

Evolving a “Wise” integration system for e-commerce transactions

Chima Adiele^{a,*}, Sylvanus A. Ehikioya^b

^a *Department of Computer Science, University of Lethbridge, Lethbridge AB, Canada T1K 3M4*

^b *Nigerian Communications Commission, Plot 72, Ahmadu Bello Way, Abuja, Nigeria*

Received 2 December 2003; received in revised form 13 February 2006; accepted 29 August 2006

Available online 5 October 2006

Abstract

The Web facilitates a global marketplace that provides an economic platform for application developers, merchants and customers to exchange goods and services from a wide range of domains. As a result, large volume of data now resides on the Web. Access to data distributed over the Web is becoming increasingly difficult because of information overload. A system that transcends the amalgamation of data, and provides easy access to data distributed over the Web is necessary. We present WISE, a correctness preserving approach to integration of Web data sources. We describe components of WISE, including a flexible semistructured data model, a common-term vocabulary, and an efficient integration algorithm that automates the integration process. We formally specify these components and show that the global integrated schema is correct, complete, minimal, and understandable.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Data semantics; Data integration; E-commerce; Formal specification

1. Introduction

The Internet has grown rapidly due mainly to the advent of the World Wide Web (Web) and increased number of Web-centric applications, such as e-commerce. E-commerce involves business activities that deal with exchange of values for goods and services where all transactions are carried out electronically. E-commerce creates a huge economic potential that is feasible because of the seamless global access that the Internet provides, the open environment the Internet supports, and the user-friendly graphical interfaces that Web tools support. Consequently, a large amount of data now resides on the Web. Access to data distributed over the Web is becoming increasingly difficult, as identifying semantically equivalent data remains a problem. The heterogeneity of e-commerce data sets makes it difficult to locate data relevant to a query or relate information from different sources. A system that transcends the amalgamation of data, and provides easy access to

e-commerce related data on the Web is desirable. Data integration provides a global view of data across organizational boundaries, while retaining the autonomy of the data sources.

1.1. Running example

Airline-A and Airline-B are two firms in the airline industry that have a service collaboration agreement with respect to baggage handling devices in their areas of operations. Airline-A has an XML database containing information about baggage handling devices. Airline-B has a relational database also containing information about baggage handling devices. With this service agreement, Airline-A and Airline-B will need to share information. However, the conflicts that exist between the underlying data sources impede the exchange of information.

Observe the conflicts that exist (both structural and semantics) between the two databases in Fig. 1. One of the main challenges for e-commerce infrastructure designers is information sharing among business partners. Information sharing involves retrieving data located in different sources thereby obtaining a global view to overcome con-

* Corresponding author.

E-mail addresses: Chima.adiele@uleth.ca (C. Adiele), ehikioya@ncc.gov.ng (S.A. Ehikioya).

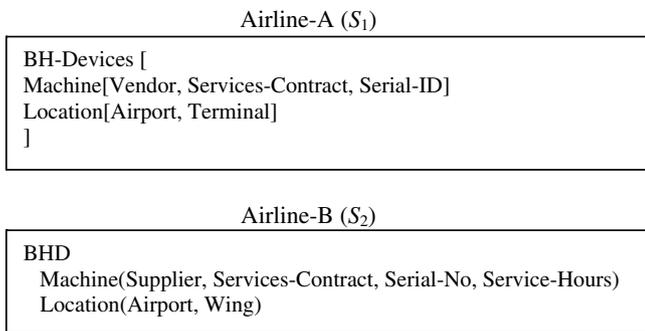


Fig. 1. Example with two source schemas.

flicts. Data integration provides a global view of data across data sources.

The problem addressed in this paper is to integrate data from Web data sources so that they can be useful for e-commerce transactions. Web data emanate from data sources that are inherently distributed and potentially heterogeneous. On one hand, enterprises tend to represent data using disparate data sources. On the other hand, users want to access such data without understanding the intricacies of the underlying data sources. As a result, interoperability problems exist in the e-commerce environment. Managing data in the e-commerce environment is becoming increasingly difficult as participants confront new business opportunities and pressures such as mergers, acquisitions, supply-chain management, and new service offerings [16]. The integration problem deals with a situation where data from two or more heterogeneous data sources are integrated. The aim of an integration system for e-commerce is to make data available and easily accessible among all participants. We need to integrate Web data sources to free the user from the responsibility of locating and interpreting e-commerce data from the broad spectrum of information that exist on the Web. This kind of environment brings together all the participants in an e-commerce transaction and hence creates more opportunities to exchange information. An e-commerce integration system provides a platform where e-commerce data from different sources can be accessed using a global schema, thus taking away the burden of locating and interpreting e-commerce data from the user. The global schema provides a virtual database that enables the user to query multiple data sources that are inherently distributed without knowing the complexity of the underlying data sources.

We present a high-level model of a schema integration system, hereinafter referred to as WISE (an acronym for Web Integration System for e-commerce). We describe the components of WISE, including a flexible semistructured data model, and an efficient integration algorithm that automates the integration process. We formally specify these components and show that the integrated schema: (1) is a correct representation of the local schemas; (2) completely represents the local schemas as all the elements in the local schemas are also in the global schema; (3) is a minimal representation of the local schemas as each unique

element in the local schema is defined not more than once in the global schema; and (4) is understandable because it is represented using easily readable and simple XML format.

This paper makes the following contributions:

- We present a correctness preserving integration system that eliminates the traditional ad hoc approach to integration. By providing a unified formalism that specifies components of an integration system, we reduce systems' complexities and offer a thorough understanding of inter- and intra-component behaviours. Our approach improves the design, and eventually, the deployment of successful integration systems.
- The design of an efficient graph-based integration algorithm that combines a flexible data model and the common-term vocabulary (*CTV*) to achieve automation takes away the burden of identifying semantic relationships in correspondence assertions¹ from the users. The goal of our methodology is to present an integration system that achieves automation, thus relieving the user from the burden of locating and interpreting e-commerce data, and reduces likely sources of error.
- Showing that the global integrated system is correct, complete, minimal and understandable enhances the quality and reliability of the integration system.

1.2. Organization

The remaining part of this paper is structured as follows. Section 2 reviews related work, thus laying the basic foundation that sets this work in perspective. Section 3 discusses data semantics. Here, we examine the characteristics of e-commerce data set that influence design decisions of an integration system. In addition, we describe a flexible semistructured data model, *Del-G*, and the *CTV*, that make explicit the content of data. In Section 4, we present a formal description of WISE and its constituent components, including an efficient graph-based integration algorithm. We show that WISE is capable of integrating Web data sources without user intervention, and hence prove its correctness, completeness and minimality. We provide integration examples in Section 5 to illustrate how our model works. Finally, we conclude in Section 6 and provide insight into future work.

2. Related work

The Stanford-IBM Manager of Multiple Information Sources (TSIMMIS) [4] focuses on the development of tools to facilitate integration of heterogeneous information sources in different structures. TSIMMIS does not believe in automation. Automating schema integration process

¹ A correspondence assertion is a statement that establishes the relationship between the semantics of objects in two different schemas.

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات