Applying Semantic Web Technologies for Tourism Information Systems

Alexander Maedche[a] and Steffen Staab[b,c]

[a] Research Center for Information Technologies at the University of Karlsruhe
Research Group Knowledge Management (WIM)
76131 Karlsruhe, Germany
maedche@fzi.de

[b] Institute AIFB University of Karlsruhe,
76128 Karlsruhe, Germany
staab@aifb.uni-karlsruhe.de

[c] Ontoprise GmbH, Haid-und-Neu-Straße 7,
76131 Karlsruhe
staab@ontoprise.de

Abstract

The Semantic Web as the next generation web is the vision of having background knowledge about the meaning of web sources stored in a machine-processable and –interpretable way. The area of tourism is highly dynamic area that currently already extensively uses the available Internet technologies. However, the shortcomings of the existing technology are that information finding and extraction as well as the interpretation of the information contained in the web sources is left to the human user. In this paper it is shown how the vision of the Semantic Web and already existing Semantic Web technologies can be used for next-generation tourism information systems.

Keywords: Semantic Web, Ontology, Metadata

1 Introduction

Looking at current information systems with tourism perspectives, but also in general, one currently finds two extremes. On the one hand side there are well-structured information repositories, neatly built and maintained at high costs, mostly from large providers of touristic packages, travel or lodging arrangements. Virtually all of these providers offer easy access to their systems via the World Wide Web, but they are mostly isolated from each other and they rarely include detail information.

On the other hand, however, there is the World Wide Web as a whole with its many small, detailed pieces of information, e.g. about opera festivals, and touristic offers in market niches, e.g. agrotourism in particular regions or regional style lodging (examples are Pousadas in Brazil or Portugal or Japanese Ryokans). Because of the vastness of the Web there lies a heavy burden on the user for accessing the latter kind of information, as well as for interpreting it and connecting it to the offers made by the large providers.

This paper sketches how the gap between the two extremes may be narrowed. The goal is to *semantically* connect currently isolated pieces of information in order to diminish the burden on the user of finding and understanding the information sources and in order to allow for individual use of tourism offers.

The key to this goal is the Semantic Web and its technologies. Tim Berners-Lee coined the vision of a Semantic Web as an extension of the current World Wide Web that does not only provide information at the syntactic level to human users, but also at a machine-understandable, semantic level to machines (cf. (Berners-Lee, 1999)).

In the Semantic Web, background knowledge about the meaning of web resources can be stored as machine-processable (meta-)data. Services for finding, integrating, or connecting information may be based on these semantic descriptions. For instance, the semantic description of Ryokans as a Japanese style of lodging returns their existence and the corresponding touristic offers even to the novice Japan visitor who is looking for accomodations and who did not know about their existence before. Analogously, the search for *classical music festivals* with corresponding travel arrangements and checking of availabilities could be integrated by semantic means such that the *opera event* in Usedom, a German island in the Baltic sea, is returned with corresponding travel and lodging arrangements though it has not been questioned for explicitly.

Thus, semantics is seen as a key factor to finding the way in the expanding web space, where currently most web resources can only be discovered via syntactic matches (e.g., keyword search). Germane to the idea of using these semantic descriptions are *ontologies*. Ontologies provide a formal conceptualization of a particular domain that is shared by a group of people. Ontologies describe vocabularies as a kind of complex (meta-)data schemata that are used in order to combine semantic metadata and offer added-value services on top of semantic descriptions.

The remainder of the paper is organized as follows. In Section 2 we introduce several scenarios that portray some pitfalls and desiderata of current tourism information systems, which could be helped with by means of the Semantic Web and its technologies. Section 3 presents some of the core ideas of the Semantic Web that are necessary to understand how the scenario applications could work. Then, we substantiate our vision with some of the methods and tools that already exist in Section 4.

2 Application Scenarios

The high-level idea of Semantic Web-based applications in tourism information systems has been sketched above. In this section, we want to elaborate on desiderata one would like to have fulfilled for tourism information systems. We here describe some application scenarios that are within reach today and that subsequent descriptions of methods and techniques in Section 4 may refer to and indicate how to solve. The following scenarios are given at increasing levels of sophistication.

