Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 3 / ISSUE 2 / AUGUST 2008 / 82-96
© 2008 Universidad de Talca - Chile

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762008000100008

Enhancing Hotel Search with Semantic Web Technologies

Magnus Niemann[1], Malgorzata Mochol[2] and Robert Tolksdorf[3]

Free University of Berlin, Networked Information Systems, Königin-Luise-Str. 24-26, D-14195 Berlin
[1] maggi@inf.fu-berlin.de, [2] mochol@inf.fu-berlin.de, [3] tolk@inf.fu-berlin.de

Received 12 October 2007; received in revised form 15 February 2008; accepted 20 May 2008

Abstract

Tourism service providers are more and more under pressure to offer products of greater complexity and diversity to meet the ever-changing demands of travelers; the individualistic consumption patterns and lifestyles makes it increasingly difficult for tourist service providers to anticipate consumer behavior and configure their services accordingly, i.e. the tourist industry must focus more on a "hybrid consumer" whose travel choice will be more complex. Although current online travel systems aim to support the customer in finding a suitable hotel or even a whole trip, most of the work is still up to the customer, who has to consider several sources of information before deciding which hotel to book. Furthermore, since the quality of a hotel room w.r.t the requirements of the end-user are multi-dimensional and cannot be easily expressed on discrete scales, the main critical issue in such cases is a price/benefit ratio which is defined by what is known, as the "best" booking. To tackle these problems an advanced search technology that considers the ratio and ranks results accordingly to the user requirements is needed. In this paper we propose a framework which uses Semantic Web technologies for an improved exploration and rating of hotels for business customers in order to reduce the search time and costs, which, in turn, results in a huge benefit for the end-users. The framework provides methods for modeling domain specific expert knowledge and integration of diverse heterogeneous data sources. Semantic technologies enable business customer to formalize their requirements and to combine those requirements with aggregated hotel information like location or features, thus achieving a selection of the hotels ranked according to the customer's requirements.

Key words: e-Tourism, Hotel search, Hotel evaluation framework, Semantic Web, *Reisewissen*

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 3 / ISSUE 2 / AUGUST 2008 / 82-96
© 2008 Universidad de Talca - Chile
This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762008000100008

1 Introduction

Travel is a domain in which the Internet has led to a new quality of service and online booking and reservation services have become widely accepted among consumers and business travelers. Furthermore, in recent years, growth rates in online tourism have moved much faster than in the overall world economy, and this trend is not expected to slow down in the near future [22]. Since travel destinations can be easily checked out in advance, it has become much simpler to choose hotels with a higher degree of precision. Particular hotels on the Web are presented with a variety of visual and factual information. In addition, the end-user may access the hotel's reservation system by entering the travel dates and getting an immediate response on the availability. Since the variety of information offered leads to a longer manual search, the age of services has become reality. There is a variety of services that integrates the information scattered across various sites, federate multiple structured and semi-structured tourism information sources on the web [10], and offer search engines for hotel rooms by providing a list of rooms available for a given period at a particular place. Often these search engines utilize databases and reservation systems to which hotels are connected. Having such an engine that provides integrated information regarding hotel vacancies clearly reduces the search time and costs, which, in turn, means a huge benefit for the end-user. At the same time, the search facilities allow for requests on quantifiable hotel data such as price, number of stars and proximity to locations of interest, whereby the search results can usually be ordered along one of the dimensions.

However, the quality of a hotel room w.r.t the requirements of the end-user are multi-dimensional and cannot be easily expressed on discrete scales. The critical issue in such cases is a *price/benefit ratio* that is defined by what is known, as the "best" booking, i.e. the cheapest hotel room is not necessarily the best. However, this effort is added to the search costs, since one has to check a large number of matching results manually to determine the room with the best *price/benefit ratio*. To face this problem an advanced search technology that considers the ratio and ranks results accordingly is needed. In such a ranking, the hotel given the first place represents the most suitable one and is not necessarily e.g. the hotel with the highest number of stars. Common search technologies utilize databases on which queries selecting only e.g. hotel offerings for a specific date with a price below 100 euros. Furthermore, they might also use some full-text index regarding specifications of the offers e.g. searching for terms like "breakfast", to consider criteria that are not in the fields of the database. The problem is that such approaches are incapable of responding to queries with a *price/benefit ratio*. To tackle such issues mechanisms are required to determine how close a given criterion is to a customer's needs. Unfortunately, neither a database query capable of handling only distances on discrete scales nor a full-text index that considers only syntactical (e.g. by stemming) or statistical distance (e.g. in a vector space) can manage this issue. A possible solution can be found within the Semantic Web initiative which provides technologies capable of enabling comparisons at a semantic level.

1.1 Semantic Web and e-Tourism

Over the last two decades the World Wide Web has rapidly evolved into a vast repository containing huge amounts of decentralized information on all matters of interest. However, with the ever increasing number of informational sources on the Web new problems are rising apace. The main challenge nowadays is to find, integrate, and process all the (available) information relevant to a particular use context. Since most of the Web's content is primarily designed to be read by humans, machines can parse Web pages for layout however, they cannot automatically process data from a particular Web site without understanding its semantics. The Semantic Web vision is here to solve this problem. It extends the current Web with information that provides well-defined meaning, thus enabling computers to return more precise search results, integrate data from different sources, and automate sophisticated tasks [2]. The objective is to use the Web like a global distributed knowledge store which can be accessed by applications.

The development of the Semantic Web is a joint effort of top scientific institutions (MIT, Stanford, ILRT etc.) and global business players (HP, IBM, Nokia etc.) [20], and is led by the World Wide Web Consortium (W3C) (Site 1) whose task is to oversee the major efforts of specifying, developing, and deploying standards and languages with the aim of expressing shared meanings. These new Internet technologies (Semantic Web technologies) are maturing and moving out from academic applications into the industry. This is demonstrated, on the one hand, by strong and growing interest in these topics by various commercial sectors and, on the other hand, by public bodies like the European Commission which supports the distribution and transfer of these technologies to the business world. One such activity is the Knowledge Web EU Network of Excellence (Site 2) which has formed an Industry Board (Site 3) to promote greater awareness and faster uptake of Semantic Web technology into European industries. The Knowledge Web together with this board aims to transfer technology from research to industry, promote ontological technologies, propose technological recommendations, and meet industrial application needs. Within this in mind certain sectors has been identified for the initial uptake of Semantic Web technologies [14] – this confirms the capability of e-tourism's application of semantic technologies and its potential to be an early adopter.

Considering tourism domain semantic technologies allow both hotel offers and requirements at a conceptual level to be described by attaching metadata to datasets that refer to parts of ontology. Ontology is an explicit specification of a shared conceptualization [9] and provides a description of the area of interest, which in our case means hotel characteristics and customer (travel) requirements. The application of ontologies and semantic technologies allow

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762008000100008

the inference, e.g. that a pool in a hotel contributes to a "spa" criterion though the word "spa" is not present in the hotel's description. When the inference is used to re-rank search results it allows "better" sorting the offers by taking into account the *price/benefit ratio* of the customer and delivering hotels (or rooms), which are more precisely fit to the customers' requirements. In this context and to take full advantage of the modern convenience of electronic business-to-business and business-to-customer trading, a number of solutions based on Semantic Web technologies have been already proposed (e.g. TIS framework [1], SATINE (Site 4) [5], [6], Harmonise (Site 5) [7], [13], OnTour (Site 6) [19]). The general goal of such projects is to semantically connect, organize and share currently isolated pieces of tourism information in order to enable better interoperability and integration of inter- and intra-company travel information systems, facilitate the user to find [12] and understand the information sources as well as to allow for individual use of travel offers.

1.2 Overview

This paper describes the results of the project *Reisewissen (Site 5)* whose aim is to develop a system to evaluate the extensive and constantly changing range of goods for the customer. Such evaluations intend to enhance the quality of a product query by considering not only the individual customer profile but also domain specific knowledge. Both domain specific knowledge and profiles, which are described in terms of ontologies, are handled using Semantic Web technologies. The main use case in the project comes from the e-tourism domain and covers searching and booking of hotels. To deal with the above-mentioned problems in terms of e-tourism use case, we propose a hotel evaluation and recommendation engine that uses Semantic Web technologies to enhance the quality of an existing hotel search engine. With the *Reisewissen* approach we aim to optimize the hotel selection process, raise the quality of the travel services, save travelers' time and significantly reduce the direct as well as indirect travel costs.

The rest of the paper is organized as followed: Sec. 2 gives a brief overview regarding the study of the common use cases in the travel domain which serves as a basis for the requirements analysis of the system. In Sections 3 and 4 we concisely describe the technical background of our semantic hotel search and introduce the framework which has been developed within the project *Reisewissen*, respectively. The article concludes in Section 5.2 with the experiences and insights gained in the course of developing the prototype as well as a vision of the further system development.

2 Requirement Analysis

Several decades of research and development in Software Engineering, one of the core disciplines of Computer Science, have shown that software projects cannot be reduced to software implementation. Rather they should focus equally on user requirements, design, testing, documentation, deployment and maintenance. To this end, the first step towards building a system or application is a system or domain analysis, whose task is to come up with a set of requirements definitions. To address the actual requirements of the travelers regarding the new hotel search and booking system we must study the common use cases in the travel domain which will ultimately serve as the basis for the system's requirements analysis of the system. Since the *Reisewissen* prototype of the hotel search portal based on semantic technologies has been developed in close collaboration with the industrial partner *ehotel AG* (Site 6) we were able to analyze a number of use cases regarding the actual demands of the *ehotel AG* customers and define the industry-based requirements.

2.1 Scenarios

To implement a hotel recommendation engine the requirements of the system depend on the meaningful use cases derived from the everyday business practices of the travel business of the *eHotel AG*. In the following some of the most relevant use cases are briefly presented to provide insight into the problems of searching for suitable hotel accommodation. These use cases show the complexity and the heterogeneity of data the system must deal with. One of these cases has been depicted in detail and will serve as a continuous example to show the functionality of the system developed.

- **1st Use Case – Tourist (traveling alone)**: Sebastian is traveling to London for the weekend. He wants to stay for three nights, from Thursday to Sunday. He plans to go sightseeing (museums, Buckingham Palace, London Eye) during the day and in the evening he would like to have dinner (preferably Indian or Chinese) and some beer in a pub. After that, Sebastian would like to visit a theatre or one of the "in" clubs in Soho. The hotel should be moderately priced and not too far away from an underground station. Preferred areas: Soho, Kensington.

- **2nd Use Case – Family (Three nights in London, arrival by car)**: On the way to their holiday in Scotland the Kuhn family (two adults with a seven-year-old daughter) wants to stop in London for sightseeing. They arrive via ferry in Dover. Since they are not used to the left-hand traffic they do not wish to stay in the city centre. Another important aspect is a parking space next to the hotel as well as a subway connection to the city. Preferred locations: south/southeast London off of the M20 motorway. Price limit: GBP 80/night, with cot.

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 3 / ISSUE 2 / AUGUST 2008 / 82-96
© 2008 Universidad de Talca - Chile
This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762008000100008

- **3rd Use Case – Business trip (business man)**: Mr. Yamamoto, European sales manager of a large Japanese car manufacturer, wants to organize a business meeting in London. Coming from Japan, he will arrive in the late afternoon, on December 14 at Heathrow airport and will stay for two days. Mr. Yamamoto prefers traditional hotels (luxury class) that nonetheless cater to the individual (less than 300 rooms). The meeting will not be held at the London's company headquarter but at Mr. Yamamoto's hotel. A meeting room for eight persons must be booked for the day after arrival. The European colleagues will arrive on December 15. Travel budgets in the European offices have been reduced, so that those arriving from Denmark, Germany and France have a maximum of GBP 120/night at their disposal. Consequently, they cannot stay in the same hotel as Mr. Yamamoto and need accommodation in a nearby hotel.

- **4th Use Case – Business trip (business woman)**: Ms B. is planning a business trip to London. On Tuesday, September 19 a three-day meeting will be held at the company headquarters. She will arrive on September 18 by plane from Frankfurt (destination - London Heathrow). The office is located in Kensington (Queens Gate Terrace). Company travel guidelines oblige the employees to use public transport rather than taxis whenever possible. Furthermore, the price limit for a single room in London should be no more than GBP 140/night including breakfast (Miles&More bonus card is available). Due to her late arrival, Ms B. places great importance on short transfer times. Beside this, she would like to have Italian or Chinese restaurants near the hotel or if this is not possible a hotel with its own restaurant. She also wants to be able to send the final version of her presentation to her colleagues via e-mail. To relax after a stressful day Ms B. would like to go swimming or to a sauna, preferably Finnish.

All these uses cases have been taken into consideration in the course of defining the system requirements while the final scenario serves as a prime example and continuous thread through this work.

2.2 Requirements

The requirements are models of the problem definition and requirement definitions simply mean "figuring out what to do before doing it". In other words, requirements definition enables appropriate decisions to be taken regarding the functionality and design of a product before time and money are invested to develop it. By bridging the gap between the needs of the market and those of the particular organization, requirement definition significantly reduces guesswork in technology product planning and helps to ensure that both business and engineering are working on the same page [16].

The analysis of the given use cases and interviews with the domain experts (travel managers, travel agents, booking services, etc.) provide the basis for the requirements of the framework to be implemented. In this section some of the crucial requirements, starting with the common and general (non-functional requirements) and followed by the specific (functional requirements) derived from the example scenarios are described.

2.2.1 Functional requirements

Reconsidering the use cases mentioned before and preliminary considering the "Business Woman" case we have identified the following functional requirements as relevant for the framework:

- **User Profile**: Instead of a simple city/date query, Ms B.'s profile of needs and wishes regarding a hotel has to be captured and formalized by the system. The profile should be constructed hierarchically and allow to capture a broad range of relevant data like price limits, food preferences and hotel/ room amenities (sauna, spa). Despite the user profile's complexity, the profile creation process must follow usability requirements. Users should be able to weight their hotel preferences, e.g. express factors like "I need business features as well as spa in the hotel, though business facilities are more important to me."

- **Data Sources**: First of all, the framework must have access to hotel data, e.g. address, room rates and information which maybe used to match a hotel to Ms B.'s profile. Since room rates and availability change frequently, the framework must poll this information at query-time from up-to-date sources like web services. Furthermore, the framework must allow for the integration of additional data sources like e.g. geo-coded data or transportation information. This information can be used to evaluate a hotel's location regarding nearby restaurants and public transport.

- **Results**: Following the non-functional requirement of "transparency" the query results, i.e. the detected and ranked (w.r.t. the suitability to Ms B.'s profile) hotels must be presented to Ms B. in a manner that clarifies the origin of a specific result. Most actual vertical search engines using matching technologies to provide optimal results do not explain the results to the user. The framework must keep track of the hotel evaluation and allow insights into the evaluation process.

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 3 / ISSUE 2 / AUGUST 2008 / 82-96
© 2008 Universidad de Talca - Chile
This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762008000100008

2.2.2 Non-functional requirements

In addition to the functional requirements, several non-functional requirements are to be taken into account when developing a platform like ours. The fulfilling of the following five non-functional requirements is a crucial task in the course of the development of the framework:

- **Openness**: The system should define extendable interfaces to allow easy addition of new data sources and expert knowledge in terms of rules.
- **Efficiency**: Due to the fact, that most of the hotel evaluation is done at query time, the algorithms and data structures must be selected for efficiency while avoiding over-specialized structures and functionalities.
- **Reliability**: The framework must be stable and reliable, e.g. inured to sudden loss of database or web service connections.
- **Transparency**: The overall process of hotel selection and evaluation must be transparent to the end user.
- **Support for domain experts**: The extension of the framework should allow domain experts (e.g. travel managers) to add their knowledge to the framework without having to rewrite current code. The framework itself should be independent of domain-specific functionality, which would allow it to be applied to other domains as well, e.g. the planning of complete travel solutions including transportation.

Taking into account the use cases and the derived requirements mentioned in the functional requirements, we have developed a hotel evaluation framework, which is specified in the next section. Since the fulfillment of the non-functional requirements.

3 Technical Realization of the Semantic Hotel Search

In the *Reisewissen* project, a prototypical framework for (but not necessarily restricted to) a semantic hotel search has been specified and implemented. The incorporation of an industry partner as a domain expert into the development process allows the building of a framework which fulfills the real world requirements of the users of the travel sector. In the following we describe the data sources used within the framework which serve e.g. as a basis for the development of the appropriate ontologies and then specify the architecture of the hotel evaluation framework together with the brief description of the particular components.

3.1 Data Sources

The need for comprehensive classification systems in tourism domain has been recognized early by many interested parties. Whereas the open interoperability specifications are not new in e-tourism domain and in particular the XML schemas of the Open Travel Alliance (OTA) (Site 7) are in widespread use, the existing data exchange formats are, however, not expressive enough to allow automatic exchange and processing of information to develop dynamic applications. To tackle such issues and considering the high interoperability requirements in the hotel information and booking systems that is compounded by the complexity of the travel domain, various heterogeneous knowledge sources have been utilized in the project.

In our opinion, the data exchange between hotels, travel portals and users (travelers and travel managers) should be based on a set of vocabularies which provide shared terms for describing relevant tourist information. Due to this fact the relevant information means not only general hotel description like address, price or room availability for a particular period of time but also all information directly associated with the traveler and type of a trip (e.g. business or holiday, with or without children, traveling by car or other types of transport). A useful travel portal should deliver detailed description of the hotels while taking into account, e.g. room amenities, technical facilities and surroundings, important points of interests as well as recreational facilities, and be able to deliver information about modes of transport, or if necessary car rental possibilities.

In the following we describe the different data sources available to our travel portal, the challenges and their treatment within the framework.

3.1.1 Hotel Data

The main data source in the project contains information of about 60.000 hotels worldwide. This information is divided into *static data*, stored permanently in a relational database, and *dynamic data* which is constantly updated and obtained from a web service. The static hotel data comprises of less frequently updated information like address, size, hotel features (e.g. restaurant, spa, massage service, wireless internet access) and room layout. The dynamic part comprises current book rates and availability – this data is obtained at the time of the customer query.

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 3 / ISSUE 2 / AUGUST 2008 / 82-96
© 2008 Universidad de Talca - Chile
This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762008000100008

Unfortunately none of this data is in a format usable by Semantic Web technologies while current hotel data lacks even a decent structure, encoding important information in free text data fields.

The biggest challenge for a semantic hotel search portal is a reasonable processing of the available data. The framework must provide the data in a Semantic Web compatible format while keeping the efficiency of the original data to guarantee fast querying. Wherever possible, the hotel information in the relational database will be augmented with aggregated information, e.g. how well a hotel fits to a certain category or how is the location of the hotel. Moreover, the dynamic hotel data will be cached to allow for fast access.

3.1.2 Location-based Data

Since one of the most important information to the customer about a hotel is – besides the hotel's internal information – the location of the hotel and surrounding points of interest (POIs), the second important data source is city-specific location-based data. Considering the scenario mentioned before, the framework has to take into account e.g. restaurants in the vicinity of the hotel according to categories (e.g. Japanese, Italian) together with other location-based data including e.g. information regarding WLAN hotspots, museums or specific neighborhood or districts.

For the London scenario, the project utilizes a free online travel guide, Open Travel Guides (Site 8), where users can add their description and rate interesting locations in London. Since this guide makes use of Resource Description Framework (RDF) (Site 9) metadata schemas for a machine-readable description of locations, including information like geo-coordinates, categories and addresses, the data could be used to enhance the semantic-based queries (in the next section we describe how this metadata is optimized to allow for a fast semantic matching). Additional geo-coded information provided by *ehotel AG* in the form of a constantly updated database of POIs has to be transformed explicitly into a semantically useful format.

The time-consuming task of processing geo-coded data, e.g. looking for Japanese restaurants near a hotel, leads to new challenges in caching the location based information. The project uses a two-dimensional, tiled caching of sorts for POI information.

3.1.3 Other Data Sources

The framework provides open interfaces to allow quick integration of new data sources. In the preliminary version of the system only a German hotel guide with more detailed hotel information has been integrated. To extend the prototypical system and, in turn, transform the hotel search portal into a concise travel guide additional data sources covering e.g. public transport information, timetables and prices can be integrated.

3.1.4 Semantic vs. Non-Semantic Data

The hope of the Semantic Web community for a vast amount of metadata-augmented web pages, especially in tourism, has given way to the insight that most of the available hotel information does not even come in a structured format. Thus, the project objects to enabling domain experts to provide arbitrary proprietary data as RDF metadata and, from this point, using semantic technologies. The ontologies developed for this semantic processing will be described in the following section.

3.2 Ontologies

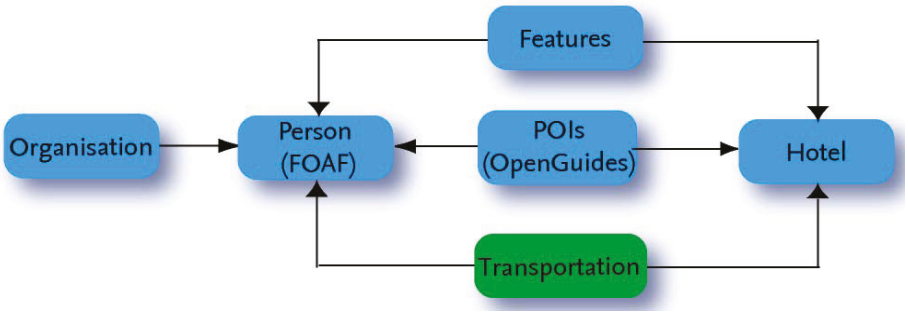


Figure 1: *Reisewissen* ontologies

Since we have been cooperating with a partner company from the tourism sector we have been able to conduct interviews with application and domain experts to find out what kinds of information are relevant for a user of a hotel booking portal. In addition to the use cases (cf. Sec. 2.1) the domain experts specialized on business travel and hotel booking systems have defined sub-domains of the application setting by specifying six main groups of information as relevant for the hotel recommendation engine: *general hotel information*, *hotel amenities*, *points of interest*, *public transportation*, *detailed user profile* and *company-specific travel guidelines*. Considering these groups and following

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 3 / ISSUE 2 / AUGUST 2008 / 82-96
© 2008 Universidad de Talca - Chile
This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762008000100008

the core requirements derived from the use cases w.r.t data sources, system functionality and results to be achieved in the research project (cf. Sec. 2.2), we decided to build two main ontologies: *person* and *hotel*, and small sub-ontologies which describe hotel *features*, *points of interest*, and means of *transportation* (cf. Figure 1).

3.2.1 Points of Interest ontology

Since location-based information plays a crucial role in the hotel search process we have taken into account geographical information regarding points of interest (POIs). In order to fully resolve the use case – London hotels with Italian or Chinese restaurants in the vicinity (cf. Sec. 2.1) – data from Open Travel Guide, where POIs information is already formalized in RDF, has been utilized. As the existing POIs classification was collaboratively developed by users of the Travel Guide portal, it resembles more a *folksonomy* than a well-defined, formal ontology. In consequence, we have had to overcome some difficulties in adapting the classification into our system. In order to tackle problems such as synonyms, plural and singular versions of the same concept, typing errors, etc., the folksonomy has been translated by using *Prolog rules* (which was defined after a manual analysis of the classification) to attain a "clean" POIs ontology.

3.2.2 Features ontology

Since travel experts are involved in the project, the researchers have gained insight not only into real world data but also certain domain specific terms and nomenclature. Based on this domain knowledge we created a feature ontology which describes general hotel features as well as room amenities (e.g. wireless internet access in rooms, handicapped accessible amenities, indoor pool).

3.2.3 Passenger transportation ontology

To represent the preferred or available transport means transportation ontology needs to be developed. To cover all means of transportation or to merge heterogeneous information sources from various cities in different countries is a nontrivial task and still "work in progress".

3.2.4 Person ontology

One of the two main ontologies is the person ontology which describes travelers and their preferences (e.g. elevator in hotel, Japanese restaurant nearby) together with company travel guidelines (e.g. London: max. GBP 100/night including breakfast) by utilizing terms from transportation, POIs, and features sub-ontologies (cf. Figure 2). In order to shorten the input time by the user of the booking system, thus to speed up the reservation process, some pre-defined general, company as well as person specific profiles have been created.

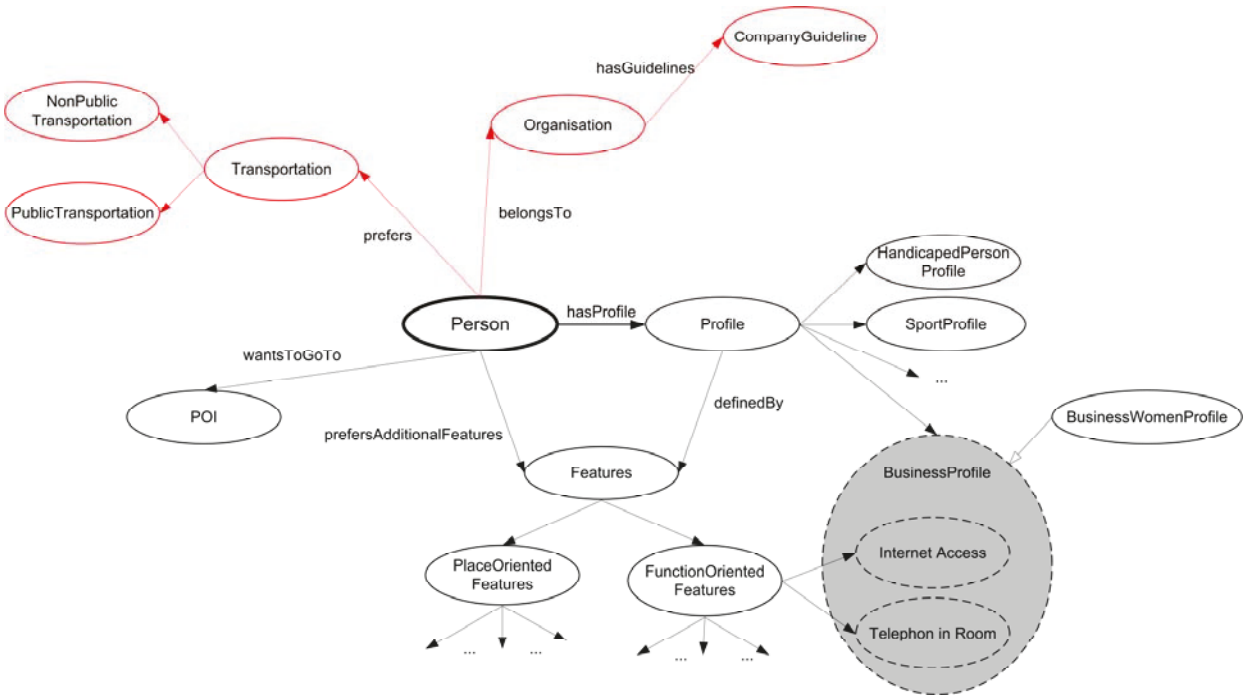


Figure 2: Partial diagram of the person ontology

3.2.5 General profiles

One example of this profile category is a general business profile including concepts from the features ontology like internet access available, telephone in room, and parking facility. A businesswomen profile could contain additional

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 3 / ISSUE 2 / AUGUST 2008 / 82-96
© 2008 Universidad de Talca - Chile
This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762008000100008

amenities like spa section, etc. Ms B. as a businesswoman (cf. Sec. 2.1) could simply choose one of these profiles already provided by the system instead of taking her time to create a personal profile from scratch.

