

An Efficient Approach to Vague Joins in the Vague Query System

T. K. Dang¹, J. Küng¹, and R. Wagner¹

Abstract: This article presents problems arisen when processing complex vague joins in the Vague Query System (VQS) and introduces a new solution to efficiently solve these problems. Join operation is one of the most expensive ones in database management systems. In context of the VQS, vague joins are prohibitively expensive in terms of IO-cost and CPU-cost because they must undergo intermediate processing steps with the sheer volume of multidimensional data in multiple feature spaces.

The VQS has been introduced to deal with the problem of empty result sets in the conventional Database Management Systems (DBMSs). It extends the query facility of the conventional DBMSs with similarity search capabilities. Concretely, when available data in a conventional database do not match a user's query precisely, the corresponding DBMS will only return an empty result set to the user. This limits the applicability of the conventional DBMSs to domains where only crisp answers are meaningful. In many other application domains, however, users also expect not only the crisp results returned but also some other results that are relevant or close to the query in a certain sense. Such applications frequently appear in the real-world domains as image/multimedia processing, CAD systems, GIS, tourist information systems, time-series databases, digital libraries, information retrieval (IR), e-commerce, etc. The VQS, however, is a Flexible Query Answering System (FQAS). It has been designed to work "on top" of the conventional DBMSs in order to return to users tuples of the query relation/view that do not match the query criteria exactly.

The main features of the VQS are to employ the concept of NCR-Tables (Numeric Coordination Representation Tables) and to introduce a new query language called the Vague Query Language (VQL) that is an extended version of the SQL. NCR-Tables store (multidimensional) semantic metadata of attributes of the query relation/view. The VQL has introduced a new operator "IS" ("similar to") to formulate similar (vague) queries in the system. Unfortunately, the operator "IS" cannot be used in a join condition. An extension of the VQL had been carried out to realize vague joins in the VQS. Nevertheless, for the sake of reducing costs, users can only get the good matches, not the best ones and the new system gives no guarantee about the goodness of the results.

This paper will propose a new approach to those vague joins. The new approach not only significantly reduces the costs (I/O- and CPU-cost), but also returns to users the best match (or approximate nearest neighbors with a certain tolerant error ϵ) in the query relation/view. Experimental results will show performance of the proposed approach.

¹ Johannes Kepler University of Linz
Altenberger Strasse 69, A-4040, Linz
AUSTRIA/Europe
khanh@faw.uni-linz.ac.at, jkueng@faw.uni-linz.ac.at, and rwagner@faw.uni-linz.ac.at