



Chapter XIV

Toward a Framework for Advanced Query Processing

Suk-Chung Yoon
Widener University, USA

INTRODUCTION

Traditionally, database systems have accepted queries specified with precise search expressions directly based on primitive data stored in databases and have returned a set of database instances (extensional answers) based on primitive data as answer sets. As the size and complexity of databases with the advance in storage technologies have been increased, we believe we need much more sophisticated query-formulation and answer-finding schemes in database systems in order to satisfy the needs of a truly intelligent information system, and to serve more new applications such as e-commerce and Web-based information systems well. In certain queries, users may prefer to express queries with more general and abstract information instead of primitive terms directly based on the data stored in a database. This type of query is referred to as *flexible query*. With flexible queries, users who may not formulate precise query conditions ask general queries involving meaningful abstract terms that do not directly come from information stored in a database. For example, a manager may wish to query, "What are the *expensive digital* products which are purchased by *young* on-line shoppers in the *Midwest* area?" The query uses abstract terms, "*expensive*," "*digital*," "*young*" and "*Midwest*" which are not stored explicitly in the database. A flexible query can be expressed in terms of predefined abstract terms such as *expensive*, *digital*, *young*, and *Midwest* that can be derived from primitive information in a database. We believe flexible queries provide users with the flexibility of expressing query conditions at a relatively high-level concept, which

relaxes the requirement of the preciseness of query conditions and allows them to ask more general questions to a database.

In certain queries, users may be interested in extracting the conditions and/or the characteristics that justify extensional answers or summarize the general features of extensional answers. This type of query is referred to as intensional query. For example, a user may wish to ask a conventional query such as “Retrieve all of the customers who bought Ford Escorts in 1999” or a flexible query such as “Retrieve all of the customers who bought *compact car*.” The constant *Compact Car* is a high-level concept. In the example, simple retrieval of the names of the customers who bought Ford Escorts or compact cars is an extensional answer, whereas the answer such as “young and single people with a middle income” is an intensional answer, which describes general characteristics of the extensional answer.

The answers for the intensional queries provide a more compact, intuitive and informative form than extensional answers could ever do. Giving us exactly what conditions must be fulfilled to get a certain extensional answer, an intensional answer can be considered as a kind of interpretation or explanation of the extensional answer. A conventional query to request intensional answers is called an intensional and conventional query, while a flexible query which requests intensional answers is called an intensional and flexible query. Both types of a conventional query and a flexible query can be intensional queries. That is, both types of queries can be answered extensionally or intensionally. In our approach, queries are classified into four types: extensional and conventional query (ECQ), intensional and conventional query (ICQ), extensional and flexible query (EFQ), and intensional and flexible query (IFQ).

The current database systems support only ECQ types of queries. We believe there will be more database applications where it is important to be able to pose and answer queries flexibly rather than with expressions directly based on primitive data. In recent years, there has been an emerging area, called data mining or knowledge discovery, that addresses the problems in finding implicit, previously unknown and potentially useful patterns from large databases (Frawley, 1991). Data mining or knowledge discovery has been the subject of intense research and development. Most of the data mining technologies have been applied in the areas of decision support and market strategies. In addition to those applications, there are other applications that would benefit from the use of the data mining techniques. This motivates the development of mechanisms for processing two basic categories of query answering: flexible query answering and intensional query answering using data mining techniques.

In this chapter, we introduce a method for flexible query answering, which is a mechanism to answer queries specified with general and abstract terms quickly and intelligently. In addition, we introduce a partially automated method for generating intensional answers to represent answer-set abstractly for a conventional query and a flexible query in the database systems.

Our approach consists of two phases: preprocessing, and execution. In the preprocessing phase, we build a set of concept hierarchies (Han et al., 1993)

12 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/toward-framework-advanced-query-processing/8282

Related Content

The Impact of an ISSP on Public Service Delivery in the Digital Era

(2019). *Information Systems Strategic Planning for Public Service Delivery in the Digital Era* (pp. 342-362).

www.irma-international.org/chapter/the-impact-of-an-issp-on-public-service-delivery-in-the-digital-era/233414

Approximate Computation of Distance-Based Queries

Antonio Corraland Michael Vassilakopoulos (2005). *Spatial Databases: Technologies, Techniques and Trends* (pp. 130-154).

www.irma-international.org/chapter/approximate-computation-distance-based-queries/29662

Visual Data Mining Based on Partial Similarity Concepts

Juliusz L. Kulikowski (2009). *Semantic Mining Technologies for Multimedia Databases* (pp. 166-181).

www.irma-international.org/chapter/visual-data-mining-based-partial/28833

An Empirical Analysis of the Object-Oriented Database Concurrency Control Mechanism O2C2

David Olsenand Sudha Ram (1999). *Journal of Database Management* (pp. 14-26).

www.irma-international.org/article/empirical-analysis-object-oriented-database/51215

Privacy in Multidimensional Databases

Francesco M. Malvestutoand Marina Moscarini (2003). *Multidimensional Databases: Problems and Solutions* (pp. 310-360).

www.irma-international.org/chapter/privacy-multidimensional-databases/26973