ONTOURISM: SEMANTIC ETOURISM PORTAL

Ying Ding
Institute of Computer Science, University of Innsbruck
Innsbruck, Austria

Christoph Herzog
EC3
Vienna, Austria

Michael Luger
Institute of Computer Science, University of Innsbruck
Innsbruck, Austria

Kathrin Prantner
E-SEC Information Security Solutions GmbH
Innsbruck, Austria

Zhixian Yan
Department of Computer Science, EPFL
Lausanne, Switzerland

ABSTRACT

Tourism is viewed as information intensive industry where information plays an important role for decision and action making. Austria, along with its long history on music, culture, and nature, unsurprisingly attracts lots of tourist and Austrian tourism industry contributes significantly to the local economy. Semantic technologies open up new ways of facilitating effective integration of (tourism) information originating from various sources on top of ontologies. This paper reports some work done in the OnTourism project (funded by Austria government). The overall goal of the project is to improve information creation, maintenance and delivery in the tourism industry by introducing semantic technologies to this domain.

Key Words: tourism, Austria, semantic technology, ontology, OnTourism

INTRODUCTION

Tourism has become globally important as it creates lots of job opportunities in many countries. It starts to become the major input for the local, regional and country wide economy. Statistics from World Travel and Tourism Council shows that in 2006, Tourism industry is expected to have US$6477 billion in demand and it will grow to US$12118 billion by 2016. Tourism activity will grow by 4.2% per annum between 2007 and 2016. In 2010 there will be around 1 billion international tourists (forecasted by the World Tourism Organization (WTO)).

Tourism is viewed as information intensive industry where information plays an important role for decision and action making (Inkpen, 1998). As the Web has changed people's daily life, it also brings the revolution to tourism industry. It has significantly influenced the way of information gathering and exchanging in the area of Tourism. More and more tourists no longer go to travel agency to book their flights, hotels and holiday arrangements which they normally do when the Web does not exist. Now they directly go to the Web and usually they get discount price and prompt service. More than 66% of American

users believe that eTourism web sites provide better services than travel agents. A third of American travellers book their travel on the Internet (eTourism Newsletter, 2006). Information technology starts to play a challenging role in the domain of tourism (Werthner & Klein, 1999; Werthner, 2003). eTourism has gradually turned into one of the most booming industrial sectors.

Austria, along with its long history on music, culture, and nature, unsurprisingly attracts lots of tourist. Austrian tourism industry contributes significantly to the local economy. Austria is ranked as the No. 8 country which wants to spend the largest amount on personal travel and tourism in 2006. Given the strategically important position of tourism industry in Austria, this paper reports some work done in the OnTourism project (funded by Austria government). The overall goal of the project is to improve information creation, maintenance and delivery in the tourism industry by introducing semantic technologies to this domain. In details, it aims at (1) applying, concretizing and evaluating semantic technologies to a specific, information-rich and economically very important domain, (2) identifying, developing and integrating reference ontologies for the tourism industry using ontology engineering methods, and (3) showing the proof-of-concept in a real-world scenario of the Austrian tourism industry.

This paper explores the possibility to apply semantic web technologies to eTourism area. It aims to create a semantic content management solution based on the existing Microsoft SharePoint employed by the Austrian Tourism call centre in order to make full usage of both semantic and social metadata. This paper is outlined as followings. Section 2 provides state-of-the-arts in eTourism and Semantic Web areas. Section 3 delineates the architecture and components necessary to realise the addition of a semantic content management as well as the interaction of these components in order to implement the identified semantic content management processes. Section 4 concludes this paper and points out some future works.