2.1 Scenario: Semantic Search Engine for Tourism

One of the first needs of a prospective tourist, but also a worker in a travel agency, is the need for searching information. This need, however, contrasts with the coverage and possibilities allowed through current state-of-the-art systems. Either systems belong to the first category mentioned above: neatly built, but with restricted coverage of topics. Or, there is the second category, the Web, that offers almost anything, but that makes the right piece of information nearly impossible to find.

What is need here are *Semantic Searches*. For instance, there may be information (given in some structured format) that state facts such as uttered in (1) and (2).

(1) *The island Usedom belongs to Mecklenburg-Vorpommern*[1].

(2) *The Usedom opera festival is a touristic highlight every summer.*

Then, a tourist who is planning to travel the region might look for places and activities
and she may ask for *classic music events* in Mecklenburg-Vorpommern. Currently, she can hardly find the information, because the information is not in a central database and keyword search may not help, because the two pieces of information are found at different places, e.g. in a database (Usedom is part of Mecklenburg-Vorpommern) and in a web page (about the opera festival in Usedom).

What is needed is Semantic Search that allows for querying distributed data, considering the semantics of concepts and instances like "Opera Festival" or "Usedom".

2.2 Scenario: Browsing Topic Portals

A lot of the information one looks for cannot be found by searching, because novices typically do not know what to search for. They may rather want to explore the possibilities offered by the travel package they have bought. This often leaves them with the problem of integrating an abundance of information, e.g. cultural events offered in a large city and displayed on different web sites with partially overlapping content. Thus, what is needed are topic or location specific portals that integrate available information and let the tourist browse and explore the local offers.

A similar problem occurs when a tourist wants to explore new types of offers, e.g. river rafting and has to sieve through multiple sites.

2.3 Scenario: Semantics-based Electronic Markets

Automatic electronic markets help where the match between providers and requesters must be made fast and/or there is a large volume of transactions. In tourism information systems both criteria apply. Late vacancies of flights or lodging easily are lost, and new offers and requests come in by the minute.

(Ygge, Akkermans, 1996) have shown how power load management may yield benefits to stakeholders in a spot market of electric power supply. Such a spot market exhibits a number of characteristics very similar to ones for tourism needs. However, the difference between touristic offers and power supply is that the former constitutes a much more complicated product. Besides of location, time and amount, parameters like quality of lodging, entertainment, geographic region, etc. play an emminent role. This requires a rich conceptual model about the tourism domain such that the benefits of electronic markets may be applied in tourism. One eventual goal could be that the final customers let their agents trade against the final providers with agencies providing the market place and the integrated information.

2.4 Scenario: Web Services for Tourists

[1] Usedom is an island, a part of the German state of Mecklenburg-Vorpommern.

Imagine registering for a conference online. The conference Web site lists the event time, date and location, along with information about the nearest airport and a hotel that offers attendees a discount. With today's Web, you have to first check to make sure your schedule is clear, and if it is you have to cut and paste the time and date into your calendar program. Then you need to make flight and hotel arrangements, either by calling reservations desks, or by going to their Web sites. There's no way you can just say, 'I want to go to that event', because the semantics of which bit is the date and which bit is the time has been lost.

The scenario one would like to have is that one gives some preferences about maximum budget and minimum of comfort, let your software find out about the constraints (including e.g. your personal datebook) and propose a complete package to you that considers the information given on the web page.

3 The Semantic Web – A Web of Metadata

In the section above we have introduced and discussed some prototypical scenarios that are on a "wish-list" for next-generation tourism information systems. The Semantic Web that is currently investigated by different research communities is a suitable candidate to support the development of next-generation tourism information systems. We have already mentioned that the Semantic Web is based on machine-readable and processable metadata. The layers of representation that allow the features described in the scenarios of section 3 and that determine the working of the tools and methods are briefly discussed in the following.

The Syntax Layer: The interchange of data represented in the Semantic Web must be facilitated through a concrete serialization syntax. XML is an obvious choice frequently used by the upper layers. However, it is important to mention that the Semantic Web is not tied to a particular syntax. Within the syntax layer Unicode[2] is used, that provides a unique number for every character, no matter what the platform, what the program, or what the language is. Beside Unicode, the usage of so-called URI (unified resource identifiers) is essential (see http://www.w3.org/Addressing/Activity).

The RDF(S) Layer: The Semantic Web concept is to do for data what HTML did for textual information systems: to provide sufficient flexibility to be able to represent all databases, and logic rules to link them together to great added value. The first steps in this direction were taken by the World-Wide Web Consortium (W3C) in defining Resource Description Framework (RDF) (Lassila et al. 1999), a simple language for expressing relationships in triples where any of the triple can be a first class web object. The RDF-Schema Specification (Brickley, Guha, 1999), which became a W3C candidate recommendation in March 2000, is an RDF application that introduces an object-oriented, extensible type system to RDF. RDF-Schema is a minimal-

[2] http://www.unicode.org/

ist model, including primitives for representing classes, properties, subproperty, sub-classes, domain & range restrictions and means for representing comments & labels.

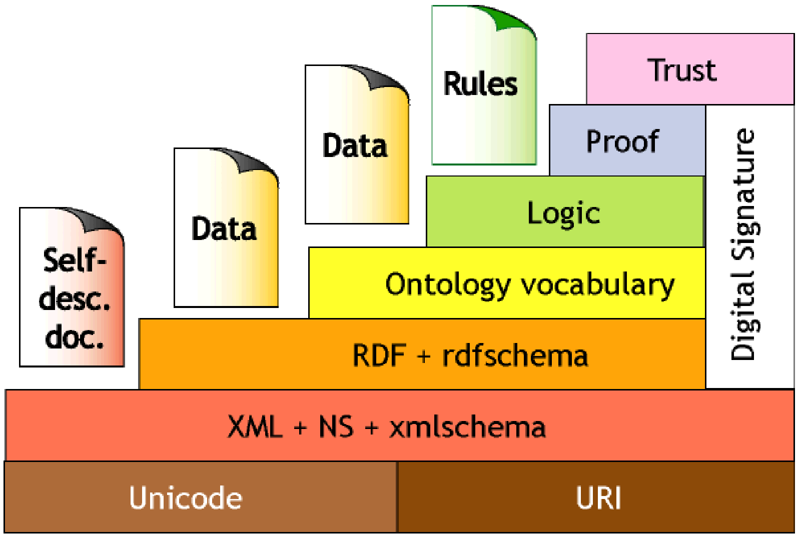

Figure 1: Semantic Web Representation Layers (cf. (Berners-Lee, 1999))

The Ontology Layer: This layer includes more complex representation primitives, such as transitive properties, cardinalities, etc.. We refer the interested reader to the recent research initiatives OIL[3] and DAML+OIL[4] that are built on top of RDF(S). E.g. OIL unifies the epistemologically rich modeling primitives of frames, the formal semantics and efficient reasoning support of description logics and is mapped to the standard Web metadata language proposals.

The Logical Layer: The logic layer consists of rules that enable inferences, e.g. to choose courses of action and answer questions. The proof layer is required to provide explanations about the answers given by automated agents that consume the provided information. Naturally, you might want to check the results deduced by your agent, this requires the translation of its internal reasoning mechanisms into an unifying proof representation language.

The Proof & Trust Layer: Proof and trust mechanisms are still to be developed. At this stage in the development of the Semantic Web, though, this problem is not tack-led. Most applications construction of a proof is done according to some fairly con-strained rules, and all that the other party has to do is validate a general proof.

---

[3] http://www.ontoknowledge.org/oil
[4] http://www.daml.org/2001/03/daml+oil-index

An example for the instantiation of the specific Semantic Web representation layers specific is given in the Figure (it is based on the actually existing web page of Paris and a specific hotel in Paris). In the upper part of the figure an ontology represented in RDF-Schema is graphically depicted. It is restricted to class definitions, that are connected via different properties, such as subclassOf, locatedIn, etc. Additionally, a rules expressing the transitivity of the locatedIn property is depicted. According to the RDF data model statements are specified for each web page and serialized using XML syntax. By defining these descriptions the web pages become machine interpretable and understandable, e.g. if somewhere else is stated that Paris is located in France and a tourist from Asia requests all hotels in France, he will receive the hotel described at the left side of the Figure, because the transitivity of the locatedIn relation is evaluated. Another advantage is that web tourism services like specific portals may be automatically generated (e.g. one may generate a “France portal” by collecting the metadata available on web servers in France).

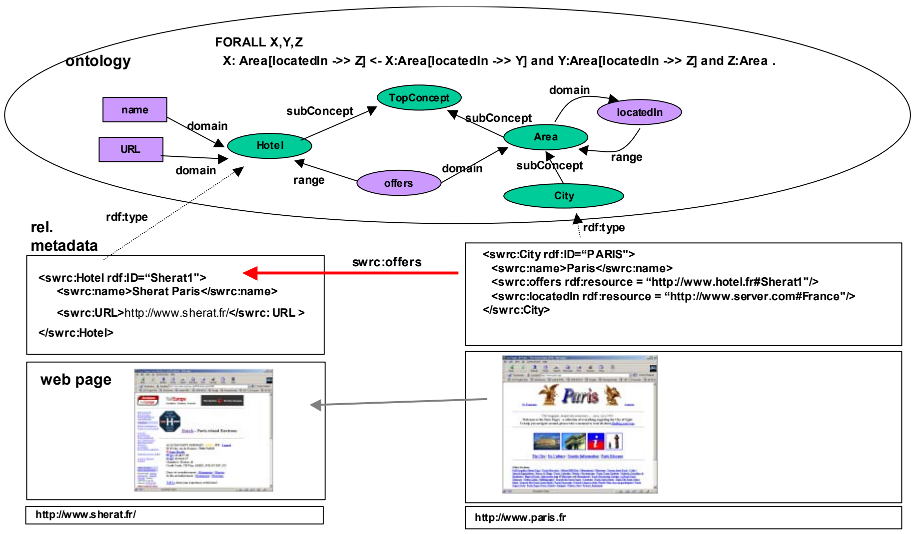

Figure 2: Example scenario for the Semantic Web

4 Semantic Web Applications

This section will elaborate on already existing methods, tools and projects that have turned some parts of the Semantic Web vision into concrete applications.

4.1 Semantic Search

Ontobroker (Fensel et al., 1999) comprises languages and tools that allow to semantically mark-up content on web pages and let the user semantically query the WWW taking advantage of semantic inferences. Ontobroker is based on the use of ontologies that guide the semantic mark-up of web documents, the querying interface and that formulate semantic rules for the domain. Thus, conventional web pages are augmented with a facility for intelligent brokering services without requiring to change the semiformal nature of web documents. Ontobroker has, e.g., been applied to needs of the knowledge acquisition community. KA^2 has been an initiative that provided semantic mark-up about researchers, academic events, etc. such that ontobroker could be used for semantic querying. Similar to Ontobroker the SHOE (Heflin & Hendler, 2000) project intended to annotate web documents with machine-readable knowledge. SHOE is a set of tools including a Knowledge Annotator, the crawler Exposé, the knowledge representation system PARKA, the PIQ (Parka Interface for Queries) and SHOE Search.

The RDFSuite (Alexaki et al., 2001) focuses on a suite of tools for RDF validation, storage and querying based on object-relational database technology. Specifically, the authors introduce a formal data model for RDF description bases created using multiple schemas. Additionally, the design of a persistent RDF Store (RSSDB) for loading resource descriptions in an ORDBMS by exploring the available RDF schema knowledge is presented. Their work also includes the definition of RQL, a declarative language for querying both RDF descriptions and schemas, and sketch query evaluation on top of RSSDB. An example application using the RDFSuite has been applied to the domain of cultural heritage.

In contrast to the systems mentioned above, within the GETESS (German Text Exploitation and Search System) project, the idea of completely automatically generating semantic markups has been pursued. For this purpose shallow natural language text processing has been combined with ontologies to generate light-weight markups for web pages. The automatically generated information has been collected and stored in a database. Finally, a user interface (also natural language driven) offered access to the semantically enriched web pages. The GETESS technology has been applied in the tourism and finance domain. A detailed introduction into the GETESS project is given in (Staab et al., 1999).

