

Chapter 14

Remarks on a Fuzzy Approach to Flexible Database Querying, Its Extension and Relation to Data Mining and Summarization

Janusz Kacprzyk

Polish Academy of Sciences, Poland

Guy De Tré

Ghent University, Belgium

Sławomir Zadrozny

Polish Academy of Sciences, Poland

ABSTRACT

For an effective and efficient information search of databases, various issues should be solved. A very important one, though still usually neglected by traditional database management systems, is related to a proper representation of user preferences and intentions, and then their representation in querying languages. In many scenarios, they are not clear-cut, and often have their original form deeply rooted in natural language implying a need of flexible querying. Although the research on introducing elements of natural language into the database querying languages dates back to the late 1970s, the practical commercial solutions are still not widely available. This chapter is meant to revive the line of research in flexible querying languages based on the use of fuzzy logic. This chapter recalls details of a basic technique of flexible fuzzy querying, discusses some newest developments in this area and, moreover, shows how other relevant tasks may be implemented in the framework of such queries interface. In particular, it considers fuzzy queries with linguistic quantifiers and shows their intrinsic relation with linguistic data summarization. Moreover, the chapter mentions so called “bipolar queries” and advocates them as a next relevant breakthrough in flexible querying based on fuzzy logic and possibility theory.

DOI: 10.4018/978-1-4666-2455-9.ch014

INTRODUCTION

Databases are a crucial element of all kinds of information systems that are in turn the “backbone” of virtually all kinds of nontrivial human activities. The growing power, and falling prices of computer hardware and software, including those that have a direct impact on database technology, have implied an avalanche growth of data volume stored all over the world. That huge volume makes an effective and efficient use of information resources in databases difficult. On the other hand, the use of databases is not longer an area where database professionals are only active and, in fact, nowadays most of the users are novice. This implies a need for a proper human-computer (database) interaction which would adapt to the specifics of the human being, mainly – in our context – to the fact that for the human user the only fully natural means of articulation and communication is natural language with its inherent imprecision.

The aspects mentioned above, the importance of which has been growing over the last decades or years, have triggered many research efforts, notably related to what is generally termed *flexible* querying, and some human consistent approaches to data mining and knowledge discovery, including the use of natural language, for instance in linguistic data summarization.

Basically, the construction of a database query consists in spelling out conditions that should be met by the data sought. Very often, the meaning of these conditions is deeply rooted in natural language, i.e., their original formulation is available in the form of natural language utterances. It is then, often with difficulty, translated into mathematical formulas requested by the traditional query languages. For example, looking for a suitable house in a real estate agency database one may prefer a *cheap* one. In order to pose a query, the concept of “cheap” has to be expressed by an interval of prices. The bounds of such an interval will usually be rather difficult to assess. Thus, a tool to somehow define the notion of “being

cheap” may essentially ease the construction of a query. The same definition may be then used, in other queries referring to this concept, also in the context of other words, as, e.g., *very*. The words of this kind, interpreted as so-called *modifiers*, modify the meaning of the original concept in a way that may be assumed context-independent and expressed by a strict mathematical formula.

It seems obvious that a condition referring to such terms as “cheap”, “large” etc. should be considered, in general, to be satisfied to a *degree* rather than as satisfied or not satisfied – as it is assumed in the classical approach to database querying. Thus, the notion of the *matching degree* is one of the characteristic features of flexible fuzzy queries.

Moreover, usually, a query comprises more than just one condition. In such a case, the user may require various combinations of conditions to be met. Classically, directly only the satisfaction of *all* conditions may be required or the satisfaction of *any one* condition may be required. However, these are in fact only some extreme cases of conceivable *aggregation* requirements. For instance, a user may be completely satisfied with the data satisfying *most* of his or her conditions.

The study of modeling of such natural language terms as “cheap”, “very” or “most” for the purposes of database querying is the most important part of the agenda of the flexible fuzzy querying research.

In this paper we will present a focused overview of the main research results on the development of flexible querying techniques that are based on fuzzy set theory (Zadeh, 1965). The scope of the chapter is further limited to an overview of those techniques that aim to enhance database querying by introducing various forms of user specified fuzzy preferences (Bosc, Kraft & Petry, 2005). We will not consider other techniques that are relevant in this area, exemplified by self-correcting, navigational, cooperative, etc. querying systems.

For our purposes, we will view a *fuzzy query* as a combination of a number of imprecisely specified (fuzzy) *conditions* on attribute values

18 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/remarks-fuzzy-approach-flexible-database/73444

Related Content

ASD-BI: An Agile Methodology for Effective Integration of Data Mining in Business Intelligence Systems

Mouhib Alnoukari (2015). *Integration of Data Mining in Business Intelligence Systems* (pp. 61-82).

www.irma-international.org/chapter/asd-bi/116808

An Experimental Approach and Monitoring Tools for Evaluating a Dynamic Cubing System

Anne Tchounikine, Maryvonne Miqueland Usman Ahmed (2016). *International Journal of Data Warehousing and Mining* (pp. 1-19).

www.irma-international.org/article/an-experimental-approach-and-monitoring-tools-for-evaluating-a-dynamic-cubing-system/171117

A Heterogeneous Network Text Attribute Fusion Method Based on Multi-Level Semantic Relation Contrastive Learning

Wei Zhang and Zhonglin Ye (2025). *International Journal of Data Warehousing and Mining* (pp. 1-18).

www.irma-international.org/article/a-heterogeneous-network-text-attribute-fusion-method-based-on-multi-level-semantic-relation-contrastive-learning/378680

Mining Scientific and Technical Literature: From Knowledge Extraction to Summarization

Junsheng Zhang and Wen Zeng (2020). *Trends and Applications of Text Summarization Techniques* (pp. 61-87).

www.irma-international.org/chapter/mining-scientific-and-technical-literature/235741

Integrating Feature and Instance Selection Techniques in Opinion Mining

Zi-Hung You, Ya-Han Hu, Chih-Fong Tsai and Yen-Ming Kuo (2022). *Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines* (pp. 800-815).

www.irma-international.org/chapter/integrating-feature-and-instance-selection-techniques-in-opinion-mining/308520