An Integrated Query Relaxation Approach Adopting Data Abstraction and Fuzzy Relation

Soon-Young Huh, Korea Advanced Institute of Science and Technology, Korea

Kae-Hyun Moon, Samsung Electronics Co., Korea

Jinsoo Park, Seoul National University, Korea

ABSTRACT

This paper proposes a cooperative query answering approach that relaxes query conditions to provide approximate answers by utilizing similarity relationships between data values. The proposed fuzzy abstraction hierarchy (FAH) represents a similarity relationship based on the integrated notion of data abstraction and fuzzy relations. Based on FAH, the authors develop query relaxation operators like query generalization, approximation, and specialization of a value. Compared with existing approaches, FAH supports more effective information retrieval by processing various kinds of cooperative queries through elaborate relaxation control and providing ranked query results according to fitness scores. Moreover, FAH reduces maintenance cost by decreasing the number of similarity relationships to be managed.

Keywords: *Cooperative Query Answering, Data Models, Knowledge Representation, Query Relaxation, Querying Databases*

1. INTRODUCTION

Query processing based on conventional database systems often fails to provide the information users really want if the user does not provide a precise query statement. Database systems may return null responses when the exact answers to queries do not exist. Conversely, the non-empty responses implying a qualified data set to queries may not satisfy the user who wants not only exact answers but also additional approximate answers. Furthermore, the schema and semantics of databases are often too complex for ordinary users to understand in their entirety to compose intended queries.

If a query processing system understands the schema and semantics of the database, it will be able to return informative responses beyond a query's requested answer set and greatly help the user obtain relevant answers in various decision support application systems. To support such intelligent query processing, a number of cooperative query answering approaches have been introduced, which provide a human-oriented interface to a database system by facilitating the relaxation of query conditions

DOI: 10.4018/jdm.2010100103

to produce approximate answers. Typically, cooperative query answering analyzes the intent of a query and transforms the query into a new query of greater scope by relaxing the original query conditions (Liu & Chu, 1993; Chu, Yang, Chiang, Minock, Chow, & Larson, 1996; Chu, Yang, & Chow, 1996; Chu & Chen, 1994; Liu & Chu, 2007; Cuppens & Demolombe, 1989; Cuzzocrea, 2005, 2007; De Sean & Furtado, 1998; Godfrey, 1997; Huh & Lee, 2001; Huh & Moon, 2000; Hung, Wermter, & Smith, 2004; Marshall, Chen & Madhusudan, 2005; Mao & Chu, 2007; Motro, 1988, 1990; Minker, 1998; Shin, Huh, Park, & Lee, 2008).

The cooperative query answering approach can be adopted as a key concept in various decision support application systems requiring intelligent and cooperative database access methods. A typical example application is a human resource management system shown in the prototype system in the paper. Specifically, in a knowledge-oriented consulting company having thousands of consultant resources spread globally, approximate query relaxation system will provide very effective consultant search capabilities, identifying appropriate candidate consultants having adequate domain knowledge and project engagement experiences for a project under consideration. To find an appropriate candidate for a marketing related project, a project manager might start by using vague search criteria such as major and career: "Find a marketing professional whose major is management or other similar field, and who has at least four years experience engaged in marketing project." Without intelligent assistance, the manager is likely to obtain either a null result to the query or an excess of answers that might not be sorted in any usable way. Additional examples benefiting from the cooperative querying can be found in a wide spectrum of applications ranging from geographic information systems to medical diagnostic systems where queries can be specified graphically or literally and incrementally on digital maps or symptom records, which greatly improves the querying capabilities. In the GIS, a pilot can ask an abstract query, "Find an appropriately-sized nearby airport where a Boeing 777 can land." The approximate query is translated to a distance range based on the position of the airport, and Boeing 777 is translated into the required runway conditions at the airport. The cooperative query processing systems will return relevant associative airport information such as runway condition and distance closeness with ranks. Also, in medical diagnostic systems, search conditions can be expanded for finding information on a rare illness.

To provide a wider range of approximate answers by relaxing search conditions, cooperative query answering requires a human expert's knowledge of the underlying database semantics (e.g., similarity strength between data values). A variety of knowledge representation frameworks have been researched, including the abstraction hierarchy (Cai, Cercone, & Han, 1993; Liu & Chu, 2005; Chu, Yang, Chiang, et al., 1996; Chu, Yang, & Chow, 1996; Liu & Chu, 2007; Huh & Lee, 2001; Huh & Moon, 2000; Shin et al., 2008), the semantic distance (Motro, 1988, 1990), and the logic model (De Sean & Furtado, 1998; Godfrey, 1997). However, each framework is limited for effective cooperative query answering with respect to the following requirements:

- *Diversity*. The knowledge representation framework should support users with varying levels of expertise in information retrieval. The novice user tends to write queries in simplistic forms because she does not have a good understanding of the semantics of the database schema and detailed schema knowledge, and thus often gets very limited results. If the user could express the requests in a simple but conceptual level and consequently obtain vague but richer results, the user would have a better understanding of the database contents and effectively reach the final results. Furthermore, by facilitating reduction of the query scope of conceptual queries, the user may pose queries at a more