3.2.6 Company specific profiles

Since the partner company involved in the research project is specialized on business trips most of the use cases presented in this paper focus on business people. The majority of companies (especially large ones) have their own travel guidelines regarding price limits, hotel features, transportation means, etc., which must be compiled with the booking process. These guidelines are semantically described in company profile(s), which can be reused for hotel booking by all company members, without the need to provide all the information every single time. Hence Ms B.'s company profile would contain the price limit for London (GBP 140/night including breakfast) and refer to local transportation since company travelers are restricted to the use of public transport (cf. Sec. 2.1).

3.2.7 Personal profiles

Since Ms B. (cf. Sec. 2.1) wishes to relax after a stressful day in a swimming pool or sauna (preferably a Finnish sauna) these two amenities are of great importance to her. To express such preferences, which, in this case, are not specified in either of the abovementioned profile types, she can complement this missing information by additionally creating her own personal profile. Furthermore, she is also given the opportunity to assign these amenities different "weights" with respect to her preferences (e.g. 70 points out of 100 for a swimming pool and 30 for a Finnish sauna). Later on, these weights would influence the ranking of best matching hotels. The three profile types altogether create a so-called main profile for a particular trip which is subsequently used in the matching process in order to provide a ranked list of hotels that best fulfill the requirements specified by the user.

3.2.8 Hotel ontology

As mentioned before, the second main ontology developed is the hotel ontology (cf. Figure 3) that is used as a foundation for semantic representation of a wide range of hotel (related) information. It encompasses terms and concepts for expressing contact data (address, phone, etc.), general hotel information (number of rooms and floors, check-out time, models of payment, etc.), price information w.r.t. room categories, as well as hotel ratings. Moreover, by utilizing concepts from features, POIs, and transportation sub-ontologies, general hotel descriptions are enhanced with information about hotel amenities, nearby points of interest, and public transportation (including detailed information concerning airports in the vicinity). Furthermore, the ontology incorporates location based information in the form of geo-coordinates.

4 *Reisewissen* Framework

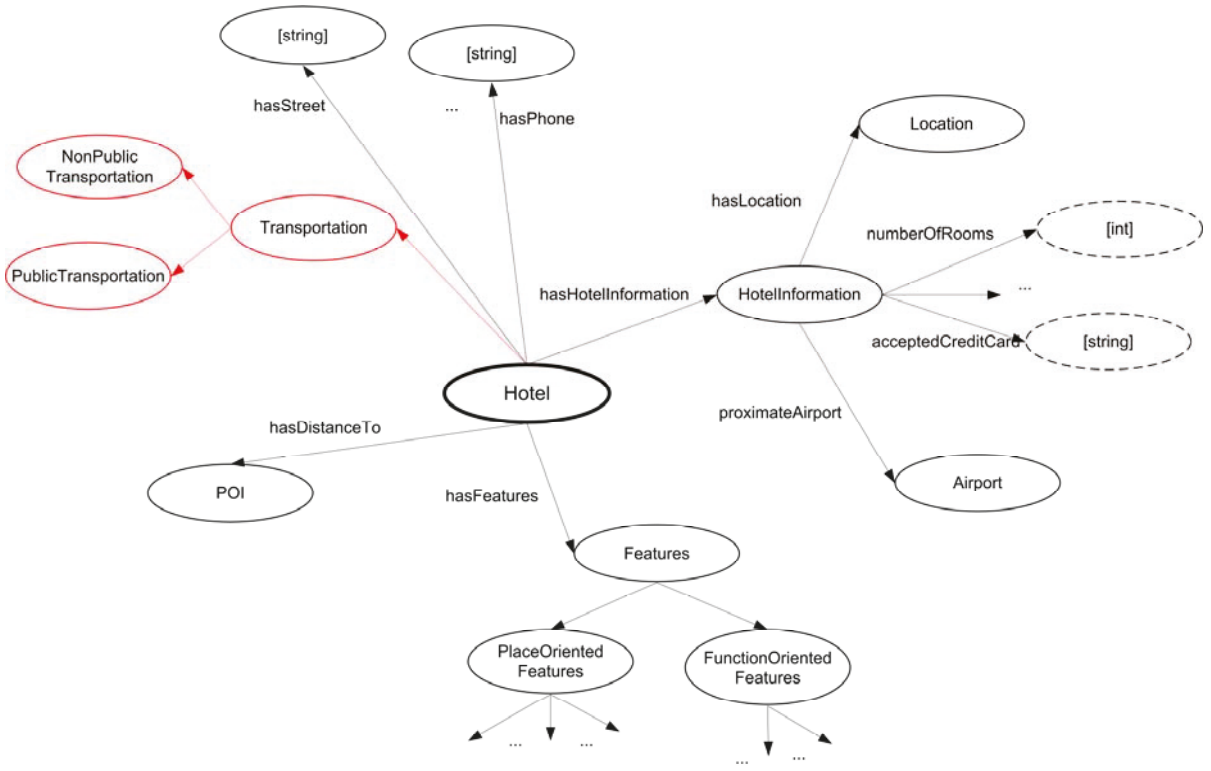


Figure 3: Partial diagram of the hotel ontology

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 3 / ISSUE 2 / AUGUST 2008 / 82-96
© 2008 Universidad de Talca - Chile
This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762008000100008

