TOURISM INFORMATION AGGREGATION USING AN ONTOLOGY BASED APPROACH

Miguel Gouveia and Jorge Cardoso

Department of Mathematics and Engineering,Universaty of Madeira, Tecnopolo, Funchal, Madeira

Abstract: Aggregating related information, from different data sources, allows the creation of data repositories with very useful information. In the tourism domain, aggregating tourism products with related tourism attractions will add value to those products. The ability to create dynamic packages is another reason to aggregate tourism information. Defining an ontology, composed by the concepts to aggregate, is the first step to create tourism aggregation systems. In this paper we define the approach and the architecture that guides to the creation of aggregated solutions that provide valued tourism information and that allow the creation of dynamic packages.

1 INTRODUCTION

The Web became a large repository where one can get information of all kinds. Some enterprises embrace this opportunity and create large data repositories. Enterprises like ChoicePoint, Experian, LexisNexis or Acxiom are some examples. They sell aggregated data that can help other enterprises to manage their business. Information aggregation like customer preferences, product prices and market tendencies can help enterprises manage the risk and reward of commercial and financial decisions. ChoicePoint, one of the top companies selling information, sells to more than half of America's top 1000 companies.

The idea of data aggregation has being applied to vast business areas, we believe that it will also have a big success in the tourism industry. The tourism domain is characterized by a significant heterogeneity market and information sources and by a high volume on online transactions (Werthner and Klein, 2004). Nowadays, there is a lot of information about tourism products throughout the Internet and other systems. There are systems that offer information about a set of tourism products types like airlines, hotels and car rental. In this group of systems we have the Computerized Reservation Systems (CRS) that are associated to a specific travel supplier and the Global Distribution System (GDS) that is a super switch connecting several

CRSs (Cardoso, 2005). From the Hotel Distribution Systems (HDS) we can get information about hotels. There are also the Destination Management Systems (DMS) that provide information about tourist regions. Besides these sets of systems, there are many web sites that offer tourism information that aren't assessable through any of the enumerated systems. Web sites about hotels that belong to small companies, car rental, golf or information about tourist regions are just some examples.

Besides the tourism information aggregation, one of the big challenges in the tourism business is ability to create dynamic packages. Dynamic package means putting together, in real time, a package of several major travel components, e.g., air flight legs, hotel nights, car rental days, etc (Kabbaj, 2003). It provides a single, fully priced package, requiring only one payment from the consumer and hiding the pricing of individual components within 5-15 seconds (Fitzgerald, 2005).

Current dynamic package applications are developed using a hard-coded approach. However such an approach for integration does not comply with the highly dynamic and decentralized nature of the tourism industry. Most of the players are small or medium-sized enterprises with information systems with different scopes, technologies, architectures and information structures. This diversity makes the interoperability of information systems and technologies very complex and constitutes a major barrier for emerging e-

marketplaces and dynamic applications that particularly affects the smaller players (Fodor and Werthner, 2004).

In this paper we will describe an architecture to aggregate tourism information in order to provide the creation of dynamic packages.

2 SEMANTIC TECHNOLOGY IN INFORMATION AGGREGATION

The process of information aggregation is not easy. Currently Europe's corporations spend over 10 billion Euros in dealing with data integration problems (Alexiev, Breu, Bruijn, Fensel, Lara, and Lausen, 2005). Companies are spending 10% to 30% of their IT budgets on integrating applications and systems internally and with their partners. The problem with information aggregation is that the information is not structured in the same manner. Each data source, or application, has a different data representation and provides different data formats for integration. HTML, XML, flat files, relational model are some of examples that we can find in an aggregation problem. Another problem is the semantic differences between data sources. We can find the same word with different meanings. For example, in one data source, customer can refer to the tourists in others it can refer to the travel agencies.

To resolve the information aggregation problem, many technologies were proposed. Database and application server vendors offer comprehensive data integration tools and platforms. However, they do not provide any support for assuring semantic coherence and consistency of the results (Alexiev, Breu, Bruijn, Fensel, Lara, and Lausen, 2005). Using ontologies and data mapping technologies, is it possible to resolve the semantic incoherence. Ontologies aim at capturing static domain knowledge in a generic way and provide a common agreement upon the understanding of that domain (Chandrasekaran, Josephson, and Benjamins, 1999).

3 ONTOLOGY BASED APPROACH

Information aggregation can remit us for two integration approaches. In the first approach we can start by selecting the data sources to integrate and then try to create an ontology, based on the metadata from the data sources to integrate. In this approach we can follow the Semantic Information Management Methodology (SIM) (figure 1).
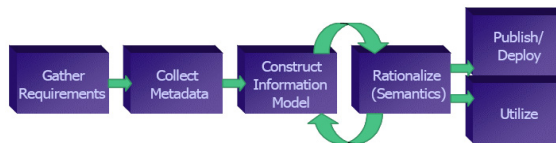

Figure 1: The SIM Methodology.

In SIM methodology first we collect the metadata of the existing data sources. Then, using this metadata, a central ontology is created capturing the meaning of the data presented in these data sources. Finally, the disparate data schemas are mapped to the ontology.

The second approach resides in thinking first in the information that we want to aggregate and create the ontology in order to create a useful knowledge base. We call this the Ontology Based Approach (figure 2). In this approach, the ontology is defined not based on existing data sources metadata but, instead, based on the solution that we want to build.

In the Ontology Based Approach, we begin with the ontology definition. Based on the defined ontology, we create the data schema that will be used to integrate all the data sources. Then, the data schema for integration is mapped to the Ontology. Finally, we must search data sources that provide the instances to populate the ontology.
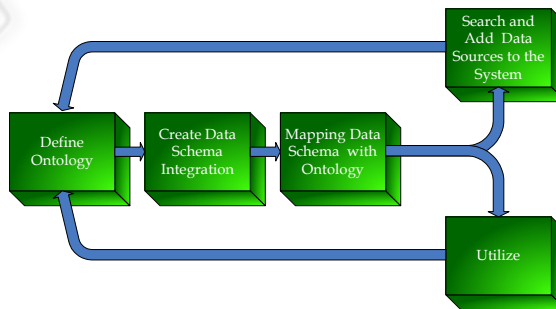

Figure 2: Ontology Based approach.

4 TOURISM INFORMATION AGGREGATION ARCHITECTURE

In this section we describe our architecture for the aggregation of information from different data sources in the tourism domain (figure 3). The aim of the architecture is to provide a framework that

allows the aggregation of tourism information following the Ontology Based Approach. The framework must access tourism data sources, extract their information, combine the data from the different sources and present it to the tourist in an aggregated form. The architecture is composed of four layers. Each one of these layers will be described next.

4.1 Semantic Layer

One of the most important components of the architecture is the ontology. It is in the ontology where we define all the concepts to aggregate. The ontology must be defined in OWL language (OWL, 2004). Using the ontology elements we can define rules. The rules must be defined in the Semantic Web Rule language (SWRL, 2004). Creating the rules in SWRL and not including them in the ontology adds flexibility to the rules definition. In run time we can activate or deactivate a specific rule. The capability of rule definition is an essential issue to allow dynamic packaging. We can define rules that restrict the tourism packages or that add discounts to a specific package definition. For example, we can define that a person who chooses to book a room in a specific hotel has a discount in a specific restaurant. The rules are managed by the RACER engine and will affect the result of the information queries. All the instances presented in the ontology must respect the defined rules. If a specific instance does not respect the rules, then it is removed from it.

To query the architecture we use the nRQL language. The nRQL language is the semantic query language used in the RACER engine. This language allows query information from the ontology defined in the OWL language.

4.2 Mapping Layer

All the data provided by the data sources must be added to the ontology defined for the architecture. This layer is responsible to transform the syntactic information, defined in XML, in semantic information, defined in OWL. The transformation process uses an XSLT document to transform XML data in OWL data. The XSLT is created using the JXML2OWL tool (Rodrigues, Rosa, and Cardoso, 2006). This tool provides an interface that allows the visual mapping between XML elements and OWL elements. As a result of the mapping we get the correspondent XSLT document.

The tool also provides the mapping rules stored in an XML file. The mapping rules define all the relation between XML elements and OWL elements. These rules are used in the query transformation process. The query transformation process has to transform nRQL queries in syntactic queries. In the transformation process we have to guarantee that all the syntactic data will be extracted in order to execute the semantic queries with success.
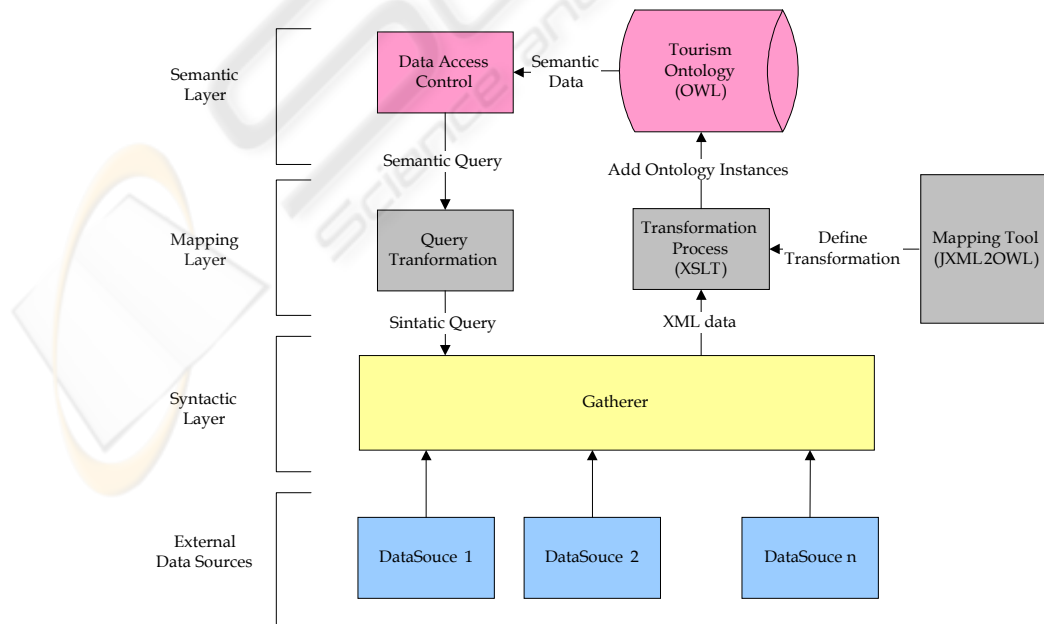

Figure 3: Tourism Information Aggregation Architecture.

4.3 Syntactic Layer

In this layer we integrate all the data sources. We use the Gatherer application (Silva and Cardoso, 2006) to perform the integration. Each one of the data sources is registered in the architecture and mapped to a pre defined XML schema. The XML schema is the one used in the mapping layer. It is created based in the ontology and is used to facilitate the data sources integration.

For each data source to integrate we have to create an XML structure that will define the data that will fulfil a specific item of the XML Schema. Thus, the Gatherer application knows where to get the information for a specific query.

4.4 External Data Sources

This layer is composed of all the data sources that will provide information to the architecture. They can be Data Bases, XML files, Web Services or simple Web Pages.

5 RELATED WORK

Semantic technologies were already used to resolve data aggregation problems. TDS Biological Modeler (Teranode, 2006) is a collaborative biology analysis application that integrates heterogeneous data sources in order to provide aggregated information for scientific analyses. In the healthcare domain the CEN/ISSS eHealth Standardisation Focus Group integrates a set of information systems to allow the exchange of meaningful clinical information among healthcare institutes (Bicer, Laleci, Dogal, and Kabak, 2005). Another example of success is the COG project (Alexiev, Breu, Bruijn, Fensel, Lara, and Lausen, 2005). The aim of this project is the integration of a set of applications existing in an automobile industry.

6 CONCLUSION

The presented architecture can be very useful to create solutions that integrate different data sources to fulfil a specific ontology. In the tourism domain, the information must be aggregated in order to allow the creation of dynamic packages. By using our architecture, we can think first in defining the information concepts that we want to aggregate. Then, search for data sources that can provide the information to integrate with them.

REFERENCES

Werthner, H. and Klein (2004), S., Information Technology and Tourism: A Challenging Relationship. Berlin etc., Springer

Cardoso, Jorge (2005, September), Dynamic Packaging using Semantic Web based Integration, Draft

Kabbaj, Mohamed Y. (2003, June), Strategic and Policy Prospects for Semantic Web Services Adoption in US Online Travel Industry, Retrieved September 12, 2006 from the MIT Center for Digital Business

Fitzgerald, C. Dynamic packaging: The impact of technology on the sale of commodity products, both online and offline, http://www.solutionz.com/pdfs/01-Dynamic_Packaging.pdf. 2005, The Solutionz Group International, Inc.

Fodor, O. and H. Werthner (2004-5). Harmonise: A Step Toward an Interoperable E-Tourism Marketplace. Int. Journal of Electronic Commerce 9 (2): 11-39

Alexiev, V., Breu, M., Bruijn, J., Fensel, D., Lara, R. and Lausen H., Information Integration with Ontologies – Experiences from an Industrial Showcase, John Wiley & Sons Ltd, 2005, p. 1-53, 55- 61 and 151-156

Chandrasekaran, B., Josephson, J. R. and Benjamins, V. (1999), Ontologies: What are they? Why do we need them?, IEEE Expert p. 20-21, Retrieved November 3, 2006 from University of Maryland

OWL - Web Ontology Language, W3C Recommendation, 10 February 2004

SWRL: A Semantic Web Rule Language, 21 May 2004 http://www.w3.org/Submission/SWRL/

Rodrigues, T., Rosa, P. And Cardoso, J. (2006), Moving from Syntactic to Semantic Organizations using JXML2OWL

Silva, B. and Cardoso J. (2006), Semantic Data Extraction for B2B Integration, International Workshop on Dynamic Distributed Systems (IWDDS), In conjunction with the ICDCS 2006, The 26th International Conference on Distributed Computing Systems. July 4-7, 2006 Lisbon, Portugal

TERANODE Design Suite, Biological Modeler (2006) http://www.teranode.com/products/tds/biological_modeler.php

Bicer, V., Laleci, G., Dogal, A., and Kabak Y. (2005), Providing Semantic Interoperability in the Healthcare Domain through Ontology Mapping, Retrieved February 17, 2006 from METU – Software Research and Development