Combining the Semantic Web with Dynamic Packaging Systems

Jorge Cardoso
Department of Mathematics and Engineering
University of Madeira
9000-390, Funchal
PORTUGAL
jcardoso@uma.pt http://dme.uma.pt/jcardoso

Abstract: - With the growth of demand for customized tourism itineraries, agencies seek technology that provides their customers the flexibility to put together unique dynamic packages from a range of alternatives, without having to be aware of the intricacy of contract rules and pricing issues. In recent years, one important type of e-tourism applications that has surfaced to support these requirements is dynamic packaging systems. Our objective is to develop a platform to enable the development of dynamic packaging systems using the latest semantic Web technologies, since it has been recognize that the tourism industry is a perfect application area for the semantic Web since information dissemination and exchange are the key backbones of the travel industry.

Key-Words: - Semantic Web, Ontologies, Semantic Information Integration, Dynamic Packaging, Travel Information Systems.

1 Introduction

Tourism Information Systems (TIS) are a new type of business systems that serve and support e-tourism and e-travel organizations, such as airlines, hoteliers, car rental companies, leisure suppliers, and travel agencies. Dynamic packaging is one class of these systems.

In this paper we are particularly interested in studying the development of dynamic packaging applications. Dynamic packaging can be defined as the combining of different travel components, bundled and priced in real time, in response to the request of a consumer or booking agent. Our approach to the development of a dynamic packaging platform encompasses the use of semantic Web technologies. E-tourism is a perfect application area for the semantic Web since information dissemination and exchange are the key backbones of the travel industry.

While the semantic Web has reached a considerable stability from the technological point of view with the development of languages to represent knowledge (such as OWL [3]), to query knowledge bases (RQL [4] and RDQL [5]), and to describe business rules (such as SWRL [6]), the industry is still skeptic on its potential. For the semantic Web to gain a considerable acceptance from the industry it is indispensable to develop real-world semantic Web-based applications to validate and explore the full potential of the semantic Web [7]. The success of the semantic Web depends on its capability to support applications in commercial settings [1].

Current dynamic packaging applications are developed using a hard-coded approach to develop the interfaces among various systems to allow the interoperability of decentralized, autonomous, and heterogeneous tourism information systems. 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 structures. This diversity makes the interoperability of information systems and technologies very complex and constitutes a major barrier for emerging e-marketplaces and dynamic packaging applications that particularly affects the smaller players (Fodor and Werthner 2004-5).

To take the development and widespread of semantic Web applications a step further, we have designed an architecture based on an infrastructure entirely designed using the technologies made available by the semantic Web, namely OWL, RQL, RDQL, Bossom [15], and SWRL.

2 Dynamic Packaging

With traditional applications travelers must visit manually multiple independent Web sites to plan their trips or vacations, register their personal information multiple times, spend hours or days waiting for response or confirmation, and make multiple payments by credit card. Consumers are discouraged with the lack of functionality. They are demanding the ability to create, manage and update itineraries. With dynamic packaging technology, travelers can build customized trips that combine customer preferences with flights, car rentals, hotel, and leisure activities in a single price.

Dynamic packaging enable consumers (or booking agent) to build a customized itinerary by assembling multiple components of their choices and complete the transaction in real time [17]. Dynamic packaging is based on an individual consumer request, including the ability to combine, multiple travel components like flights, hotels, car rentals, and any other tourism related component in real time and provides a single, fully priced package, requiring only one payment from the consumer and hiding the pricing of individual components within 5-15 seconds [18]. The products available to customer can be stored in local inventories or external sources.

information in the knowledge base, the information is transformed into a set of ontology instances. At this level, we have an e-Tourism ontology describing tourism domain information such as flights, hotels, leisure activities, etc. Since all data sources refer to the same ontology, theoretically there are not syntactic neither semantic conflicts.

The inference and query layer supplies an interface that allows making inference and querying the knowledge-base. Inference is carried out using semantic packaging rules. The query module allows finding information describing travel products stored in the knowledge-base.

Finally, the dynamic packaging layer is responsible for reading the packaging rules specifications and generating valid packages, i.e. travel packages that comply with the packaging rules.

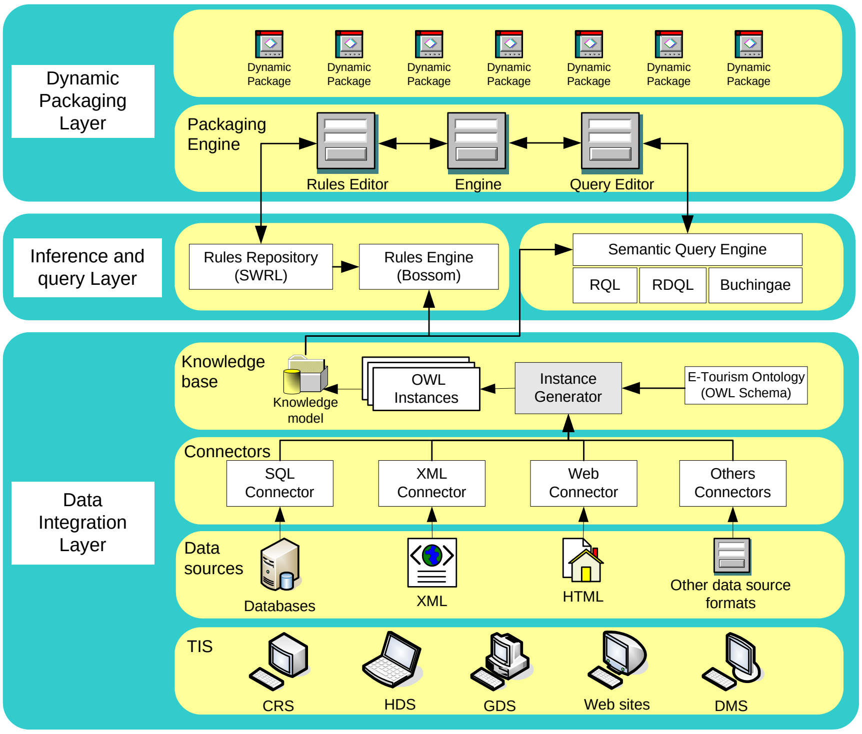

Figure 1. Dynamic Packaging Architecture

2 Dynamic Packaging Architecture

The architecture of our system is composed of three main layers: the integration layer, the inference and query layer, and the dynamic packaging layer. The layers are articulated in the following way. The integration layer includes all the data and information needed by our semantic dynamic packaging system. It typically includes data stored in relational databases (other type of data source are also supported). At this level, we can find information which describes travel or tourism. This information is accessed using connectors that retrieve information from the data sources using a variety of protocols. The information is stored in knowledge base. Before storing the

3.1 Data Integration Layer

One important requirement for dynamic packaging solutions is the existence of an infrastructure to integrate data in an automated way, allowing querying in a uniform way across multiple heterogeneous systems containing tourism related information [18]. The key point of differentiation between dynamic and traditional vacation packages is the ability for the travel consumer to dynamically access data stored into several, separate inventory management systems [19]. Meyer [20] reiterates that a key characteristic of dynamic packaging is to be able to combine services which are described in local inventories or in external sources. The data integration layer uses an ontology to create a shared global knowledge model for all the data sources made available by the tourism information systems. In the next sections we analyze what kind of information systems need to be integrate, what type of data sources are made available, and what is our approach to allow querying in a uniform way multiple heterogeneous tourism information systems.

3.1.1 Tourism Information Systems

One of the challenges that dynamic packaging applications face is the integration of the tourism information systems (TIS), namely, Computerized Reservation Systems, Global Distribution Systems, Hotel Distribution Systems, Destination Management Systems, and Web sites.

A Computerized Reservation System (CRS) is a travel supplier's own central reservation system [21]. A CRS enables travel agencies to find what a customer is looking for and makes customer data storage and retrieval relatively simple. These systems contain information about airline schedules, availability, fares, and related services. Some systems provide services to make reservations and issue tickets.

A Global Distribution System (GDS) is a super switch connecting several CRSs. A GDS integrates tourism information about airlines, hotels, car rentals, cruises and other travel products. It is used almost exclusively by travel agents. There are currently four major GDS [21]: Amadeus, Galileo, Sabre, and Worldspan.

Hotel Distribution Systems (HDS) work closely with GDSs to provide the hotel industry with automated sales and booking services. A HDS is tied into a GDS, allowing hotel bookings to be made in the same way as an airline reservation [21].

Destination Management Systems (DMS) supply information interactively accessible about a destination, enabling tourist destinations to disseminate information about products and services as well as to facilitate the planning, management, and marketing of regions as tourism entities or brands [22]. These systems offer a guide to tourist attractions, festivals and cultural events, coupled with online bookings for accommodation providers. Two of the most well known DMS include Tiscover (Austria) and Gulliver (Ireland).

The Internet is revolutionizing the distribution of tourism information and sales. Previously, many companies had to use their booking systems as platforms from which to distribute their products via existing channels, such as GDSs. Recently, companies have chosen the strategy to sell products on their own Web sites to avoid using a GDS [23]. This is the simplest and cheapest strategy to sale products. A recent survey [24] revealed that over 95% of hotel chains had a Web site, with almost 90% of these providing technology to allow customers to book directly.

3.1.2 Data Sources

Data source integration is a research topic of huge practical importance for dynamic packaging. Integrating distributed, heterogeneous, and autonomous tourism information systems, with different organizational levels, functions, and business processes to freely exchange information can be technologically difficult and costly.

Dynamic packaging applications need to access tourism data sources to query information about flights, car rentals, hotel, and leisure activities. Data sources can be accessed using the Internet as a communication medium. Some wrapping process may be needed to achieve this, but that is beyond the scope of this paper. The sources can contain HTML pages presents in Web sites, databases, specific formatted files, such as XML, or flat files. To develop a robust dynamic packaging application it is important to classify each data source according to its type of data since the type of data will influence our selection of a solution to achieve data integration. For dynamic packaging applications, tourism data sources can host three major types of data: unstructured data, semi-structured data, and structured data.

We will see that the use of an ontology will allow us to integrate data with different structures, resolving the structural heterogeneity of data sources.

3.1.3 Connection layer

The connection layer maintains a pool of connections to several data sources (e.g. relational databases, XML files, HTML online Web pages, etc.). We use a connection layer to achieve two goals: abstraction and efficiency. On one hand, the connection layer adds a level of abstraction over the data sources and it is responsible for presenting a single interface to the underlying data sources. On the other hand, the connection layer provides connection pooling to considerably increase application processing. When data is required from the connection layer, connections to the data sources must be established, managed, and then freed when the access is complete. These actions consume time and resources. The use of a connection layer minimizes the opening and closing time associated with making or breaking data source connections.

3.1.4 Knowledge base

As a solution to the problem of integrating heterogeneous data sources we provide a uniform access to data. To resolve syntactic and structural heterogeneity we map local data sources schema into a global conceptual schema. Since semantic problems can remain, we use ontologies to overcome semantic heterogeneity. To this end, we specify a formal ontology about the specific knowledge domain of tourism to be shared among several external data sources.

The main component of the knowledge base layer is the Instance Generator. The data extracted by the connection layer is formatted and represented using an ontology.

Ontology Creation. The development of an ontology-driven application typically starts with the creation of the ontology schema. Our ontology schema contain the definition of the various classes, attributes, and relationships that encapsulate the business objects that model the tourism and travel domain.

Our e-tourism ontology provides a way of viewing the world of tourism. It organizes tourism related information and concepts and allows achieving integration and interoperability through the use of a shared vocabulary and meanings for terms with respect to other terms. Our ontology was built to answer to three main questions that can be asked when developing dynamic packages for a tourist: What, Where, and When.

- **What**. What can a tourist see, visit and what can he do while staying at a tourism destination?
- **Where**. Where are located the interesting places to see and visit?
- **When**. When can the tourist visit a particular place?

After conducting an analysis of ontology editors, we have select Protégé [25] to construct our ontology. The main components of the e-tourism ontology are concepts, relations, instances, and axioms. A concept represents a set or class of entities within the tourism domain. Activity, Organization, Weather, and Time are examples of concepts used. These concepts were represented in OWL [3] in the following way:

```
(...)
<owl:Class rdf:ID="Activity"/>
<owl:Class rdf:ID="Organization"/>
<owl:Class rdf:ID="Weather"/>
<owl:Class rdf:ID="Time"/>
<owl:Class rdf:ID="Directions"/>
<owl:Class rdf:ID="Transportation"/>
(...)
```

The class Activity (which answers to the question ‘What’) refers to sports, such as skiing, sightseeing or any other activity, such as shopping or visiting a theatre. The class Organization (which answers to the question ‘Where’) refers to the places or locals where the tourist can carry out an activity. Examples of infrastructure that provides the means for exerting an activity include restaurants, cinemas, or museums. The class Time and Weather (which answers to the question ‘When’) refers to the time and weather conditions which allow a tourist to carry out an activity at a certain place. The ontology also includes relations which describe the interactions between concepts or concept’s properties. For example, the concepts Fishing and Hiking are sub-concepts of the concept Sport.

```
(...)
<owl:Class rdf:ID="Fishing">
    <rdfs:subClassOf>
      <owl:Class rdf:about="#Sport"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="Hiking">
    <rdfs:subClassOf>
      <owl:Class rdf:about="#Sport"/>
    </rdfs:subClassOf>
</owl:Class>
(...)
```

Ontology population. By ontology population we refer to a process, where the class structure of the e-Tourism ontology already exists and is extended with instance data (individuals). This can be done either by a computer or by a human editor. In our case, the e-Tourism ontology instances are created automatically by the instance generator. The ontology and its instances is a semantic knowledge-base that integrates information coming from several external data sources. As we have seen in section 3.1.2, data describing the resources may be stored in relational databases, flat files, XML files, and HTML web pages.

3.2 Inference and query layer

The query layer provides a query interface to the e-Tourism knowledge base formed with all the ontology instances automatically generated. The query interface understands three distinct semantic query languages: RQL [4] (RDF Query Language), RDQL [5] (RDF Data Query Language), and Buchingae [26]. These languages allow querying ontology classes, navigating to its subclasses, and discovering the resources which are directly classified under them. Our initial objective was to make available to users a language that would enable to query the native representation of our knowledge base, i.e. OWL, but no suitable query language of this type exists yet.

Using this layer, travel agents are able to query tourism related information. For example, the following query expressed in RDQL allows selecting the hotels that have a cost lower than 60 euros.

```
SELECT ?x,?c,?z
WHERE
   (?x <http://apus.uma.pt/ET.owl#Hotel>
      ?y),
   (?x <http://apus.uma.pt/ET.owl#Name>
      ?c),
   (?y <http://apus.uma.pt/ET.owl#Cost> ?z)
   AND ?z<60
```

The inference engine is implemented with a rule management system. Adopting a rule management system allows to extract and isolate dynamic packaging logic from procedural code. Since the rules associated with tourism information may change quite often, these changes cannot be handled efficiently by representing rules embedded in the source code of the application logic. The option to detach dynamic packaging rules from the source code gives travel agents an effective way for creating a rule base and for building and changing rules.

In our approach, the rules are defined in SWRL (Semantic Web Rule Language) or Buchingae. They correspond to axioms about classes (concept) or their properties of the instance stored in the OWL knowledge-base. By applying these rules on the set of facts it is possible to infer new facts.

SWRL was designed to be the rule language of the semantic Web enabling rule interoperation on the Web. It provides the ability to write Horn-like rules expressed in terms of OWL concepts to reason about OWL individuals.

Since SWRL rules are fairly well-known, we give an example of a Buchingae rule. The rule states that travelers that buy a travel package with a flight, a hotel reservation, and a car rental are eligible to receive a 10% discount on the final price of the package,

```
prefix builtin =
http://www.etri.re.kr/2003/10/bossam-builtin
#;
prefix RUD = http://apus.uma.pt/ET.owl#;
namespace is http://www.etri.re.kr/samples#;
rulebase rb01
{
    (...)
    rule R01 is
      if
        packageProduct(?x, RUD:Flight) and
        packageProduct(?x, RUD:Hotel)
        packageProduct(?x, RUD:CarRental)
       then
         discount(?x, RUD:TenPercent)
}
```

A large number of rule engines are available as open source software. Some of the most popular engines include Jess, Algernon, SweetRules, and Bossam. We chose Bossam [15], a forward-chaining rule engine, as the first rule engine candidate for our semantic course management system since it supports OWL inferencing, it works seamlessly with Java, is well documented, and is very easy to use and configure.

3.3 Dynamic packaging layer

Dynamic packages are automatically created by the dynamic packaging engine. Our architecture includes not only the dynamic packaging engine, but also the rule editor and the query editor. The configuration of the dynamic packaging engine involves the following activities. During the rule development phase, the rule designer defines packaging rules using the rule editor application. The rule editor, a component that provides an interface to the rule repository, supports the creation and modification of packaging rules. Packaging rules are codified and stored in an integrated repository, providing a central point for definition and change, which can later drive dynamic package construction. The construction of packages may also involve querying the knowledge base. This is especially important when a dynamic package has already been put together according to the packaging rules and it is necessary to add information describing each product. This information can easily be obtained from the knowledge base.

4 Conclusion

The industry and its main players are waiting to see how real-world applications can benefit from the use of semantic Web technologies. The success of the Semantic Web vision is dependant on the development of practical and useful semantic Web-based applications. As a contribution to increase the widespread of these new technologies, we have developed the architecture of a Semantic Dynamic Packaging System based entirely on semantic Web technologies (such as OWL, RQL, RDQL, and SWRL). The concept of dynamic packaging is to bundle all the components selected by a traveler to produce one reservation and entails only one payment from the customer. The system can semantically integrate and extract heterogeneous data from tourism data sources describing travel products; answers to complex semantic queries, and is able to carry out reasoning using explicit semantic rules. The system supplies an integrated environment where travel agents can easily create dynamic packages for their customers.

References:

[1]. Cardoso, J., et al., Academic and Industrial Research: Do their Approaches Differ in Adding Semantics to Web Services, in Semantic Web Process: powering next generation of processes with Semantics and Web services, J. Cardoso and S. A., Editors. 2005, Springer-Verlag: Heidelberg, Germany. p. 14-21.

[2]. Berners-Lee, T., J. Hendler, and O. Lassila, The Semantic Web, in Scientific American. 2001.

[3]. OWL, OWL Web Ontology Language Reference, W3C Recommendation. 2004, World Wide Web Consortium, http://www.w3.org/TR/owl-ref/.

[4]. Karvounarakis, G., et al. RQL: a declarative query language for RDF. in Eleventh International World Wide Web Conference. 2002. Honolulu, Hawaii, USA. p. 592-603.

[5]. RDQL, Jena RDQL, http://jena.sourceforge.net/RDQL/. 2005.

[6]. Ian Horrocks, et al., SWRL: A Semantic Web Rule Language Combining OWL and RuleML, http://www.daml.org/2003/11/swrl/. 2003.

[7]. Lassila, O. and D. McGuinness, The Role of Frame-Based Representation on the Semantic Web. Linköping Electronic Articles in Computer and Information Science, 2001. 6(5).

[8]. OWL-S, OWL-based Web Service Ontology. 2004.

[9]. Cardoso, J. Issues of Dynamic Travel Packaging using Web Process Technology. in International Conference e-Commerce 2004. 2004. Lisbon, Portugal.

[10]. Shum, S.B., E. Motta, and J. Domingue, ScholOnto: an ontology-based digital library server for research documents and discourse. International Journal on Digital Libraries, 2000. 3(3): p. 237-248.

[11]. Roure, D., N. Jennings, and N. Shadbolt, Research Agenda for the Future Semantic Grid: A Future e-Science Infrastructure http://www.semanticgrid.org/v1.9/semgrid.pdf. 2001.

[12]. Swoogle, Search and Metadata for the Semantic Web - http://swoogle.umbc.edu/. 2005.

[13]. Kumar, A. and B. Smith, On Controlled Vocabularies in Bioinformatics: A Case Study in Gene Ontology, in Drug Discovery Today: BIOSILICO. 2004. p. 246-252.

[14]. Mandal, C., V.L. Sinha, and C. Reade, Web-based Course management and Web Services. Electronic Journal of e-Learning, 2004. 2(1): p. 135-144.

[15]. Bossom, Bossom engine for the semantic Web, http://projects.semwebcentral.org/projects/bossam/. 2005.

[16]. Meinel, C., H. Sack, and V. Schillings. Course management in the twinkle of an eye - LCMS: a professional course management system. in Proceedings of the 30th annual ACM SIGUCCS conference on User services. 2002. Providence, Rhode Island, USA: ACM Press. p. 281-283.

[17]. Lofgren, S., Metadata for Improving Commercialisation of Dynamic Tourist Packages, http://www.ibit.org/dades/doc/864_ca.pdf. 2005.

[18]. 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.

[19]. Solutions, T., Making sense of dynamic packaging, http://www.triseptsolutions.com/Dynamic_Packaging.pdf. 2005.

[20]. Meier, A., Analysts and Investors day, Mallorca, Spain. http://www.kuoni.com/corporate-site/investor-relations/pdf/AM_Technology.pdf. 2005, Published by Kuoni Group.

[21]. Inkpen, G., Information Technology for Travel and Tourism. 2 ed. 1998, Essex, England: Addison Wesley Longman Ltd.

[22]. Buhalis, D., eTourism - Information technology for strategic tourism management. 2002: Longman. 408.

[23]. Dombey, A. Separating the emotion from the fact - The effects of new intermediaries on electronic travel distribution. in ENTER Information and Communications Technologies in Tourism Conference. 1998. Istanbul, Turkey.

[24]. O'Connor, P., Online Pricing - An Analysis of Hotel Company Practices. Cornell Hotel & Restaurant Administration Quarterly, 2003. 44(1): p. 10-19.

[25]. Knublauch, H., et al. The Protégé OWL Plugin: An Open Development Environment for Semantic Web Applications. in Third International Semantic Web Conference (ISWC 2004). 2004. Hiroshima, Japan.

[26]. Buchingae, RDF Query Survey, http://www.w3.org/2001/11/13-RDF-Query-Rules/#Buchingae. 2005.