Semantic Heterogeneity in Multidatabase Systems: A Review and a Proposed Meta-Data Structure

Te-Wei Wang, Florida International University, USA
Kenneth E. Murphy, Willamette University, USA

ABSTRACT

This article reviews current research on the issue of semantic conflict resolution in multidatabase system design. It is observed that in multidatabase systems, semantic conflicts need to be resolved at both schema level and instance level. Based on the literature review, a new taxonomy for differentiating semantic conflicts and a meta-data representation incorporating the taxonomy are proposed. It is argued that the new meta-data representation is effective for summarizing local schemata, and hence it can serve as a common protocol for multidatabase systems that require instance level conflict resolution.

Keywords: meta-data representation; multidatabase; semantic conflict resolution; semantic heterogeneity.

INTRODUCTION

Semantic heterogeneity or semantic conflict is the main source of problems in multidatabase design. In this article, a brief review of previous work in semantic conflict identification is presented which leads to the creation of a taxonomy for resolving conflicts in multidatabase design that is more inclusive when compared to existing frameworks, for example that of Batini et al. (1986). A meta-data structure, based on this taxonomy, will be proposed that can be used as a point of reference (a common protocol) for semantic conflict resolution.

For last three decades, a significant quantity of multidatabase research has focused on resolving the problem of semantic heterogeneity or semantic conflicts. Semantic heterogeneity is often present in multidatabase systems because of the lack of global schema definition. The situation is similar to common misunderstandings that occur in everyday interpersonal communication. Misunderstandings can result from two people who speak different languages. They cannot understand one another un-
less interpreters are present. Even when interpreters are used, concepts that cannot be precisely translated remain. In fact, the level of shared understanding between the parties after communication depends heavily on the knowledge of the interpreters. Even if persons participating in the conversation are speaking the same language, misunderstandings could persist due to the ambiguity of language. Based on this analogy, it is apparent that not all semantic conflicts can be systematically resolved. A good conflict resolution system should have the intelligence to separate resolvable and irresolvable conflicts. Given such an intelligent conflict resolution process, corresponding procedures can be created to integrate results from different data sources and to report inconsistencies that cannot be resolved.

To create multidatabases without semantic conflicts, a significant amount of research has focused on schema integration at the conceptual level (Lim & Chiang, 2000). Real world examples such as the Cyc knowledge base in Carnot (Collet, Huhns & Shen, 1991; Singh et al., 1997) and the CORDS multidatabase (Martin & Powley, 1997) all use similar schema-integration concepts to provide multidatabase systems with an integrated view at the logical level. However, in practice, semantic conflicts exist not only at the logical or conceptual level, but also at the instance or run-time level. That is, in practice, many conflict resolutions may need to be performed at query run-time. To facilitate run-time semantic conflict resolution, the integration engine should have the ability to construct consistent meta-data at run-time. In this article, a meta-data structure for the purpose of capturing the run-time meta-data generated by such a run-time integration problem is proposed. We organize this article by addressing the following questions in sequence:

- What is a multidatabase system?
- What are the methods currently used in multidatabase systems to resolve semantic heterogeneity?
- What are different types of semantic heterogeneity?
- Can a “better” taxonomy for classifying semantic heterogeneity be found, resulting in a meta-data structure to assist in addressing semantic conflicts?
- Is the meta-data structure proposed sufficient for practice?

WHAT IS A MULTIDATABASE SYSTEM?

Database systems are a major resource in most corporations today. Organizations have been buying state-of-the-art software platforms on which database systems are built. Typically, a database system is designed to address an organization’s needs at a fixed point in time. Organizations, however, have information needs that are dynamic. The original design of a particular database system could soon be and often is quickly outdated. This situation is inevitable in fast growing or multi-site organizations where different subunits have developed their own database systems. When the greater organization requires integrated data, problems emerge because database systems cannot “talk” or query one another directly. A mediator is required to coordinate the communication and/or data exchange process. The design and implementation of such a mediator is an essential component of so-called multidatabase systems.

A multidatabase system provides integrated access to heterogeneous, autonomous local databases (Bright, Hurson &
Related Content

Matching Attributes across Overlapping Heterogeneous Data Sources Using Mutual Information
Huimin Zhao (2010). *Journal of Database Management* (pp. 91-110).
[www.irma-international.org/article/matching-attributes-across-overlapping-heterogeneous/47421/](www.irma-international.org/article/matching-attributes-across-overlapping-heterogeneous/47421/)

The Impact of Conceptual Data Models on End-User Performance
[www.irma-international.org/article/impact-conceptual-data-models-end/51109/](www.irma-international.org/article/impact-conceptual-data-models-end/51109/)

Compression Schemes with Data Reordering for Ordered Data

Excess Entropy in Computer Systems
[www.irma-international.org/chapter/excess-entropy-in-computer-systems/85465/](www.irma-international.org/chapter/excess-entropy-in-computer-systems/85465/)

Component-Based Generalized Database Index Model
[www.irma-international.org/chapter/component-based-generalized-database-index/11127/](www.irma-international.org/chapter/component-based-generalized-database-index/11127/)