

HANDBOOK OF RESEARCH ON

# FUZZY INFORMATION PROCESSING IN DATABASES



JOSÉ GALINDO

VOLUME I

# Detailed Table of Contents

**Foreword** .....xxiii

**Preface** .....xxvi

**Section I**  
**Introduction**

**Volume I**

**Chapter I**

Introduction and Trends to Fuzzy Logic and Fuzzy Databases ..... I

*José Galindo, University of Málaga, Spain*

This chapter is a basic chapter for novel researchers in the area of fuzzy logic. It introduces the main concepts, like fuzzy sets and fuzzy numbers, linguistic labels, membership functions, the representation theorem and the extension principle, fuzzy set operations like union and intersection (t-norms and t-conorms), negations, fuzzy implications, different comparison operations, fuzzy quantifiers, and the possibility theory. With respect to the fuzzy databases, this chapter gives a brief introduction to this topic and enumerates a list of six research topics in this field.

**Chapter II**

An Overview of Fuzzy Approaches to Flexible Database Querying ..... 34

*Slawomir Zadrozny, Polish Academy of Sciences, Poland*

*Guy de Tré, Ghent University, Belgium*

*Rita de Caluwe, Ghent University, Belgium*

*Janusz Kacprzyk, Polish Academy of Sciences, Poland*

An overview of main trends in the research on fuzzy querying techniques, including both querying techniques for traditional databases as well as for fuzzy databases, is described.

### **Chapter III**

Introduction to Fuzzy Data Mining Methods.....	55
--	----

*Balazs Feil, University of Pannonia, Hungary*

*Janos Abonyi, University of Pannonia, Hungary*

This chapter gives a comprehensive view about the links between fuzzy logic and data mining, following nine steps of knowledge discovery. It defines and studies interesting methods, like fuzzy clustering, fuzzy classification, fuzzy association rule mining, and visualization of the results.

## **Section II Fuzzy Queries**

### **Chapter IV**

Handling Bipolar Queries in Fuzzy Information Processing.....	97
---	----

*Didier Dubois, IRIT, Université de Toulouse, France*

*Henri Prade, IRIT, Université de Toulouse, France*

Bipolar queries distinguish between negative and positive preferences in the processing of flexible queries. Negative preferences express what is more or less impossible or feasible, and they specify flexible constraints restricting feasible or tolerated values. Positive preferences are less compulsory and rather express wishes, giving measurements between indifferent and preferred values.

### **Chapter V**

From User Requirements to Evaluation Strategies of Flexible Queries in Databases .....	115
--	-----

*Noureddine Mouaddib, Université de Nantes, France*

*Guillaume Raschia, Université de Nantes, France*

*W. Amenel Voglozin, Université de Nantes, France*

*Laurent Ughetto, Université de Rennes 2, France*

This chapter studies the whole process of fuzzy querying, from the query formulation to its evaluation, proposing index structures in the evaluation of fuzzy queries. After introducing different ways for expressing flexibility in queries, the chapter reviews current methods for evaluating fuzzy queries. Finally, SAINTETIQ is presented, a data summarization model that produces a hierarchy of summaries given a relational table and additional metadata.

### **Chapter VI**

On the Versatility of Fuzzy Sets for Modeling Flexible Queries .....	143
--	-----

*P. Bosc, IRISA-ENSSAT, Université de Rennes 1, France*

*A. Hadjali, IRISA-ENSSAT, Université de Rennes 1, France*

*O. Pivert, IRISA-ENSSAT, Université de Rennes 1, France*

This work advocates the interest of extending usual Boolean queries with preferences using fuzzy sets, highlighting the expressiveness of fuzzy sets with the division operator in the context of regular databases. Some useful examples are exposed using the fuzzy query language SQLf.

## **Chapter VII**

Flexible Querying Techniques Based on CBR .....	167
---	-----

*Guy de Tré, Ghent University, Belgium*

*Marysa Demoor, Ghent University, Belgium*

*Bert Callens, Ghent University, Belgium*

*Lise Gosseye, Ghent University, Belgium*

The goal of this work is to enhance case-based reasoning (CBR) modeling a gradation in similarity of the cases. Thus, a new case is compared to previous cases in order to predict the corresponding unknown data values for the new case using possibility theory. This flexible CBR can be used to enhance flexible querying of regular databases under some conditions. Briefly, a real-world application is shown for information retrieval in a juridical database.

## **Chapter VIII**

Customizable Flexible Querying for Classical Relational Databases .....	191
---	-----

*Gloria Bordogna, CNR IDPA, Italy*

*Guiseppa Psaila, University of Bergamo, Italy*

The Soft-SQL project is presented, an extension of SQL for fuzzy queries to classic relational databases. Perhaps the most interesting characteristic is to provide tools allowing users to directly specify the context-dependent semantics of soft conditions. For example, a cheap flat in Milan does not have a similar price to a cheap flat in Tokyo.

## **Chapter IX**

Qualifying Objects in Classical Relational Database Querying .....	218
--	-----

*Cornelia Tudorie, University "Dunărea de Jos", Galati, Romania*

The author studies fuzzy queries in order to rank the resulting objects (object qualification). After a discussion on different kinds of fuzzy conditions in a fuzzy query, a new particular condition is proposed: the relative object qualification as a query selection criterion, that is, queries with two conditions in which the first one depends on the results of the second one, for example, "Retrieve the inexpensive cars among the high-speed ones."

## **Chapter X**

Evaluation of Quantified Statements Using Gradual Numbers .....	246
---	-----

*Ludovic Liétard, IRISA/IUT & IRISA/ENSSAT, France*

*Daniel Rocacher, IRISA/IUT & IRISA/ENSSAT, France*

The chapter is devoted to the evaluation of quantified statements that can be found in many applications, for example, in fuzzy querying databases. It introduces the main techniques to evaluate such statements and proposes a new theoretical background for the evaluation of quantified statements with one or two fuzzy conditions: "most of the employees are well paid" and "most of the young employees are well paid." The work shows that the context of fuzzy numbers provides some nice characteristics.

## Chapter XI

FSQL and SQLf: Towards a Standard in Fuzzy Databases .....	270
--	-----

*Angélica Urrutia, Universidad Católica del Maule, Chile*

*Leonid Tineo, Universidad Simón Bolívar, Venezuela*

*Claudia Gonzalez, Universidad Simón Bolívar, Venezuela*

The goal of this chapter is to propose a unified SQL-based language for fuzzy relational databases. The authors study the two more general approaches in this field, SQLf and FSQL. They study the characteristics and definitions of these languages, and also the current implementations based on both languages.

## Chapter XII

Hierarchical Fuzzy Sets to Query Possibilistic Databases .....	299
--	-----

*Rallou Thomopoulos, INRA, France*

*Patrice Buche, INRA, France*

*Ollivier Haemmerlé, IRIT, France*

Within the framework of flexible querying of possibilistic databases, based on the fuzzy set theory, this chapter focuses on the case where the vocabulary used both in the querying language and in the data is hierarchically organized, which occurs in systems that use ontologies. A hierarchical fuzzy set is defined as a fuzzy set whose definition domains are hierarchies. Besides this, two applications are presented.

## Chapter XIII

Query Expansion by Taxonomy .....	325
-----------------------------------	-----

*Troels Andreasen, Roskilde University, Denmark*

*Henrik Bulskov, Roskilde University, Denmark*

An overview of the use of taxonomies and ontologies in querying is presented, with a special emphasis on similarity derived from the ontology, where key concepts are organized and related. Queries can be expanded with these similarity measures, thereby causing query evaluation to be based on concepts from the ontology domain rather than on words or numbers in the query.