STATE OF THE ARTS
This section outlines the state-of-the-art of relevant application and technologies in the area of eTourism and Semantic Web.

eTourism can be defined as the analysis, design, implementation and application of IT and e-commerce solutions in the tourism and travel industry. It is the leading commercial ICT application in B2C e-commerce with the majority of stakeholders and players having implemented or at least defined their future strategy (Stonebraker and Hellerstein, 2001). Content Integration in eTourism must be seen as the most important task and challenge for tourism intermediaries, since all business processes are centred around creating, maintaining and delivering information to various groups of users for whom it serves different purposes (Werthner, 2003; Werthner and Ricci, 2004). Content integration in eTourism is fundamentally different from data integration (Werthner, 2003). Given the heterogeneous environment of IT infrastructures, the National Tourism Organisation has the challenge to aggregate the information and knowledge from all these heterogeneous systems; especially in Austria, where we have some major key players within the eTourism sector (e.g., Nethotels, Tiscover and Feratel). The current technologies are insufficient in combining the contents of regional and national intermediaries and the diverse content providers. Conventional online content residing in web content management systems needs to be augmented with structures representing the meaning of the data.

Semantic technologies provide methods and concepts facilitating integration of (tourism) information originating from various sources on top of so-called ontologies (formal domain conceptualisations). Semantic technologies offer a significant potential for better cross-system integration and a more versatile linkage of available multimedia tourism data based on ontology references and other types of semantic mark-up (such as geo-referencing of data). Previous projects, such as European-funded Harmonise (IST-FP4) and Harmo-TEN (IST-FP5) are targeted at tourism data harmonisation. These projects aim at working out fairly generic methods of data integration. Semantic Web technologies centre around the idea of adding semantics to the existing data in the Internet and making them machine-readable. This leads to the next generation of the Web and allows distributed information processing based on

unambiguous descriptions of resources and services. Ontologies are the backbone of Semantic Web technologies. They establish a common understanding of a domain by making the shared conceptualization explicitly in a machine-processable manner. An ontology represents the domain knowledge by describing its concepts or entities, and the relationships between them in a precise, detailed way (Gruber, 1993). Ontology languages for specifying ontologies and describing instances have been proposed over the past years, some of them became standards. For example, the Resource Description Framework (RDF/S) and the Web Ontology Language (OWL) have been recommended by the W3C (Bussler, Davies, Fensel and Studer, 2004) for this purpose.

Ontological support for eTourism aims for data interoperability. A major initiative is Harmonise, which was developed within the Harmo-TEN IST project (www.harmo-ten.org). Harmonise takes a mediator approach, translating data between two systems with different data schemes. The mediator program acts as a semantic gateway between the systems, permitting the receiver to view the source as an extension of its own database, without concern for the differences in names and representations of data. Therefore, the Harmonise solution (Fodor and Werthner 2005) allows actors to retain their existing data models unchanged, while being able to communicate and interoperate with other participants from the Harmonise virtual network. Due to the large number of partners in this network, building a mediator for any two actors in a tourism industry network quickly becomes infeasible. So a shared minimal set of concepts representing the most relevant concepts for main parts of the tourism industry (e.g., accommodation and travel), has been established by the group of experts in the Harmonise consortium. Data reconciliation among different tourism information systems is obtained with respect to this domain ontology. The Harmonise ontology is given in RDF/S with the description in English and contains over 200 concepts. Another project is maintained by the World Tourism Organization (WTO) which offers a thesaurus on tourism and leisure activities in English, French and Spanish with 1,800 descriptors in each language. This is the first multilingual approach in this area. However, the current state of domain ontologies is quite limited and unsatisfactory due to the specific targeting to various different organizations, sectors and applications (Prantner, Ding, Luger, Yan and Herzog, 2007).

Summing it up, current content creation, maintenance and delivery processes in the tourism industry lack machine-accessible semantics describing the meaning of content items and their networked structure. This causes problems and affects both content providers/intermediaries and content users. It prevents implementing efficient, automated content management processes as well as semantic-based applications. OnTourism aims at reducing / solving these problems and, therefore, opening up new ways of handling, delivering and using tourist information.

SEMANTIC CONTENT MANAGEMENT ARCHITECTURE

The OnTourism project aims to manage documents in a document repository with semantically meaningful data. This semantic annotated data is used in a semantic search application to generate more precise and meaningful search results than a simple full text indexing could accomplish. Furthermore, a social tagging approach complements this semantic search, where a correlation analysis between user selected keywords (tags) and semantic annotation in form of ontology elements will be used to make the ontology more accessible for the average user. Finally, a social bookmarking component contributes to the idea of "keeping found things found", i.e., being able to quickly re-find interesting documents by applying those search terms that make the most sense for the user. This section gives an overview of the semantic content management architecture and a brief overview of the system's components.

