

Schema Mapping for Interoperability in XML-Based Multidatabase Systems*

Li Ruixuan Lu Zhengding Xiao Weijun Li Bing Wu Wei

College of Computer Science and Technology, Huazhong University of Science and Technology
(HUST), Wuhan, 430074, P. R. China

rxli@public.wh.hb.cn

Abstract

The main aim of a multidatabase system (MDBS) is to achieve data integration and interoperability among distributed and heterogeneous database systems. But data model heterogeneity and schema heterogeneity make this a challenging task. Multidatabase users can only view the global schemas whose real data come from local database systems. Thus, mappings from global schemas to local schemas should be established. At the same time, the recent emergence of XML has shown great attractability as a new standard for data representation and exchange on the web. Data model based on XML becomes more suitable for integrating different types of systems. This paper firstly introduces a multidatabase common data model based on XML, named XML-based Integrated Data Model (XIDM). Then, an approach of schema mappings based on XIDM in MDBSs has been presented. Finally, the illustration and implementation of schema mappings in a multidatabase prototype – Panorama are also discussed.

Key words: Interoperability, Schema mapping, Integrated data model, Multidatabase systems, eXtensible markup language (XML)

1. Introduction

Nowadays, due to the increasing development of web data management and collaborative work, the traditional single database system cannot satisfy the requirements of data sharing and interoperation in web environment. At the same time, the existing databases cannot be discarded totally, so developing some collaborative software that can access and process the data of multiple pre-existing databases is a necessary trend. A multidatabase system (MDBS) is a layer of software that can integrate a collection of pre-existing, heterogeneous, distributed database systems called local database systems (LDBSs) [1]. It mainly solves the problem that how to achieve the schema integration and data interoperability among multiple LDBSs. Due to the data model heterogeneity and schema heterogeneity, there may exist kinds of

differences and conflicts in different local database schemas. In order to access the data in these multiple database systems transparently, an approach that can reduce these differences and resolve these conflicts must be needed [2]. A general way of integrating these schemas is through a global schema that is constructed by schema transformation from local schemas of participating database systems. Global schema is a virtual knowledge base in MDBSs. Multidatabase users can only query the global schemas whose data come from local database systems. So, schema mappings from global schemas to local schemas should be employed.

To represent and deal with multidatabase export schemas and global schemas, common data model is needed. In the old fashion multidatabase systems, such as Pegasus [3], Interbase [4], CORDS [5], MIND [6], relational data model, object-oriented model and object relational model played the most important roles. The growing popularity of the web has increased the need to search, display, manipulate and exchange information among different data sources, including database systems and file systems. An attempt to achieve this through data standardizing representation gives rise to the advent of XML (eXtensible Markup Language) techniques. As a data description language, XML can describe not only structured data, but semi-structured data. Thus, a common data model based on XML is a better selection for MDBSs.

Researches on data models and heterogeneous information integration over the last several years have focused on handling semi-structured data, mapping between XML and relational databases, and querying web data or XML models, such as OEM model and LoreL query language in Lore [7], YAT XML algebra and YATL query language in YAT system [8], XML data model and XMAS query language in MIX mediator system [9], X-Ray approach for integrating XML with relational database systems [10]. These researches have made great contributions to modeling semi-structured or XML data, storing and managing semi-structured data, wrapping relational data sources as XML views, and

* This work was supported by National High Performance Computing Foundation of China under grant 99319.