## Section III

### Implementation, Data Models, Fuzzy Attributes, and Applications

## Chapter XIV

How to Achieve Fuzzy Relational Databases Managing Fuzzy Data and Metadata .....	351
--	-----

*Mohamed Ali Ben Hassine, Tunis El Manar University, Tunisia*

*Amel Grissa Touzi, Tunis El Manar University, Tunisia*

*José Galindo, University of Málaga, Spain*

*Habib Ounelli, Tunis El Manar University, Tunisia*

This chapter is addressed mainly to database administrators and enterprises interested in the fuzzy capabilities in their current databases. It presents three migration approaches from real relational databases

toward fuzzy relational databases. These strategies offer different possibilities, from the possibility of fuzzy queries using the FSQL language to storing fuzzy data. Of course, each possibility poses different troubles that must be solved.

## **Chapter XV**

A Tool for Fuzzy Reasoning and Querying ..... 381

*Geraldo Xexéo, Universidade Federal do Rio de Janeiro, Brazil*

*André Braga, IBM Brazil, Brazil*

CLOUDS is a library and user interface organizing uncertainty in database systems, a tool that allows the creation of fuzzy reasoning systems over classic, nonfuzzy relational databases. It defines a fuzzy extension to SQL queries and was incorporated into a geographic information system.

## **Chapter XVI**

Data Model of FRDB with Different Data Types and PFSQL ..... 407

*Aleksandar Takači, University of Novi Sad, Serbia*

*Srdan Škrbić, University of Novi Sad, Serbia*

This chapter introduces a way to extend the relational model with mechanisms that can handle imprecise, uncertain, and inconsistent attribute values using fuzzy logic. Furthermore, a query language called PFSQL is described for this fuzzy database model, with fuzzy capabilities and the possibility to specify priorities in every simple fuzzy condition. The priorities of PFSQL are compared with the thresholds of FSQL.

## **Volume II**

### **Chapter XVII**

Towards a Fuzzy Object-Relational Database Model ..... 435

*Carlos D. Barranco, Pablo de Olavide University, Spain*

*Jesús R. Campaña, University of Granada, Spain*

*Juan M. Medina, University of Granada, Spain*

The authors introduce a fuzzy object-relational database model including fuzzy extensions of the user-defined data types and the collection types. Then they study a way to flexibly compare complex data types and an extension of collection types allowing partial membership of its elements. An application in the image-retrieval field is briefly exposed.

### **Chapter XVIII**

Relational Data, Formal Concept Analysis, and Graded Attributes..... 462

*Radim Belohlavek, Binghamton University–SUNY, USA and Palacky University,*

*Czech Republic*

Formal concept analysis with graded (fuzzy) attributes is studied. It is a particular method of analysis of fuzzy relational data, and here an overview of foundations of this formal concept analysis is presented together with concept lattices and attribute implications.

**Chapter XIX**

Fuzzy Spatial Data Types for Spatial Uncertainty Management in Databases..... 490  
*Markus Schneider, University of Florida, USA*

The author proposes some fuzzy spatial data types, introducing fuzzy points, fuzzy lines, and fuzzy regions in the two-dimensional space. This chapter also studies fuzzy topological predicates for fuzzy querying with an SQL-like spatial query language.

**Chapter XX**

Fuzzy Classification in Shipwreck Scatter Analysis..... 516  
*Yauheni Veryha, ABB Corporate Research Center, Germany*  
*Jean-Yves Blot, Portugal Institute of Archaeology, Portugal*  
*Joao Coelho, Portugal Institute of Archaeology, Portugal*

This is an application of fuzzy sets theory in the area of maritime archaeology. Specifically, the authors show how fuzzy classification using SQL is applied in shipwreck scatter analysis to obtain a user-friendly representation of the wear-type parameters of fragments of ceramics from an ancient shipwreck. This data mining method helps to classify fragments of ceramics by detecting intrinsic classes and neighborhood relations, keeping high precision of data classification in comparison to classical methods. The authors state that this framework can be relatively easily integrated with conventional relational databases, which are widely used in existing archaeological information systems.