Figure 1 shows the OnTourism architecture and the modules. There are 6 modules. The *SharePoint Server module* contains some already existing functions which are partially a simplification and may in reality translate to a series of function calls. The *Ontology Module*, where the OnTourism metadata ontology and its instances (as applied to documents) are stored. The Ontology module builds upon the ontology and tools developed within the project. The *Tagging Module*, where user selected keywords (tags) describing the documents in the repository are stored together with secondary data structures for access

optimisation. The *Mapping Module*, which maintains the statistical mapping between folksonomy tags and ontology elements as well as secondary data structures for faster mapping and clustering. The *Search Module*, which executes queries on the ontology, folksonomy and full text level and combines the results into a unified list with a weighted relevance measure. The *Bookmarking Module*, which allows users to bookmark and describe documents in order to contribute to the idea of "keeping found things found".
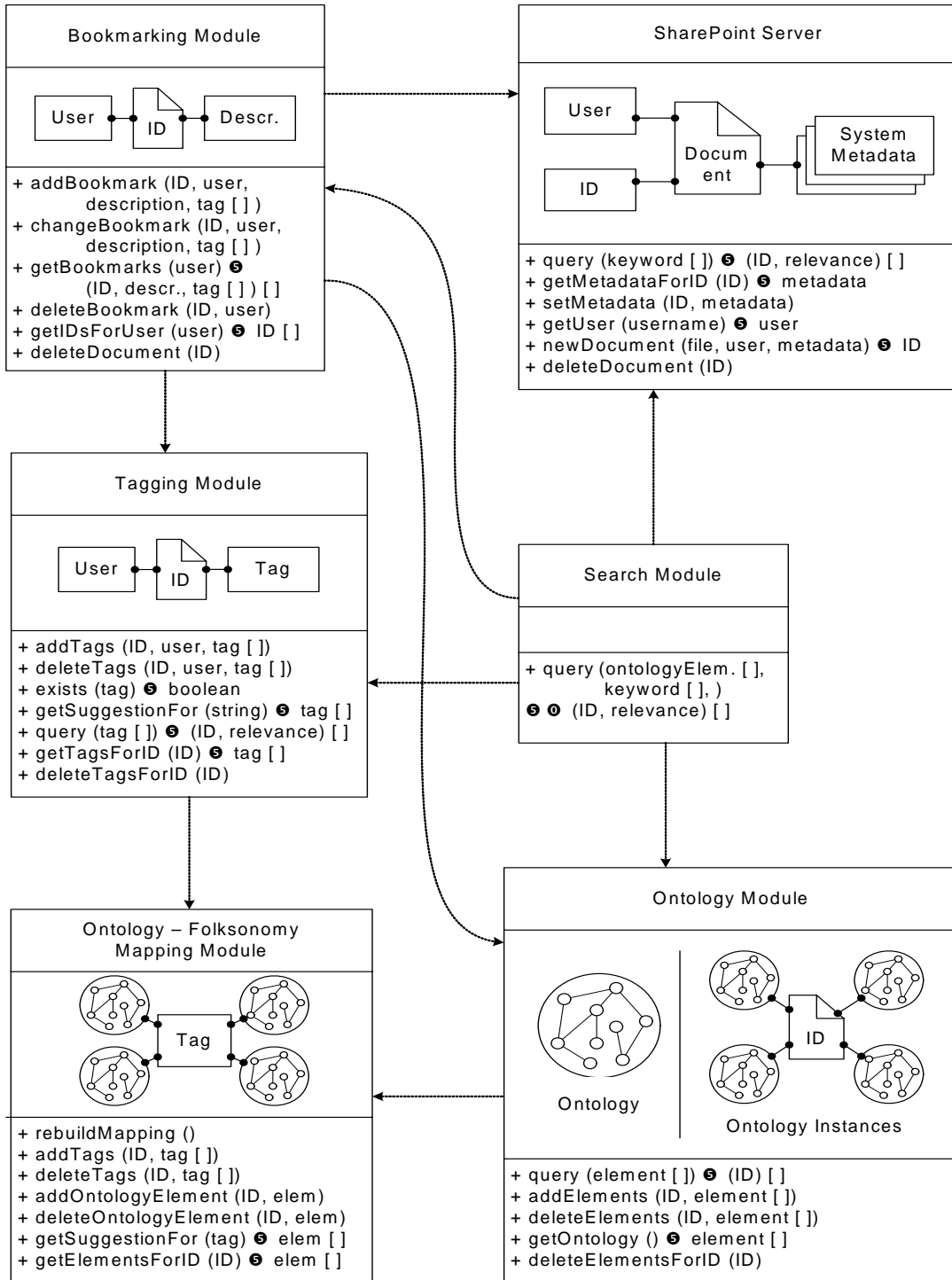
Figure 1
OnTourism Architecture Overview

