Methodology of Schema Integration for New Database Applications: A Practitioner’s Approach

Joseph Fong
City University of Hong Kong

Kamalakar Karlapalem
Hong Kong University of Science & Technology

Qing Li
Hong Kong Polytechnic University

Irene Kwan
Hong Kong Polytechnic University

A practitioner’s approach to integrate databases and evolve them so as to support new database applications is presented. The approach consists of a joint bottom-up and top-down methodology; the bottom-up approach is taken to integrate existing database using standard schema integration techniques (B-Schema), the top-down approach is used to develop a database schema for the new applications (T-Schema). The T-Schema uses a joint functional-data analysis. The B-schema is evolved by comparing it with the generated T-schema. This facilitates an evolutionary approach to integrate existing databases to support new applications as and when needed. The mutual completeness check of the T-Schema against B-Schema derive the schema modification steps to be performed on B-Schema to meet the requirements of the new database applications. A case study is presented to illustrate the methodology.

There has been a proliferation of databases in most organizations. These databases are created and managed by the various units of the organization for their own localized applications. Thus the global view of all the data that is being stored and managed by the organization is missing. Schema integration is a technique to present such a global view of an organization’s databases. There has been a lot of work done on schema integration. Batini et al. (1986) and Özsu amd Valduriez (1991) present surveys of work in this area. But all these techniques concentrate on integrating database schemas without taking into consideration of new database applications. This paper presents a practical approach to schema integration to support new database applications by comparing the existing databases against data requirements of the new applications. If the existing databases are inadequate to support new applications, then they are evolved to support them.

In any schema integration methodology all the database schemas have to be specified using the same data model. The proposed approach uses an extended entity relationship(EER) data model. Therefore, the first step in the schema integration methodology is to translate a non-EER database schema to an EER database schema. A joint bottom-up and top-down approach for schema integration to support new database applications is proposed. The bottom-up approach is taken to integrate existing databases using standard schema integration techniques whereas the top-down approach is used to
come up with the database schema for the new applications. The top-down approach uses the joint functional-data analysis. The B-schema generated by bottom-up approach is evolved by comparing it with the T-schema generated by the top-down approach. This facilitates a streamlined approach to evolve integrated databases to support new applications.

Conventional approaches that have been widely used in database community for database design can be classified as top-down, and are therefore suitable for designing databases from scratch to support new applications. On the other hand, research in the area of heterogeneous distributed databases over the last decade has emphasized on bottom-up approaches towards global schema derivation by integrating existing databases. These two kinds of approaches are complementary in many aspects, and thus can be combined into a unified framework for schema integration.

Fong et al. (1994) developed a hierarchical comparison scheme using three major criteria for comparing relationships in two schemas. The paper classified the relationship integration by taking into account the degree of relationship, roles and structural constraints as the main features to guide the comparison of relationships. Fong (1992) applied information capacity equivalence as a measure of correctness for judging transformed schemas in schema integration. It presents a classification of common integration based on their operational goals and derive from them the instances level of equivalence of schemas after integration.

**Top-down schema design techniques**

Traditional database design has focused on data elements and their properties, and the approaches taken by database professionals were data-driven; the entire focus of the design process is placed on data and their properties (Korth and Silberschatz, 1991; Ullman, 1982; Elmasr and Navathe, 1989). Typically, a data-driven (DD) approach first creates a conceptual schema by analyzing data requirements, which is then followed by logical and physical schema design; the applications that use the database will be developed after the database is created. An alternative kind of design that has been very popular in information systems design is termed as function-driven (Senn, 1989). In these kinds of approaches, the main focus is on applications rather than data. More specifically, functional analysis starts with application requirements to generate functional schemas, which are then mapped into application specifications. These form the basis for the subsequent application program design. In functional analysis, databases are seen as isolated repositories of information used by individual activities; the vision of data as a global resource of the enterprise is not present.

More recently, the idea of applying functional analysis techniques and concepts from traditional information systems area into conceptual database design has become increasingly popular, and has resulted in so-called Joint Data- and Function-Driven (JDFD) approach, which is more powerful than pure DD approach (Batini et al., 1986). As shown in Figure 1, JDFD produces the conceptual database structure and the function schema in parallel, so that the two design processes influence each other. More specially, the JDFD approach makes it possible to test whether data and function schemas are mutually consistent and complete. Note that both pure DD and JDFD types of approaches are used for designing new databases to support new applications.

**Bottom-up schema integration techniques**

Schema integration is a relatively recent problem that has appeared in the context of distributed, heterogeneous databases (Sheth and Larson, 1990; McLoed and Heimbigner, 1980). It takes place in a bottom-up fashion, requiring that an integrated global schema be designed from the local schemas, which refer to existing databases. Figure 2 summarizes the schema integration activity which has as input the local schemas and local transactions, and has as output the global schema as well as the specifications of data and query-mapping from global to local databases.

Though different researchers have proposed different solution procedures for conducting schema integration, they can be eventually considered to involve a mixture of the following activities: pre-integration, comparison, conforma-
Related Content

Comprehension of Hierarchical ER Diagrams Compared to Flat ER Diagrams
www.igi-global.com/chapter/comprehension-hierarchical-diagrams-compared-flat/23017?camid=4v1a

Towards Google Earth: A History of Earth Geography
www.igi-global.com/chapter/towards-google-earth/7923?camid=4v1a

NetCube: Fast, Approximate Database Queries Using Bayesian Networks
www.igi-global.com/chapter/netcube-fast-approximate-database-queries/8017?camid=4v1a

OO and EER Conceptual Schemas: A Comparison of User Comprehension
www.igi-global.com/article/eer-conceptual-schemas/51140?camid=4v1a