4.2 Topic Portals

The idea of ontology-based knowledge portals has been described in (Staab & Maedche, 2000), (Maedche et al., 2001). Knowledge portals provide views onto domain-specific information on the World Wide Web, thus facilitating their users to find relevant, domain-specific information. The construction of intelligent access and the provisioning of information to knowledge portals, however, remained an *ad hoc* task requiring extensive manual editing and maintenance by the knowledge portal providers. In order to diminish these efforts ontologies are used as a conceptual backbone for providing, accessing and structuring information in a comprehensive approach for building and maintaining knowledge portals. In (Staab & Maedche, 2000) two examples for running knowledge portals have been presented, namely the research case study KA2-Portal and the commercial case study, TIME2Research knowledge portal. The TIME2Research knowledge portal aims at streamlining the process that the technical analysts performs, because it allows for collaborative knowledge provisioning. The KA2-Portal has been developed on top of the KA^2 application, serving the knowledge acquisition community with relevant information about members, events, topics and publications – but for the purpose of browsing rather than for semantic search.

The COHSE (Goble et al., 2001) project defines and deploys a Conceptual Open Hypermedia Service. It consists of an ontological reasoning service which is used to represent a sophisticated conceptual model of document terms and their relationships, a Web-based open hypermedia link service that can offer a range of different link-providing facilities in a scalable and non-intrusive fashion, and integrated to form a conceptual hypermedia system to enable documents to be linked via metadata describing their contents and hence to improve the consistency and breadth of linking of WWW documents at retrieval time (as readers browse the documents) and authoring time (as authors create the documents).

4.3 Semantics-based Electronic Markets

First work on the application of matchmaking between agents has been described by (K. Sycara et al., 1999). In their work a comprehensive agent framework is described allowing to set up semantics-based electronic markets. Along similar lines (Eiter et al., 2001) have used the GRAPPA framework within the Human Resource Network (HRNet), a large-scale application for the mediation of jobs. The idea behind this application is that complex structured, semantically driven job descriptions serve as input for the open job matching task within companies.

4.4 The Transactional Web - Web Services

A first step into the direction of the transactional Web has been established within the Herakles project (Knoblock et al., 2001), where static data source structure descriptions have been combined with dynamic services. In their work they have applied the general framework for creating information assistants to build an example travel assistant. The resulting system helps a user plan out a business trip from beginning to end. Thus, starting the system, it first looks up the upcoming meetings in your calendar. After having selected a meeting it extracts the dates for the meeting, looks up the location, checks the weather, and even makes a recommendation about whether you should fly, drive or take the train to the meeting. Additionally, it makes recommendations about whether you should park at the airport or drive and park your car and about how you should get to your final destination from the airport. The underlying techniques of their approach are so-called wrappers extract information pieces from structured web pages and constraint-based reasoning for the integration of the multiple information sources, programs, and constraints.

Beyond the approach presented by (Knoblock et al., 2001) are fully automatically generated, dynamic service descriptions as described in (Fensel et al., 1999), (McIlraith et al., 2001) (Denker et al., 2001). This includes the automated Web service discovery, execution, composition and interoperation.

5 Conclusion

In this paper we have introduced the vision of the Semantic Web and its potential benefits to tourism information systems. In tourism much data is only loosely structured or given in texts – a mode that is too weak for rich querying. The Semantic Web provides additional mechanisms adequate for dealing with these structures, enabling new applications for commerce and communications, especially.

References

S. Alexaki, V. Christophides, G. Karvounarakis, D. Plexousakis, K. Tolle: The ICS-FORTH RDFSuite: Managing Voluminous RDF Description Bases. In Proceedings of the Second International Workshop on the Semantic Web - SemWeb'2001. Hongkong, China, May 1, 2001, pp. 1-13. CEUR Workshop Proceedings. http://ceur-ws.org

T. Berners-Lee: Weaving the Web. Harper, 1999.

D. Brickley, R. Guha. Resource description framework (RDF) schema specification. Technical report, W3C. W3C Proposed Recommendation, 1999. http://www.w3.org/TR/PR-rdf-schema/.

