

Data Engineering

June 1989, Vol. 12, No. 2

Special Issue on Imprecision in Databases

Contents

Letter to the TC Members	1
<i>L. Kerschberg (TC Chair)</i>	
Letter from the Issue Editor	2
<i>A. Motro</i>	
The Possible Approach to Handling of Imprecision in Database Systems	4
<i>H. Prade and C. Testemale</i>	
FIIS: A Fuzzy Intelligent Information System	11
<i>M. Zemankova</i>	
Representation and Access of Uncertain Relational Data	21
<i>A. Tzvieli</i>	
Incomplete Information in Logical Databases	29
<i>T. Imielinski</i>	
Approximate Retrieval: A Comparison of Information Retrieval and Database Management Systems	41
<i>C. Eastman</i>	
From Browsing to Querying	46
<i>A. D'Atri and L. Tarantino</i>	
A Trio of Database User Interfaces for Handling Vague Retrieval Requests	54
<i>A. Motro</i>	
Call for Papers	64

Letter from the Editor

Most research in database systems is based on assumptions of precision and specificity of both the data stored in the database, and the requests to retrieve data. In reality, however, both may be imprecise or vague. Considerable amount of research has focused on the issues of imprecision and vagueness in databases, and this issue of Data Engineering is devoted to this topic.

The wide variety of approaches to issues of imprecision is evidenced in part by the proliferation of the terms used to describe data and retrieval specifications which are not entirely “crisp”, including vague, uncertain, imprecise, incomplete, fuzzy, approximate, and ambiguous. Still, much of the work in this broad area falls into one or more of these three categories.

The first category includes formal models with constructs for representing imprecise data and for expressing imprecise retrieval requests. Suggested representations for imprecise data include intervals of values, “fuzzy” values (with appropriate definitions), values accompanied by certainty factors, and null values (of one kind or another). Retrieval requests for matching imprecise data often include a threshold value, that provides the degree of specificity of the request. Fuzzy values, similar to those allowed in the representation, are also allowed in queries. Note, that this category includes the considerable body of work on null values.

The second category includes systems that enable specification of imprecise queries in databases that contain only crisp data. To satisfy such queries these systems employ some kind of mechanism for determining proximity among data items of the same domain. The query language then provides constructs for specifying the desirable values that should be retrieved, and a process of “weak matching” is used to satisfy queries. This approach has been demonstrated with several experimental database systems. Note, that this category also includes various information retrieval systems (such as those intended for bibliographic searches).

The models and systems in the first two categories assume that the user has a precise notion of what he or she is looking for. The last category includes systems for dealing with retrieval requests whose imprecision is a result of vagueness of the retrieval goal itself (vagueness is in the user’s mind, so to speak). Systems in this category include various kinds of browsers, that allow users to explore the contents of the database even without specific retrieval goals, and interactive query construction aids, that assist in crystalizing vague retrieval goals into specific queries.

Four of the seven papers in this issue fall into the first category. Prade and Testemale review a general approach for handling imprecision, based on possibility theory. Zemankova describes FIIS, a knowledge-based system extended to deal with various aspects of imprecision. Tzvielli discusses various problems associated with the representation and access of uncertain data. And Imielinski discusses incompleteness in logical databases. In the second category, Eastman compares approximate retrieval in information retrieval systems and in database management systems. In the third category, D’Atri and Tarantino discuss three

styles of interaction that are suitable for users who lack sufficient knowledge to express formal queries. Finally, I review three user interfaces in the second and third categories.

As these papers demonstrate, this important area is currently very active, with many of the research projects now addressing recent technologies and the opportunities and problems they suggest. Examples include issues of fuzziness and incompleteness in knowledge-rich databases, and browsing interfaces for object-oriented databases.

Ami Motro
University of Southern California
May 1989