that have been considered as being very important in all reports referenced. In other words, this book deals with some questions and tasks that the community of researchers of databases has been pointing out for several years.

The use of fuzzy logic to manage the imprecise and/or uncertain information in databases is even older than the aforementioned reports. It is a research line that is widely consolidated and has been developing for over 20 years. Traditionally, two major categories of work lines have been considered.

a. Those dealing with the problems of flexible querying to databases, in general, consider that the user expresses the query by using imprecise terms; the result is often a set of elements of the database affected to an accomplishment degree.
b Those addressing the description of data models include imprecise and/or uncertain attributes, relationships, and structures represented by fuzzy sets and fuzzy logic.

The book we are presenting includes some interesting and innovative chapters belonging to both work lines, starting with a chapter by the editor introducing basic concepts on fuzzy logic and fuzzy databases.

The chapter by S. Zadrożny, G. de Tré, R. de Caluwe, and J. Kacprzyk has a special interest since it offers a very wide review of fuzzy flexible querying. D. Dubois and H. Prade discuss the possibility of expressing negative preferences. R. Thomopoulos, P. Buche, and O. Haemmerlé study the use of ontologies in hierarchical queries. On the other hand, different approaches to solving flexible querying are also considered: case-based reasoning by G. de Tré et al., relative object qualification by C. Tudorie, and flexible queries using taxonomies by T. Andreasen and H. Bulskov.

The evaluation strategies for fuzzy queries are also studied by W. A. Voglozin, G. Raschia, L. Ughetto, and N. Mouaddib, whereas P. Bosc, O. Pivert, and A. Hadjali present a new study about the expressiveness of fuzzy sets illustrated by the fuzzy division operator. G. Xexéo and A. Braga give a new tool for fuzzy reasoning and querying applied to geographic information systems. In other chapters, M. Schneider defines fuzzy spatial data types, L. Liétard and D. Rocacher give an exhaustive list for the evaluation of quantified statements, and G. Bordogna and G. Psaila define a language for fuzzy querying in classical relational databases.

Formal extensions of the fuzzy relational data model are dealt with in the chapter by A. Takači and S. Škrbić, introducing a query language with the possibility to specify priorities in fuzzy statements, in the chapter by R. Belohlavek, about formal concepts, and in the chapter by R. Belohlavek and V. Vychodil, about similarities.

The interesting problem of implementing fuzzy database languages and systems is also dealt with in the book in works by A. Urrutia, L. Tineo, and C. Gonzalez, and M. A. Ben Hassine et al.

The chapter by C. Barranco, J. Campaña, and J. M. Medina studies the object relational approach and the fuzzy object-oriented data model. The above-mentioned chapter by Zadrożny et al. presents both the relational and object-oriented cases where fuzzy queries are made to fuzzy data models.

As we already mentioned, data mining appeared as a major research area inside the database field in the first challenge report 15 years ago, and it has been successively included inside this category in all subsequent reports. The use of fuzzy sets and fuzzy logic in data mining has been widely extended,

and in fact, before data mining was properly considered as a research area, some of its problems were addressed by means of fuzzy approaches. In this sense, let us remember the well-known fuzzy extensions of the *K*-means method for clustering problems, or the use of fuzzy rules in classification models. However, it has been in the last decade that this research line was consolidated by appearing in a wide variety of suggestive results, such as new fuzzy clustering approaches, different fuzzy association-rule definitions, new fuzzy classification techniques, and so forth.

This book also offers interesting results about fuzzy data mining topics.

First, the chapter by B. Feil and J. Abonyi is an excellent theoretical review. The use of fuzzy decision trees is studied in one chapter by M. J. Beynon. The extraction of fuzzy association rules is discussed by W.-H. Au and Y. Wang in their respective chapters. C. Fiot studies sequential pattern discovery, S. L. Wang et al. study fuzzy functional dependencies, and A. K. Sharma, A. Goswami, and D. K. Gupta define fuzzy inclusion dependencies. Fuzzy classification is studied by A. Meier, G. Schindler, and N. Werro.

Subjects associated with data mining such as data cleaning and decision making are presented in the works of H. H. Shahri and M. J. Beynon, respectively. S. Turgay proposes an agent-based fuzzy data mining structure, and missing data are studied by J. I. Peláez, J. M. Doña, and D. La Red by using the so-called fuzzy imputation approach.

Additionally, the diversity of the theoretical results cited above, including topics such as fuzzy information processing, should generate many different applications, which can be mainly found in the chapters by Y. Veryha et al., Y. Chen et al., and R. Carrasco et al., as well as in the examples and demonstrations of other chapters.

Summarizing, we can state that the present book offers an excellent perspective about what is now being investigated about uncertainty and imprecision management by means of fuzzy sets and fuzzy logic in the field of databases and data mining. Furthermore, all chapters include good introductions to their respective topics, good lists of references, and some key terms with concise and useful definitions. Therefore, we are sure that this handbook will be very informative and useful for a broader class of researchers, students, and companies related to the database world.

Professor Dr. M. Amparo Vila and Professor Dr. Miguel Delgado
University of Granada
Granada, Spain

REFERENCES

Abiteboul, S., Agrawal, R., Bernstein, P., Carey, M., Ceri, S., et al. (2005). The Lowell Database Research Self-Assessment. *Communications of the ACM, 48*(5), 111-118.

Bernstein, P., Brodie, M., Ceri, S., DeWitt, D., Franklin, M., et al. (1998). The Asilomar Report on Database Research. *ACM SIGMOD Record, 27*(4), 74-80.

Silberschatz, A., Stonebraker, M., & Ullman, J. D. (Eds.). (1991). Database systems: Achievements and opportunities. *Communications of the ACM, 34*(10), 110-120.

Silberschatz, A., Stonebraker, M., & Ullman, J. D. (Eds.). (1996). Database research: Achievements and opportunities into the 21st century. *ACM SIGMOD Record, 25*(1), 52-63.

Silberschatz, A., Zdonik, S., et al. (1996). Strategic directions in database systems: Breaking out the box. *ACM Computing Surveys, 28*(4), 764-778.