**Chapter XXI**

Fabric Database and Fuzzy Logic Models for Evaluating Fabric Performance..... 538  
*Yan Chen, Louisiana State University, USA*  
*Graham H. Rong, Massachusetts Institute of Technology, USA*  
*Jianhua Chen, Louisiana State University, USA*

A Web-based fabric database is introduced in terms of its physical structure, software system architecture, basic and intelligent search engines, and various display methods for search results. This application uses effective fuzzy linear clustering methods to predict the fabric drape coefficient from fabric mechanical and structural properties, and the fabric tailorability with good prediction accuracy. Finally, a neuro-fuzzy computing technique for evaluating nonwoven fabric softness is presented.

**Chapter XXII**

Applying Fuzzy Data Mining to Tourism Area ..... 563  
*R. A. Carrasco, Universidad de Granada, Spain*  
*F. Araque, Universidad de Granada, Spain*  
*A. Salguero, Universidad de Granada, Spain*  
*M. A. Vila, Universidad de Granada, Spain*

This chapter proposes the use of an extension of the FSQL language for fuzzy queries as one of the techniques of data mining, which can be used to solve the problem of offering the better place for soaring given the environment conditions and customer characteristics. After doing a process of clustering and characterization, the method is able of classify new items in a cluster.



## Section IV

### Fuzzy Data Mining

#### Chapter XXIII

Fuzzy Classification on Relational Databases .....	586
--	-----

*Andreas Meier, University of Fribourg, Switzerland*

*Günter Schindler, Galexis AG, Switzerland*

*Nicolas Werro, University of Fribourg, Switzerland*

A context model with fuzzy classes is proposed to extend relational database systems. More precisely, fuzzy classes and linguistic variables and terms, together with appropriate membership functions, are added to the database schema. In order to formulate unsharp queries, the authors present the fCQL, a fuzzy classification query language, whose statements are transformed into SQL statements.

#### Chapter XXIV

Incremental Discovery of Fuzzy Functional Dependencies .....	615
--	-----

*Shyue-Liang Wang, New York Institute of Technology, USA*

*Ju-Wen Shen, Chunghwa Telecom Lab, Taiwan*

*Tzung-Pei Hong, National University of Kaohsiung, Taiwan*

Mining fuzzy functional dependencies from fuzzy databases based on similarity relations is studied, while methods are proposed to validate and incrementally search these dependencies. A detailed example is given to illustrate the process of the mining algorithm. In addition, numerical results are given to show the monotonic characteristics of the fuzzy functional dependencies.

#### Chapter XXV

Data Dependencies in Codd's Relational Model with Similarities.....	634
---	-----

*Radim Belohlavek, Binghamton University-SUNY, USA and Palacky University,*

*Czech Republic*

*Vilem Vychodil, Binghamton University-SUNY, USA and Palacky University,*

*Czech Republic*

This chapter deals with fuzzy logic extensions of the relational model that consist of adding similarity relations to domains, truth degrees attached to the table rows (ranked tables), and considers functional dependencies in these extensions. It presents a particular extension and functional dependencies in this extension that follow the principles of fuzzy logic in a narrow sense. This extension is compared to several other extensions proposed in the literature.

#### Chapter XXVI

Fuzzy Inclusion Dependencies in Fuzzy Databases .....	658
---	-----

*Awadhesh Kumar Sharma, MMM Engineering College, Gorakhpur, UP, India*

*A. Goswami, I.I.T., Kharagpur, India*

*D. K. Gupta, I.I.T., Kharagpur, India*

This chapter introduces one definition of fuzzy inclusion dependencies in fuzzy databases, a fuzzy constraint that we can see as a fuzzy foreign key between two given fuzzy relations. Inference rules on such dependencies are derived and an algorithm has been proposed for the discovery of these fuzzy inclusion dependencies.



