Towards the Semantic Web in e-Tourism: Lack of Semantics or Lack of Content?

Martin Hepp
University of Innsbruck
Digital Enterprise Research Institute
Innsbruck, Austria
martin.hepp@deri.org

Katharina Siorpaes
University of Innsbruck
Digital Enterprise Research Institute
Innsbruck, Austria
katharina.siorpaes@deri.org

Daniel Bachlechner
University of Innsbruck
Digital Enterprise Research Institute
Innsbruck, Austria
daniel.bachlechner@deri.org

ABSTRACT
The Semantic Web aims at making the wealth of information available on the Web accessible to more precise search and automated information extraction and processing, based on a machine-readable representation of meaning in the form of ontologies. One common assumption is that the Semantic Web can be made a reality by gradually augmenting existing Web data by ontological annotations. In this paper, we describe the results of a quantitative analysis of Web content about Austrian accommodations and show that the data necessary to make the vision of the Semantic Web a reality is widely not available on the current Web, not even in a human-readable form. We discuss possible causes and argue that Semantic Web services technology will mitigate the problem, since business entities are more likely to expose functionality than data in competitive markets.

Categories and Subject Descriptors
K.4.4 **[Computers and Society]:** Electronic Commerce; I.2.4 **[Artificial Intelligence]:** Knowledge Representation Formalisms and Methods; J.1 **[Administrative Data Processing]:** Business; J.7 **[Computers in Other Systems]**

General Terms
Management, Human Factors, Documentation, Measurement

Keywords
Semantic Web, Semantic Web services, e-Tourism, Web, Annotation, Ontologies.

1. INTRODUCTION
One common assumption is that the Semantic Web can be achieved by gradually augmenting the existing data by annotations, and that the main problem of today's web is the "needle in the haystack" problem: everything is there, but we only have insufficient methods of finding and processing what's on the Web. In this paper, we show that, at least in the domain of eTourism, this assumption is inappropriate, because the available Web resources do not contain sufficient information, even from a perspective of a human user. As a consequence, even a perfect annotation of existing Web content would not allow the vision of the Semantic Web to become a reality in the domain of e-Tourism. We analyzed websites operated under the direct control of the accommodation management and also such maintained inside tourism portals. At this point, we abstract from the task of annotation itself, i.e. to which degree the process of adding machine-readable meaning to existing content. In section 2, we describe the methodology for our analysis. Section 3 summarizes our data and highlights core findings. In section 4, we discuss the implications of our findings for Semantic Web research. Section 5 concludes the paper.

Demos and Posters of the *3rd European Semantic Web Conference (ESWC2006)*, Budva, Montenegro, June 11-14, 2006.

2. METHOD
First, we identified information categories that are relevant for consumers looking for travel accommodation. For this, we reused a survey by the Austrian Chamber of Commerce [1], which describes the information needs of consumers. We added the category "availability", which is not listed explicitly in the survey, since it is a core information need in typical travel-related scenarios. Then, we created an ordinal scale for the amount of information per category, ranging from 0 (no information) to 5 (comprehensive coverage of all aspects). Second, we obtained the official directory of all legal accommodations located in the state of Tyrol (n=4,665). Third, we took a random sample (n=100) of the listed accommodations, and for each entry in this sample, searched the Internet for an official Web page. If we could not find a Web resource or if we had doubts about the identity, we called the owner or operator of the accommodation for clarification. Fourth, we checked the leading Austrian tourism portal Tiscover (http://www.tiscover.at) for entries covering the very same sample. Fifth, we manually analyzed the content of both the respective vendor-operated Web resources and the Tiscover entries, and graded the amount of available information using the predefined ordinal scale. Sixth, we aggregated the results and determined the amount of Web resources and portal entries that provide at least a "sufficient" amount of information in the respective category according to the grading scheme. Sufficient was defined in the sense that all information is given that an average consumer needs in order to determine his or her perceived utility of an available accommodation, i.e. to make a reservation decision.

3. RESULTS
In this section, we present the results of the survey and highlight significant observations. We clearly distinguish between vendor-provided data and portal data, since the portal Tiscover, which was part of our analysis, is a managed Web site and does not grant external access to the full internal database. The potential contribution of Semantic Web technology for tourism *portals* is likely smaller as compared to individual Web sites, since portals put a lot of effort into resolving inconsistent data representation

and thus mitigate the problem of data heterogeneity and other obstacles that Semantic technology promises to overcome.