The main component of the OnTourism architecture is the document repository (see Figure 2), which builds upon the existing document repository of Austria Call Centre. This existing document repository is based on a Microsoft Sharepoint server and is in our approach extended by semantic metadata (ontology storage) and social metadata (folksonomy storage). Furthermore the observable co-occurrence between the two kinds of metadata is captured by a mapping component to generate a statistical mapping between the folksonomy tags and the ontology elements. This mapping is further used in the user interface in order to make the ontology more accessible for the users.
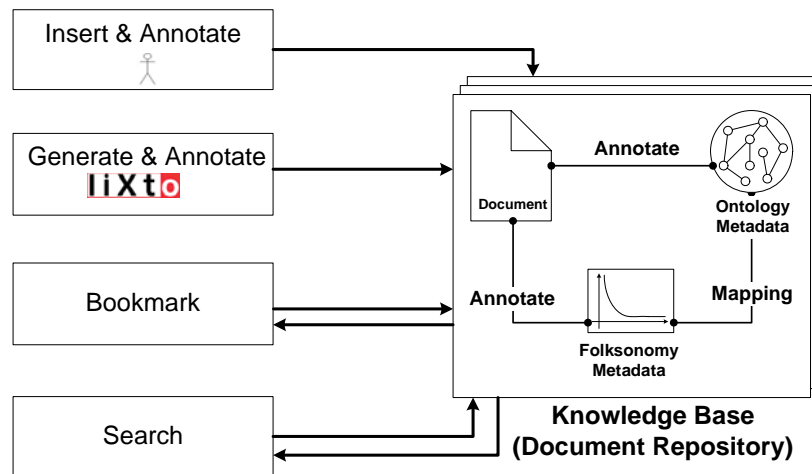
Figure 2
OnTourism Main Components

The first step in the OnTourism semantic content management process is the semantic annotation of documents stored in the repository. In the process of annotating documents newly inserted into the repository as well as editing the metadata of already existing documents, the user interface allows the user to annotate the documents with freely chosen keywords (eventually constituting a folksonomy) as well as with elements from the OnTourism metadata ontology. This selection is guided by the mapping between folksonomy and ontology, where the system suggests to the user probably appropriate ontology elements upon selecting folksonomy tags. After the user selects the appropriate metadata for the document (Ontology, Folksonomy and SharePoint system metadata), the document is inserted into the SharePoint system (if it is a new document) and the appropriate metadata is stored in the ontology and folksonomy storage modules. Finally, the mapping module updates the ontology – folksonomy mapping.

The documents in the document repository of Austria Call Centre are used as well as generated mainly by call centre agents. Having up-to-date documents about rapidly changing topics like current events is clearly beyond the capability of the call centre personnel. Therefore, in order to cope with the continuous stream of information, especially emerging from the rapidly changing cultural events, an automatic solution to generate latest documents about current events is introduced. These documents are automatically created and inserted into the SharePoint server at set intervals. We estimate about 200 documents to be automatically generated per week.