G. Denker, J. R. Hobbs, D. Martin, S. Narayanan, R. Waldinger: Accessing Information and Services on the DAML-Enabled Web. In Proceedings of the Second International Workshop on the Semantic Web - SemWeb'2001. Hongkong, China, May 1, 2001, pp. 67-77. CEUR Workshop Proceedings. http://ceur-ws.org

T. Eiter, D. Veit, J. Müller, M. Schneider: Matchmaking for Structured Objects. In: Y. Kambayashi, W. Winiwarter, M. Arikawa (Eds.): DaWaK 2001, Berlin: Springer, LNCS 2114, pp. 186–194, 2001.

D. Fensel, J. Angele, S. Decker, M. Erdmann, H.-P. Schnurr, S. Staab, R. Studer, A. Witt: On2broker: Semantic-based access to information sources at the WWW. Proceedings of WebNet, 1999, pp. 366-371.

D. Fensel, V. R. Benjamins, E. Motta, B. Wielinga: UPML: A Framework for Knowledge System Reuse. Proceedings of IJCAI 1999, Morgan Kaufmann, pp. 16-23.

C. Goble, S. Bechhofer, L. Carr, D. De Roure, W. Hall: Conceptual Open Hypermedia = The Semantic Web? In Proceedings of the Second International Workshop on the Semantic Web - SemWeb'2001. Hongkong, China, May 1, 2001, pp. 44-50. CEUR Workshop Proceedings. http://ceur-ws.org

J. Heflin, J. A. Hendler: Dynamic Ontologies on the Web. Proceedings of the AAAI/IAAI 2000. AAAI Press/MIT Press, pp. 443-449.

S. Handschuh, S. Staab, A. Maedche: CREAM — Creating relational metadata with a component-based, ontology-driven annotation framework. To appear in: Proceedings of the First Semantic Web Working Symposium SWWS, Stanford, USA, 2001.

C. Knoblock, S. Minton, J. Ambite, M. Muslea, J. Oh, M. Frank: Mixed-Initiative, Multi-Source Information Assistants. WWW-10 – Proceedings of the Tenth International World Wide Web Conference. Hong Kong, May 1-5, 2001. ACM Press, 2001, pp. 697-707.

O. Lassila, R. Swick. Resource description framework (RDF) model and syntax specification. Technical report, W3C. W3C Recommendation. http://www.w3.org/TR/REC-rdf-syntax, 1999.

A. Maedche, S. Staab: Ontology Learning for the Semantic Web. IEEE Intelligent Systems, 16(2), 2001.

A. Maedche, S. Staab, R. Studer, N. Stojanovic, Y. Sure: SEmantic PortAL - The SEAL Approach. In: Creating the Semantic Web. D. Fensel, J. Hendler, H. Lieberman, W. Wahlster (eds.) MIT Press, Cambridge, MA.

S. McIlraith, T. Son, H. Zeng: Semantic Web services. In IEEE Intelligent Systems, 16(2), March-April 2001, pp. 46–53.

S. Staab, A. Maedche: Knowledge Portals — Ontologies at Work. In: AI Magazine, 21(2), Summer 2001.

S. Staab, C. Braun, I. Bruder, A. Düsterhöft, A. Heuer, M. Klettke, G. Neumann, B. Prager, J. Pretzel, H.-P. Schnurr, R. Studer, H. Uszkoreit, B. Wrenger. GETESS - Searching the Web Exploiting German Texts. CIA '99 – Proceedings of the 3rd international Workshop on Cooperating Information Agents. Upsala, Sweden, July 31-August 2, 1999, pp.113-124, LNAI 1652, Berlin, Heidelberg: Springer, 1999.

G. Stumme, A. Maedche: FCA-Merge – Bottom-Up Merging of Ontologies. In: Proceedings of the 17th Int. Joint Conference on Artificial Intelligence (IJCAI'01), Seattle, USA, 2001, San Francisco/CA: Morgan Kaufmann.

K. Sycara, M. Klusch, S. Widoff, J. Lu: Dynamic service matchmaking among agents in open information environments. ACM SIGMOD Record, 28(1), March 1999, ACM Press, pp. 47 – 53.

F. Ygge, H. Akkermans. Power load management as a computational market. In Second International Conference on Multi-Agent Systems, pages 393-400, 1996.