3.1 Representation of Accommodations in the WWW

Out of the 100 accommodations in the sample, 60 maintain a Web site individually for this accommodation, either operated by the hotel owner or managed by a service provider. Additionally, all of these 60 are members of the Austrian tourism portal Tiscover. 33 are only represented in the Tiscover tourism portal. 5 % cannot be found at all on the Web but their existence could be verified by phone, and 2 % do either not exist any longer or could not be found at all.

3.2 Coverage of Information Categories

In this section, the percentage of Web sites that contain at least a sufficient amount of information in the various categories is visualized in Figure 1. As detailed above, sufficient means a rating of 3 points or more in the grading scheme.

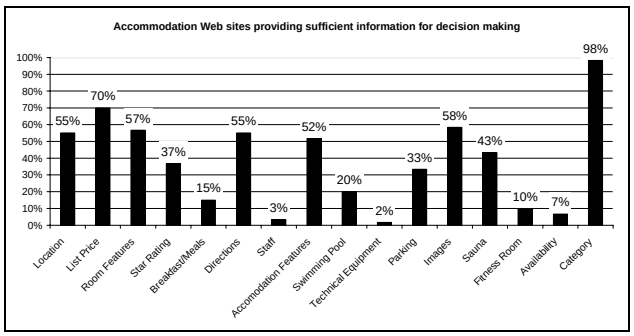

Figure 1. Percentage of specific Web sites with sufficient information for decision making.

3.3 Availability of Information in Tourism Portals

In this section, we summarize the coverage of consumers' information needs among the 93 hotels that are listed in the Tiscover portal. We were surprised by the following observations:

1. Only 27 % of those hotels from our sample that are listed in Tiscover give price information (whereas 70 % of the hotel Web pages contain at least a list price). 2. Only a third of those hotels have sufficient room feature descriptions in Tiscover, while more than half (57 %) of the vendor-operated Web pages contain such detail. 3. 20 % of all listings in Tiscover give sufficient information about technical equipment, while only 2% of the Web pages contain such detail. 4. Three times as many hotels give current availability information on Tiscover (22 %) as compared to vendor operated Web sites (7 %). Still, the biggest part of all Tyrolean hotels does not provide current availability information anywhere on the Web.

4. DISCUSSION

In our sample only 60% of the accommodations maintain a website, while 93% are represented in a tourism portal. Thus, it is very likely that a portal membership is more feasible than a self-maintained website. This is an important quality when assessing the potential of Semantic Web technologies in this sector. From the results we can see that the self-operated websites of accommodations lack information. Only 7 % offer room availability information, which is the most important fact when searching for a suitable offer. The remaining 93 % of accommodation Web sites require a user to either call or communicate by e-mail with the provider in order to get availability information. This is a serious obstacle for making the Semantic Web a reality in the E-tourism domain. The situation inside the tourism portal Tiscover is remarkably better, but in several ways still surprisingly insufficient. For eight out of ten hotels, no current availability data is available, and for 73 % of the hotels not even a list price can be retrieved. The predominance of tourism portals is a challenge for the Semantic Web, since the internal databases are not accessible, and the discovery and matchmaking of consumer request and available supply is hidden inside the portal. As a consequence, the Semantic Web cannot be made a reality in the sector we analyzed by annotating information on Web pages ("data-centric Semantic Web"). Rather, turning the Web into the Semantic Web requires annotating exposed functionality, i.e. services.

5. CONCLUSIONS

We have presented evidence that the current quantity of information available on the Web would not allow the Semantic Web to become a reality in the domain of accommodations in Tyrol. Our analysis has shown that dedicated Web sites do not contain sufficient information that would allow potential guests to make a reservation decision without additional e-mail or phone communication. Similarly, information in managed tourism portals is insufficient, while, additionally, those portals keep control over all functions and the data inside their databases, which are not exposed on the Web. Since we have no reason to assume that the encapsulation of information inside systems will decrease, we can assume that the Semantic Web will only become a reality, in the domain of e-Tourism, if it includes the annotation of functionality and not just published information. In short, the vision of the Semantic Web will not become a reality without Semantic Web services tech-nology, e.g. WSMO or OWL-S.

6. REFERENCES

[1] S. Dolnicar and T. Otter. Marktforschung für die oesterreichische Hotelklassifizierung (Market research for the Austrian hotel classification schema; in German). Vienna, Austrian Chamber of Commerce, 2001.

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

[3] Dogac, A., Y. Kabak, G. Laleci, S. Sinir, A. Yildiz, S. Kirbas and Y. Gurcan: Semantically Enriched Web Services for the Travel Industry. SIGMOD Record 33 (3), 21-27, 2004