Chapter IV

XML Schema Integration to Facilitate E-Commerce

Louise Lane and Kalpdrum Passi
Laurentian University Sudbury, Canada

Sanjay Madria
University of Missouri, USA

Mukesh Mohania
IBM India Research Lab, India

ABSTRACT

XML has become the de facto standard for Information Exchange protocol for e-commerce and many workgroup applications such as Enterprise Resource Planning (ERP). The availability of large amounts of heterogeneous distributed web data necessitates the integration of XML data from multiple XML sources for many reasons. Currently, there are many e-commerce companies, which sell similar products but represent them using different XML schemas with possibly different ontologies. When any two such companies merge, there is a need for a uniform schema integration methodology. In some applications like comparison-shopping, there is a need for an illusionary centralized homogeneous information system. In this chapter, we propose an XML Schema integration methodology. We define an object-oriented data model called XSDM (XML Schema Data Model) and present a graphical representation of XML Schema for the purpose of schema integration. We use a three-layered architecture for XML Schema integration, with each layer presenting an integrated view of the concepts that characterize the layer below. The three layers included are namely pre-integration, comparison and integration. During pre-integration, an analysis of the schemas to be integrated occurs. During the comparison phase of integration, correspondences as well as
conflicts between elements are identified. During the integration phase, restructuring and merging of the initial schemas takes place to obtain the global schema. We define integration policies for integrating element definitions as well as their data types and attributes. The policies are also applicable in integrating DTD schemas with other DTD/XML Schemas.

INTRODUCTION

XML (eXtensible Markup Language) is used to describe semi-structured data, irregular or incomplete data whose structure may be subject to unpredictable changes. Unlike traditional semi-structured data, XML documents are self-describing, thus XML provides a platform-independent means to describe data and therefore, can transport data from one platform to another (Bray et al., 1998). XML documents can be both created and used by applications. The valid content, allowed structure and metadata properties of XML documents are described by their related schema(s) (Thompson et al., 2001). An XML document is said to be valid if it conforms to its related schema. A schema also gives additional semantic meaning to the data it is used to tag. The schema is provided independently of the data it describes. Any given data set may rely on multiple schemas for validation. Any given schema may itself refer to multiple schemas.

In ecommerce, XML documents can be used to publish everything from product catalogs and airline schedules to stock reports and bank statements. XML forms can be used to place orders, make reservations and schedule shipments. XML eliminates the need for custom interfaces with every customer and supplier, allowing buyers to compare products across many vendors and catalog formats, and sellers to publish their catalog information once to reach many potential buyers. XML can also enable online businesses to build on one another’s published content and services to create innovative virtual companies, markets and trading communities. With a global view of the Internet-wide shopping directories, a query system can locate all merchants carrying a specific product or service, and then query each local schema in parallel to locate the best deals. The query system can sort the offers according to criteria set by the buyers—the cheapest flight, the roomiest aircraft, or some weighted combination. The traditional method used for B2B information exchange is through Electronic Data Interchange (EDI), which is complex, expensive and necessitates a custom integration solution between each pair of trading partners. A query based system that uses XML as the common format to enterprise integration is simpler and more open than traditional EDI as it eliminates the proprietary message formats used by each company. Creating virtual enterprises by enabling businesses to build on one another’s services involves modeling enterprises as collections of services, some internal to a particular business, others
Related Content

Integration of Libre Software Applications to Create a Collaborative Work Platform for Researchers at GET
[www.igi-global.com/article/integration-libre-software-applications-create/2609?camid=4v1a](www.igi-global.com/article/integration-libre-software-applications-create/2609?camid=4v1a)

Virtual Telemedicine Using Natural Language Processing
[www.igi-global.com/article/virtual-telemedicine-using-natural-language/41727?camid=4v1a](www.igi-global.com/article/virtual-telemedicine-using-natural-language/41727?camid=4v1a)

Whose Questionnaire is It, Anyway?
[www.igi-global.com/article/whose-questionnaire-anyway/40341?camid=4v1a](www.igi-global.com/article/whose-questionnaire-anyway/40341?camid=4v1a)
Enhancing Web Data Mining: The Study of Factor Analysis
www.igi-global.com/chapter/enhancing-web-data-mining/162891?camid=4v1a