A Global Object Model for Accommodating Instance Heterogeneities

Ee-Peng Lim\textsuperscript{1} and Roger H.L. Chiang\textsuperscript{2}

\textsuperscript{1} Centre for Advanced Information Systems
School of Applied Science, Nanyang Technological University
Singapore 639798, SINGAPORE

\textsuperscript{2} Information Management Research Centre
Nanyang Business School, Nanyang Technological University
Singapore 639798, SINGAPORE

Abstract. To completely address database integration problems in the context of multidatabase\textsuperscript{[10]} and data warehousing systems, one has to examine various integration and query requirements. Due to various reasons such as poor data quality in local databases, ongoing local database updates, and instance heterogeneities, some instance differences have to be accommodated by the integrated databases. We have therefore proposed a new object-oriented global data model, called $OO_{RA}$, that can accommodate attribute and relationship instance heterogeneities in the integrated database. In addition, the $OO_{RA}$ model has been designed to allow database integrators and end users to query both the local and resolved instance values using the same query language.

1 Introduction

To fully address the schema and instance integration issues in both multidatabase and data warehousing systems, one has to examine the database integration process at the macro level. Throughout the entire database integration process, inter-database heterogeneities should be handled appropriately. While there has not been a well accepted database integration methodology, we proposed to divide the entire integration process into three phases, namely Analysis, Derivation, and Evolution.

- **Analysis:** Analysis is essentially a knowledge acquisition phase. In this phase, database integrators are expected to understand pre-existing databases at both the conceptual and implementation levels. Database integrators are also required to find out from the integrated database users their global application requirements in order to derive the global schema and instances.

- **Derivation:** The actual derivation of global schema and integrated instances is done in this phase. Once the derivation is done, queries on the integrated database can be evaluated. It is in this phase a complete mapping from local schemas to the global schema, as well as a mapping from local instances to global instances are specified.

- **Evolution:** Due to the autonomy of local database systems, updates to the local databases may violate the mapping from local instances to global
instances. Evolution therefore refers to the ongoing refinement of integrated
databases as the local database schemas and instances evolve. It becomes
the most important phase to maintain a multidatabase or a data warehouse.

Among the above three phases, evolution has been largely ignored in the
database integration research primarily due to two reasons. Firstly, most re-
searchers focus on schema integration issues\cite{13,4,11,14}. While a lot of schema
integration issues have to be investigated for different databases during the
derivation phase, it is uncommon to investigate schema integration issues during
the evolution phase due to rare modification to pre-existing local schemas. Sec-
ondly, research on the integration of instances has been pre-occupied by query
processing issues instead of local database updates during the evolution phase.
In this paper, we argue that instance integration may not be complete in the
derivation phase. During the evolution phase, one also has to consider local
database updates which lead to new instance conflicts that cannot be handled
by pre-defined integration methods. Hence, new global data models that can
accommodates instance heterogeneities become necessary.

Literature Review

Most previous database integration research focused on resolving schema con-
flicts. Lately, as researchers begin to address instance integration problems, sev-
eral solutions of instance conflict resolution have been proposed \cite{7,8,13}. We
review some data modeling research in handling instance conflicts as follows.

- **Polygen model** \cite{13} was proposed to capture source information of attribute
  values that come from different local relations. A source value is associated
  with every attribute value of the tuples of polygen relations. The source
  information captured include the sites the attributes originated from and the
  intermediate sites at which they are processed. The model, however, does not
  provide the mechanism to accommodate or resolve instance heterogeneities.

- **TS Relational model** \cite{5} was proposed to accommodate entity and attribute
  conflicts in a relational integrated database. A special source attribute is
  assigned to every relations. An extended relational algebra has been proposed
  to manipulate the TS relations. Like the Polygen model, TS Relational model
  is not designed to represent resolved instance values.

- **Role-based Multidatabase model** \cite{9} considered the different roles (or rela-
tions) assumed by real-world objects. Queries on a role-based multidatabase
  are decomposed into queries on different combinations of roles. Apart from
  not handling resolved instance values, the role-based multidatabase model
  does not classify between tolerable and intolerable relationship and attribute
  value conflicts (see Sect. 2).

Research Objectives

Our research addresses the problem of accommodating instance heterogeneities
(conflicts) in the global data model adopted for integrated databases. There are
a number of reasons for accommodating instance heterogeneities: