Sangam - A Solution to Support Multiple Data Models, Their Mappings and Maintenance *

Kajal T. Claypool†, Elke A. Rundensteiner†, Xin Zhang†, Su Hong†, Harumi Kuno‡, Wang-chien Lee§, and Gail Mitchell§

(†)Department of Computer Science
Worcester Polytechnic Institute
Worcester, MA 01609-2280
{kajal | rundenst | xinz | suhong}@cs.wpi.edu

(‡)Hewlett-Packard Company
1501 Page Mill Road
Palo Alto, CA 94304
harumi_kuno@hp.com

(§) Verizon Communications
40 Sylvan Rd.
Waltham, MA 02451
{wang-chien.lee | gmitchell}@verizon.com

1. OVERVIEW

Networked environments like the Internet have catalyzed a phenomenal growth in the publication of data bringing with it an increasing need to share and integrate information. Alas, the format of data to be integrated varies from company to company and sometimes from person to person. To accomplish tasks such as data sharing, exchange, and integration, we may need to map an XML schema to a relational schema to drive transformation of XML elements into relational data or map an XML schema of one application to that of another, or a query posed against a high-level semantic model may need to be mapped into an equivalent query posed against a logical database schema.

In this project, we present a novel model management system, Sangam [1] that provides a complete and integrated solution to the large problem, as well as an infra-structure for creating advanced tools and operators. While previous work has dealt with meta-data management of data models, the key thrust of our work now lies in the *modeling of maps* in a model management system. A map is represented in Sangam as first-class citizens thereby allowing users to operate and manipulate them via a set of generic tools and operators. Our system is capable of (1) describing different data models such as relational, OO and XML; (2) describing cross-model mappings to map, for example, a XML schema to a relational schema (see Figure 1); (3) describing re-structuring within one model, for example within XML documents; and (4) discovering mappings with the aid of pre-defined maps and additional domain knowledge between two given schemas to aid e-business to communicate with their own individual XML documents. As a proof of concept we provide (1) a generation tool that can automate mappings between application schemas and the subsequent mapping between the application data; and (2) a set of generic modification operators to reflect schema changes in the source and/or the target.

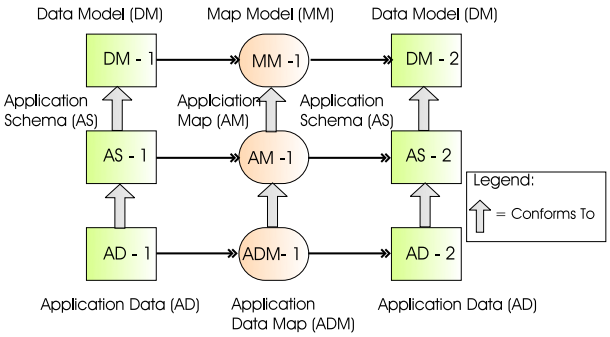

Figure 1: Four-Tier Meta Model Architecture for Data Models and the Map Models.

2. DEMONSTRATION

Our Sangam prototype is developed using Java JDK 1.2, Oracle8i (with object extensions), and IBM XML Parser 4J. Our examples for the demonstration focus on mapping XML to relational model and vice versa; and on re-structuring within the XML and relational model. We will demonstrate:

- importing DTD or relational schema as meta-data;
- translating DTD to a relational schema using pre-defined maps;
- re-structuring DTD [2] or relational schemas using pre-defined maps;
- discovering maps between two DTDs using domain knowledge;
- generating code for the map to guide data transformation;
- propagating schema change from source to target and vice versa by allowing in-place modification of the map.

3. REFERENCES

[1] K. T. Claypool and E. A. Rundensteiner. Sangam: A Model Management System. Technical Report WPI-CS-TR-01-15, Worcester Polytechnic Institute, March 2001.

[2] X. Zhang, W.-C. Lee, G. Mitchell, and E. A. Rundensteiner. Clock: Synchronizing Internal Relational Storage with External XML Documents. In *ICDE-RIDE 2001*, 2001.

*This work was supported in part by the NSF NYI grant #IIS-97-29878, #IIS-97-96264 and #IIS-99-88776, IBM and Verizon Laboratories Inc..

ACM SIGMOD 2001 May 21-24, Santa Barbara, California USA