## **Chapter XXVII**

A Distributed Algorithm for Mining Fuzzy Association Rules in Traditional Databases..... 685

*Wai-Ho Au, Microsoft Corporation, USA*

A new distributed algorithm for mining fuzzy association rules from very large databases is proposed. This algorithm has a very effective measure to distinguish interesting associations from uninteresting ones. Each site scans its own database partition to obtain the number of tuples characterized by different linguistic variables and linguistic terms. Afterward, they exchange their own local counts with all the other sites to find the global values.

## **Chapter XXVIII**

Applying Fuzzy Logic in Dynamic Causal Mining..... 706

*Yi Wang, Nottingham Trent University, UK*

This chapter applies fuzzy logic to a dynamic causal mining algorithm, which is a combination of mining rules and system dynamics for discovering causality patterns in a target system. The final goal is that fuzzy logic assists the user to make better decisions, and also assists in a better understanding of future behavior of this target system.

## **Chapter XXIX**

Fuzzy Sequential Patterns for Quantitative Data Mining..... 727

*Céline Fiot, University of Montpellier II – CNRS, France*

Sequential-pattern methods handle sequence databases, extracting frequently occurring patterns related to time and transforming large amounts of data into useful comprehensible knowledge. After introducing various fuzzy sequential-pattern approaches and the general principles they are based on, a complete framework is defined for mining fuzzy sequential patterns handling different levels of consideration of quantitative information. This framework is applied to two real databases: Web access logs and a textual database.

## **Chapter XXX**

A Machine Learning Approach to Data Cleaning in Databases and Data Warehouses..... 745

*Hamid Haidarian Shahri, University of Maryland, USA*

The data cleaning process is a duplicate elimination problem, for example, in data integration and warehousing. Here, neuro-fuzzy techniques are mixed to produce a unique adaptive framework for data cleaning, which automatically learns from and adapts to the specific notion of similarity at a meta-level. It can be utilized in the production of an intelligent tool to increase the quality and accuracy of data.

## **Chapter XXXI**

Fuzzy Decision-Tree-Based Analysis of Databases..... 760

*Malcolm J. Beynon, Cardiff University, UK*

This chapter offers a description of fuzzy decision-tree-based research, including the exposition of small and large fuzzy decision trees to demonstrate their construction and practicality. Basically, a fuzzy decision tree is a set of fuzzy if-then decision rules allowing a linguistic interpretation of the considered problem and managing the possibility for imprecision in the used data values.

## **Chapter XXXII**

Fuzzy Outranking Methods Including Fuzzy PROMETHEE .....	784
--	-----

*Malcolm J. Beynon, Cardiff University, UK*

This chapter describes the rudiments of fuzzy outranking methods, with particular attention to fuzzy PROMETHEE, a multicriteria decision-making technique using fuzzy information. Alternative fuzzy PROMETHEE approaches are described, with one used in two real-life applications. Starting with known data about a series of possible alternatives, a preference ranking of them can be achieved.

## **Chapter XXXIII**

Fuzzy Imputation Method for Database Systems.....	805
---	-----

*J. I. Peláez, University of Málaga, Spain*

*J. M. Doña, University of Málaga, Spain*

*D. La Red, National University of the Northeast, Argentina*

Missing data are often an actual problem in real data sets. Imputation is a method to fill in missing data with plausible values to produce a complete data set. This work analyzes the performance of the different traditional data imputation methods. A new fuzzy imputation approach is proposed using the ordered weighted average (OWA) operators by Yager and the majority concept.

## **Chapter XXXIV**

Intelligent Fuzzy Database Management in Multiagent Systems.....	822
--	-----

*Safiye Turgay, Abant Izzet Baysal University, Turkey*

In this chapter, an agent-based fuzzy data mining structure is defined to process and evaluate data and to build a rule structure for the system. Within the developed system, the focus was on the operation feature of the fuzzy data mining structure, which is the same for each agent composing the system. The suggested association rules are derived from a relational database.