Chapter XII
Hierarchical Fuzzy Sets to Query Possibilistic Databases

Rallou Thomopoulos  
INRA, France

Patrice Buche  
INRA, France

Ollivier Haemmerlé  
IRIT, France

ABSTRACT
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. We give an overview of previous works concerning two issues: first, flexible querying of imprecise data in the relational model and, second, the introduction of fuzziness in hierarchies. Concerning the latter point, we develop an aspect where there is a lack of study in current literature: fuzzy sets whose definition domains are hierarchies. Hence, we propose the concept of hierarchical fuzzy set and present its properties. We present its application in the MIEL flexible querying system for the querying of two imprecise relational databases, including user interfaces and experimental results.

INTRODUCTION
In flexible querying systems, fuzzy sets are used to represent preferences in selection criteria. For instance, in the framework of a database about microbiological risk assessment in foods, the users may ask for milk as a first choice or yogurt as a second choice. In possibilistic databases, an imprecise datum is represented by a possibility distribution. For instance, in some kinds of human diseases, the bacterium Escherichia coli is suspected to be responsible, but other bacteria like Listeria are not excluded. Behind those two different purposes, the same homogeneous formalism is used: the fuzzy set theory. In both cases, a relation order is defined on a domain of values. In this chapter, we study the case when the domain of values is not “flat” but hierarchically organized, using the “kind of” relation. For instance, food products, like milk or yogurt, are part of a hierarchy of substrates, in
which whole milk is a kind of milk. In the same way, the bacteria *Escherichia coli* and *Shigella* are part of a hierarchy of micro-organisms. We call a fuzzy set defined on a hierarchy a hierarchical fuzzy set (HFS). Contrary to the classical case when the domain of values is “flat,” in this case, the assumption that the values are independent does not hold. Two order relations (the preference/possibility order relation and the “kind of” relation) must be put in adequacy. Several issues thus have to be addressed:

- Does the preference/possibility degree associated with a given value in a fuzzy set have implications on the degrees associated with other values of the domain, particularly more specific or more general values?
- What would be the meaning of two comparable values (with the meaning of the “kind of” relation) associated with different preference/possibility degrees?
- Can the “kind of” relation be used to enlarge the user’s query in order to obtain more answers while respecting the preference order defined by the user in the selection criteria?

We have designed and realized two instances (for two different relational databases) of a flexible querying system, called MIEL, involving hierarchical fuzzy sets. Both databases contain imprecise data and deal with risk assessment in food, respectively, microbial risk and chemical risk.

The need for flexible querying, imprecise data representation, and studying fuzzy sets when the domain of values is hierarchically organized is justified, in both databases, by three characteristics of the data:

- Although composed of several thousand entries (10 for the microbial database and 50 for the chemical database), data are not abundant enough to answer every query and therefore there is a need for flexible querying in order to complement exact answers with pertinent answers (i.e., semantically close).
- Data include imprecise values. For instance, the level of contamination of a given food by a given contaminant is not precisely known but is included in a given interval or is inferior to a given threshold.
- Symbolic data are often organized in taxonomies: for example, taxonomies of food products (Ireland & Moller, 2000), of bacteria (Ballows, Truper, Dworkin, Harder, & Schleifer, 1992), and so forth.

The MIEL fuzzy querying system has been especially designed for end users who are not specialists of computer science. They express their query through a set of prewritten queries we call views. These views can be complemented by the users through the simple graphical user interface of the MIEL system. That interface allows the users to specify their projection attributes and their selection criteria. The taxonomies of the symbolic data can also be browsed by the end users in order to express their selection criteria as hierarchical fuzzy sets.

In this chapter, first we provide some background on the topic and recall some broad definitions useful for understanding the main focus of the chapter. Second, we define and explain the concept of hierarchical fuzzy set and compare it to the bibliography. Third, we present the MIEL flexible querying system which uses the concept of hierarchical fuzzy set. Fourth, the instantiations of the MIEL system for the querying of two imprecise databases in the field of risk assessment in food are presented and we give some experimental results. Fifth, current projects and future trends are presented and then we conclude this chapter.

**BACKGROUND**

We are concerned in this chapter with the combination of two topics: first, flexible querying of imprecise data, which includes flexible querying techniques, the representation of imprecise data, and the combination of both previous topics in the
Related Content

Framework for a Geographic Districting DSS using an Intelligent Object-oriented Model
[www.igi-global.com/article/framework-geographic-districting-dss-using/51180?camid=4v1a](www.igi-global.com/article/framework-geographic-districting-dss-using/51180?camid=4v1a)

A Study of a Generic Schema for Management of Multidatabase Systems
[www.igi-global.com/article/study-generic-schema-management-multidatabase/51169?camid=4v1a](www.igi-global.com/article/study-generic-schema-management-multidatabase/51169?camid=4v1a)

BACIIS: Biological and Chemical Information Integration System
[www.igi-global.com/article/baciis-biological-chemical-information-integration/3338?camid=4v1a](www.igi-global.com/article/baciis-biological-chemical-information-integration/3338?camid=4v1a)

Direct Perfect Hashing Functions for External Files
[www.igi-global.com/article/direct-perfect-hashing-functions-external/51084?camid=4v1a](www.igi-global.com/article/direct-perfect-hashing-functions-external/51084?camid=4v1a)