Preface

The increasing use of information systems by governmental agencies and businesses has created mountains of data that contain potentially valuable knowledge. Admittedly, these data do constitute "golden mines" which should be swiftly and efficiently processed and interpreted to be useful. Users (e.g. decision-makers) would like to efficiently reveal relevant data. Moreover, users are often not interested in large sheets of figures, but in knowledge that is usually overshadowed by large amount of data.

People can relatively easily answer imprecise questions like, *is it true that most of tall persons in the room wear blue or green shirts*? Different hues of these colours as well as the meaning of the vague term *tall people* are not limitations for solving this task. However, if we want to know, which of these two sentences: *most of young commuters commute short distances; most of medium aged commuters commute short distances* better explains the commuting behaviour, then we have to adapt this query to mine the truth value form the data. The same holds for querying *cheap hotel with good references and if possible near to the city centre* and common-sense reasoning: *if customer buys products very often, then provide high discount*.

The initial research in the theory of fuzzy sets and fuzzy logic was motivated by the perception that traditional computing techniques are not effective in dealing with problems, in which vagueness, imprecision and subjectivity are immanent, and therefore should not be neglected. These types of uncertainty are commonly called fuzziness.

According to Prof. Zadeh, four principal rationales for handling fuzziness exist. Two of them, which are relevant for this book, are: "don't need rationale" and "don't know rationale". In the former, the tolerance for imprecision is in accord with the remarkable human capability to solve variety of tasks without precise calculations. For example, summarizing data by short questions of natural language; creating queries with flexible conditions and approximate inference. In the latter, the values of attributes are not known with sufficient precision to justify the use of traditional databases for storing these data. Many data cannot be adequately expressed as precise numbers or as one linguistic term, due to non-sharp

boundaries of observations, tendency of people to estimate or guess answers in surveys and tolerance intervals of measurement instruments. Therefore, the data are often vague and include both quantitative and qualitative elements. Storing these data as crisp values might cause loss of valuable information.

Keeping in mind the aforementioned facts, fuzzy queries, fuzzy inference processes, linguistic summaries and managing fuzzy data in information systems could be the option. We have chosen these areas, because businesses of all sizes and governmental agencies cope with them in their work. The motivation for this book has arisen from the author's experience in teaching courses of fuzzy logic for business informatics and database design and in research and development of information systems and data mining applications mainly for the official statistics purposes. Furthermore, many small- and medium-sized enterprises cannot afford sophisticated tools or experts for information systems and data mining, even though they are aware of limitation of sharp boundaries in data analyses. Many tasks can be solved in a classical way, but their complexity becomes high. The complexity of the problem can be reduced by including the intensity of the examined property. This permits us to discern elements with the same property, based on the intensity matching it.

Roughly, the intent of the book could be depicted in Fig. 1. The usual scenario is that user wants to retrieve data or summarized information from a database. Furthermore, user might be interested to classify data. Often user is not aware of the nature of collected data or cannot determine sharp criteria. In addition, all data including vague ones are usually stored as crisp values.

In the book we examine these approaches theoretically as well as on the municipal statistics data. The latter is illustrated in appendixes. These data are suitable source, due to larger number of municipalities, which are often very similar in several attributes. Second reason is that some of attributes are fuzzy in their nature, but are limited to crisp values.

We should not expect that domain experts are familiar with the fuzzy logic theory. Therefore, the book demonstrates developing user-friendly interfaces to allow users exploring advantages of fuzzy logic in their tasks. Furthermore,
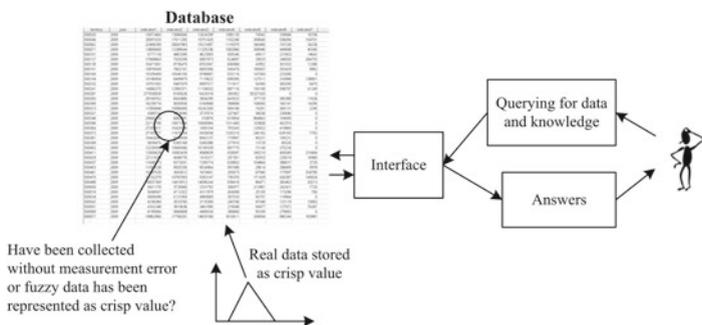

Fig. 1 Intent of the book

companies usually keep data in relational databases. We should keep this in mind during adapting database to cope with the fuzzy data.

The main target audience of the book are students, researchers and practitioners working in fields of data analysis, database design and business intelligence. This book does not go deeply into the foundation and the mathematical theory of fuzzy logic and relational algebra (e.g. theorems' proofs). Hence, intermediate knowledge of fuzzy logic and relational databases is recommended.

The book is divided into six chapters in the following way. Chapter 1 is focused on the theory of fuzzy sets and fuzzy logic to a level, which is advisable to know in order to proceed to next chapters. Readers skilled in fuzzy logic theory can skip this chapter.

Chapter 2 is devoted to flexible queries. The following aspects of flexible queries are examined: constructing fuzzy sets for query conditions; aggregation operators for commutative and non-commutative conditions with and without priorities; dealing with empty and overabundant answer problems and issues related to practical realizations.

Chapter 3 is dedicated to linguistic summaries. We start with the basic linguistic summary and build more complex ones. To meet this goal, selecting appropriate aggregations, implications for preferences and issues related to construction of membership functions are examined. Quality measures of created summaries are also considered. Finally, several possible applicabilities are discussed.

Chapter 4 presents fuzzy logic control architecture adjusted to the aims of business and governmental agencies. It shows fuzzy rules, construction of fuzzy sets and procedures for solving inference tasks by generalized *modus ponens*. In the first part we explain reasoning procedures. In the second part fuzzy expert systems are discussed. In the last part classification by IF-THEN rules is examined.

Chapter 5 covers fuzzification of classical relational databases. We briefly review classical relational databases and fuzzy database models. The emphasis is on storing fuzzy data in classical relational databases in a way that existing data and normal forms are not affected. Furthermore, practical aspects of user-friendly interfaces for storing, updating, querying and summarizing are examined.

Chapter 6 shortly discusses possible integration of fuzzy queries, summarization and inference related to crisp and fuzzy databases. Use of these approaches in a complementary, rather than competitive way, can support variety of tasks.

Finally, we suppose that the book will provoke at least some interest to continue research and also will be of support for developing tailored applications communicating with users by easy-to-use interfaces. Maybe the next generations of relational database management systems and applications will include many fuzzy characteristics and users will enjoy easy-to-use interfaces for fuzzy queries, fuzzy inferences, fuzzy summarization, fuzzy recommending and so on, without the necessity of knowing mathematics of fuzzy logic. We hope that the book will contribute to this field with a membership degree greater than 0.25.

Bratislava   Miroslav Hudec
April 2016

Fuzziness in Information Systems
How to Deal with Crisp and Fuzzy Data in Selection,
Classification, and Summarization
Hudec, M.
2016, XXII, 198 p. 91 illus., Hardcover