OnTour: Tourism Information Retrieval based on YARS

Katharina Siorpaes
Digital Enterprise Research Institute
Technikerstraße 21a
Innsbruck, Austria
+43 512 507-6488

katharina.siorpaes@deri.org

Daniel Bachlechner
Digital Enterprise Research Institute
Technikerstraße 21a
Innsbruck, Austria
+43 512 507-6488

daniel.bachlechner@deri.org

ABSTRACT

Semantic Web enabled information retrieval can help to overcome problems caused by information flooding in the World Wide Web. Various industry sectors are faced with the challenge to set up and maintain websites that provide information required by the consumer. Within the scope of the OnTour project, a system based on a fast and flexible Semantic Web backbone has been developed focusing on e-tourism. The major benefits of the OnTour approach, such as its simplicity, modularity, and extensibility, can already be highlighted with the help of the system's current prototype.

Categories and Subject Descriptors

H.3.3 [**Information systems**]: Information Search and Retrieval – *Clustering, Information filtering, Query formulation, Relevance feedback, Retrieval models, Search process, Selection process*

General Terms

Experimentation

Keywords

Semantic Web, RDF, YARS, Information Retrieval, Tourism

1. INTRODUCTION

Ontologies and ontology-based information retrieval have the potential to significantly improve the process of searching information on the World Wide Web. Concept search and browsing can ease the burden of searching the web using keyword based techniques. This is especially important in information-based industries, such as e-Tourism. The predominant goal of the OnTour project is to show that the application of Semantic Web technologies in a real-life scenario can have substantial effects on current flaws such as information flooding in Web-based tourism environments. The OnTour system is a starting point for a field study in cooperation with the tourism industry in Austria to investigate the application of Semantic Web technologies in a real-life scenario. The affinity to new information technologies and the increasing importance of the internet make e-tourism a perfect candidate for this study. There are many challenges but

only a small selection that have major impact on the tourism industry and are therefore relevant for the OnTour project is cited here as examples. In this demo proposal, we describe a first prototype based on the RDF store YARS. The system's architecture is accentuated in section 2. In section 3, we explain how the demonstration itself will be conducted. In section 4, we provide an outlook to future work. Finally, we present our conclusions in section 5.

1.1 Major Challenges

The tourism industry has an above-average growth rate as well as a great economical importance throughout Europe. It is notably information dependent and needs new ways to deal with its vitality and scale. At present primary tourism products entail high information search costs for the clients as well as the suppliers. According to [4] such informational market imperfections lead to the establishment of long and costly information and value chains. In the highly competitive European tourism market it is vital for primary suppliers to provide a fair amount of information online. Due to the SME structure of the European hotel and restaurant sector, which includes roughly 8.5% of all European enterprises, it isn't affordable for most of the suppliers to set up, maintain, and advertise their own websites. A study in the tourism market described in [2] shows that only 60% of accommodations maintain their own websites, while 93% are members in tourism portals. [3] points out that currently the only alternative for primary suppliers to participate in e-tourism are structured information repositories which are built and maintained at high costs mostly from large providers of all inclusive packages. But besides the high costs, further problems such as long winded data maintenance, low adaptation rate and general inflexibility as well as suboptimal target groups argue against it.

1.2 Proposed Solution

OnTour[1] is a lightweight system based on the RDF data store YARS [1]. The approach ultimately makes it possible to negotiate a large part of the major restraints known in connection with Web-based tourism and in particular the challenges mentioned above. YARS relies on optimized index structures and thus enables fast storage and retrieval of large amounts of data while retaining the small footprint and lightweight architecture approach of OnTour. Describing data in RDF helps to overcome a list of problems known in Web-based information systems with respect to quality and processability.

---

[1] http://ontour.deri.org

2. SYSTEM ARCHITECTURE

OnTour enables querying a set of accommodation facilities of different types and categories by specifying a list of constraints. The primary user interface is Web-based. In this section, we describe the user interface, the data store, and the internal commutation of the system.