Since the manual metadata annotation of this large number of documents would require too much of the call centre personnel's time, an automatic annotation solution for these documents becomes necessary. The solution is based on the use of Lixto Software (www.lixto.at). Information about current events is crawled from several web sources. This is enriched with relevant information from further web sources and compiled into RDF documents. A description of each document in terms of the OnTourism

ontology is also generated. These documents, along with the generated metadata, are automatically inserted into the OnTourism document repository.

As mentioned before a social bookmarking system complements the semantic content management process, providing (a) a source for semantically less clear, yet meaningful metadata for documents and (b) contributing to the idea of "keeping found things found". Therefore, the users have the possibility to bookmark documents (and other resources) and apply their own, freely chosen keywords (tags) to annotate the documents. Feedback in the user interface guides the user towards selecting terms already existing in the folksonomy. After getting search results, the user can bookmark any of these results from the search interface. The bookmarking component asks for an (optional) description of the document and offers the user the possibility to apply tags to the document which are cascaded to the tagging component and the ontology – folksonomy mapping component.
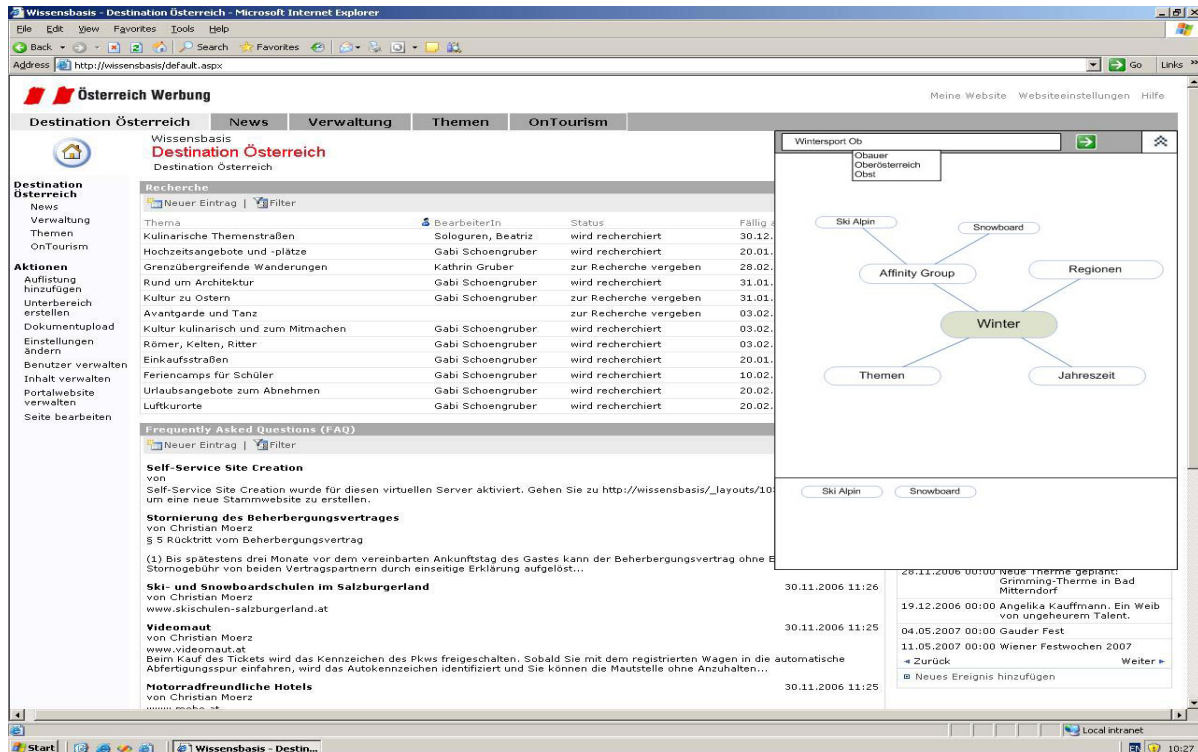

Figure 3
OnTourism portal screenshot

The most important part of OnTourism's semantic content management processes is the search process. Using semantically meaningful metadata from an ontology as well as freely chosen keywords which are considered meaningful by the users for search in the document repository, the precision and recall of the search should be better than with only a normal syntactical search based on full text indexing. The user can select freely chosen keywords for the search. The user interface, utilising the tagging module, proposes tags from the folksonomy to choose. Based on this choice, the user interface utilises the mapping module to suggest probably appropriate ontology elements to the user (see Figure 3). Having specified any or all of the three search components (ontology elements, folksonomy tags, other keywords), the composite query is sent to the search module which in turn queries the ontology module, the tagging module and the full text indexing function of the SharePoint server. The search module unites the search results and ranks them, where the results from the ontology module have the highest weight and the results from the full text indexing have the lowest weight.

DISCUSSION, CONCLUSION AND FUTURE WORK

Today's information management concepts and solutions for the complex tasks of tourism intermediaries are still at the low-level from a semantic point of view. Information exchange takes place in predefined scenarios only and requires a lot of bilateral agreements regarding content, exchange formats, and business process rules. This hampers the users/requesters of those information, thus it limits searching and choosing "the right" tourist offers. This paper presents some research work done within the OnTourism project. It identified the importance of applying semantic technologies to advance information creation, maintenance and delivery in the tourism industry. It presented the overview of the semantic content management architecture and its key components based on the real life scenarios provided by Austria Call Centre. The rational for developing OnTourism portal are the followings: (1) to make it possible to be integrated with other existing tourism data (based on pre-defined ontologies and ontology alignment); (2) to make it easier for end users to find what they want as they have three ways of search (ontology, tags and keywords); (3) to make the content provider (Austria Call Center employee) to better manage their data. It has already shown the impact on the content provider itself as the content are annotated based on pre-defined ontologies and they become easier to manage, browse and search. Although some training efforts need to be spent to train Call Center employee to understand ontology, to annotate data based on ontologies, but the long-term business implications on such system can be foreseen when their data need to be integrated with other tourism data and end users increases their expectation to find proper or precise information they request. In the long run, the quality of services of Austria Call Center should be increased.

Future work will be focused on further implementing of this architecture and testing the whole system with the users from Austria Call Centre. Furthermore, evaluation will be conducted by comparing this system with identified related works. The implication from business and technical point of view should be explored as well.

ACKNOWLEDGEMENT

REFERENCES

Bussler, C., Davies, J., Fensel, D. and Studer, R.(2004). The *Semantic Web: Research and Applications*. Springer: Berlin.

eTourism newsletter. (2006). *eTourism web sites preferred to traditional travel agents*. http://www.etourismnewsletter.com/ecommerce.htm#behaviour. [Accessed the 23rd of December 2007, 11:03]

Fodor, O. and Werthner, H. (2005). Harmonise – a Step towards an Interoperable e-Tourism Marketplace. *International Journal of Electronic Commerce*, 9(2). 11-19.

Gruber, T. (1993). Toward principles for the design of ontologies used for knowledge sharing. *Proceedings of the Padua workshop on Formal Ontology*, March, 1993.

Inkpen, G. (1998). *Information technology for travel and tourism*. Edinburgh: Addison Wesley.

Prantner, K.; Ding, Y.; Luger, M.; Yan, Z. and Herzog, C. (2007). Tourism Ontology and Semantic Management System: State-of-the-arts Analysis. *IADIS International Conference WWW/Internet 2007*, Vila Real, Portugal, 5-8 Oct 2007.

Stonebraker, M. and Hellerstein, J.M. (2001). Content Integration for E-Business. *Proceedings of ACM SIGMOD/PODS 2001* (pp. 552-560). Santa Barbara, California, USA.

Werthner, H. (2003). Intelligent Systems in Travel and Tourism, *Proceeding of the 18th International Joint Conference on Artificial Intelligence (IJCAI-03)* (pp. 1620-1628). Acapulco, Mexico, August 9-15, 2003.

Werthner, H. and Klein, S. (1999). *Information technology and tourism: a challenging relationship.* Berlin: Springer.

Werthner, H. and Ricci, F. (2004). Electronic Commerce and Tourism. *Communications of the ACM*, 47(12): 101-105.