2.1 User Interface

The Web interface provides the user with a form that allows them to restrict the set of results by filtering them according to location and type constraints. For the purpose of debugging the query sent as well as the results received from the data store are displayed in text areas. The result values can be specified before submitting a query by checking the particular boxes. The result can also be shown in XHTML at the bottom of the page. In this case all result values are displayed by default. XSLT is used to transform the result returned by the data store according to the guidelines defined in the XSL files. With the help of XSL and CSS files the actual output can be defined and adjusted to respective needs easily with-out having to change or even understand the rest of the code. Ensuring the ease of OnTour's use has always been an important factor during the development process of the user interface. The context-aware form adapts automatically without reloading the page to changes in the task being performed. Intuitive handling as well as fault tolerance have always been major issues and will become even more important as the prototype is evolving.

2.2 Data Store

YARS is a data store for RDF which allows querying based on a declarative query language. The YARS servlet uses RDF/N3 for encoding facts as well as queries and provides an HTTP interface for performing query, add, and delete operations. By default YARS delivers results in form of RDF/N3 but the HTTP Accept Header can be used to make the YARS servlet return in XML format. YARS is a system that combines methods from information retrieval and databases to allow better query answering performance. In contrast to other implementations such as Jena2 and Sesame, YARS provides optimized index structures. The data set in use for the current version of the prototype including more than 4600 Tyrolean hotels was kindly made available by the Austrian Federal Economic Chamber. The data structure is very simple and comprises only two concepts, accommodation facilities and related occupational groups.

2.3 Communication

The N3QL queries are assembled dynamically on the client's side before they are sent to YARS using the XMLHTTP API. The communication between the Web interface and YARS happens asynchronously to ensure maximum usability. The HTML presentation is directly connected to XML data for interim updates without reloading the page. The results, delivered in XML, are further processed and displayed on the client's side.

3. DEMONSTRATION

The demonstration of the OnTour system consists of two parts. First the architecture is explained and visualized, second the functionality of the prototype is demonstrated. A variety of queries are executed and their results discussed during the demonstration. Furthermore the advantages of the approach at hand are brought out clearly from both a tourist's and a tourism operator's point of view. Within the scope of the demonstration it is also shown how the major challenges described in the introduction can be overcome with a fully-fledged tourism information system based on the notions of OnTour.

4. FUTURE WORK

To show the full potential of the RDF-based OnTour system, associate information has to be combined with the existent data sets. Geographical information is a possibility as well as information about classification criteria, leisure infrastructure or events. In the medium run information depth, scalability and performance have to be analyzed and optimized. This happens hand in hand with substantial extensions of the knowledge base in terms of the quantity and quality as well as the necessary adaptations concerning the user interface and its usability. The modular design and the simplicity of the prototype guarantee easy exchange and extension of components in the future. The lightweight architecture also allows the system's adaptation to the requirements of newly emerging standards and devices.

5. CONCLUSION

E-tourism is a perfect candidate for Semantic Web because it is information-based and depends on the WWW, both as a means of marketing and transaction channel. Ontology-based information retrieval makes it possible to handle the known challenges in connection with Web-based information systems in a more efficient way. Even though OnTour is a prototype carrying a relatively small stock of data based on a small ontology, it is able to demonstrate its use very well and more importantly can give an interesting outlook on future work.

6. REFERENCES

[1] Andreas Harth, and Stefan Decker. Yet Another RDF Store: Perfect Index Structures for Storing Semantic Web Data With Contexts. Research Paper, 2005.

[2] Martin Hepp, Katharina Siorpaes, and Daniel Bachlechner. Towards the Semantic Web in E-Tourism: Can Annotation Do the Trick?. In *Proceedings of the 14th European Conference on Information System (ECIS 2006)*, Gothenburg, Sweden, June 12-14, 2006.

[3] Alexander Mädche, and Steffen Staab. Applying Semantic Web Technologies for Tourism Information Systems. Research Paper, 2001.

[4] Hannes Werthner, and Francesco Ricci. E-Commerce and Tourism. In *Communications of the ACM*, 47(12):101-105, New York, December 2004. ACM Press.