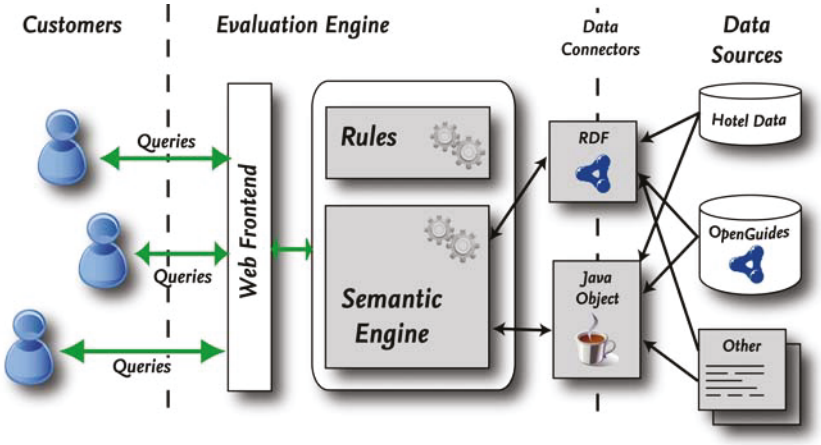


Figure 4: *Reisewissen* Evaluation Framework

The main task in the *Reisewissen* project consists of implementing a prototypical framework to be used within a semantic hotel search engine. Figure 4 shows the main architecture of the framework together with its potential customer and data sources which form the foundation for hotel and other relevant information. With the help of data connectors (which allow for queries on a data source), the evaluation engine accesses the heterogeneous information in a confirmative manner way and enriches it by using expert knowledge and rules. Customers define their profiles by specifying their requirements and wishes w.r.t. the accommodation, which steers the evaluation of the hotels. The semantic engine for the evaluation of the hotel resources returns a ranking of hotels with detailed information regarding the degree of the hotel/profile matching, which then provides a basis for further refinement and re-evaluation. The entire hotel evaluation framework is implemented in Java, while the rule and complex evaluator bits in Prolog. For the implementation of the Semantic Web-specific technologies we have made use of the Jena Framework (Site 10). The framework is divided into two parts: *domain-dependent* and *domain-independent*. The latter contains generic evaluation functions and aggregators, as well as the evaluation engine, while a domain dependent part provides the specific implementations of evaluation functions for the travel and hotel domain. Thus the framework may be used in completely different application and domain without big adaptation just by changing the domain dependent part. In the following, the main parts of the evaluation will be (mostly) described in the context of the given scenario.

4.1 Data Connectors

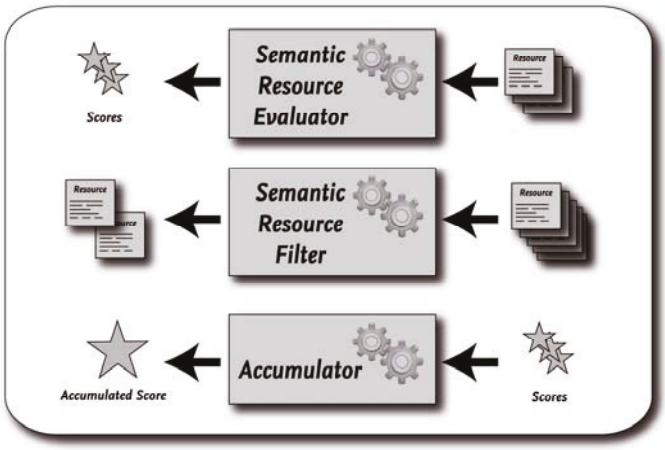


Figure 5: Semantic Engine

Each data source is linked to the framework using a data connector, the main task of which is to provide query capability for the data source. The query results will be returned either as RDF metadata or as simple Java objects. In the case of RDF the data is an instance of the according ontology (hotel features, POI etc.). The transformation from data in relational databases is done using the D2RQ framework [3]. In the Java case the data structures returned are optimized for efficiency, so that for example geo-location references like distances or the number of Chinese restaurants in the surrounding of a hotel can be computed efficiently. Besides those transformations, a data connector is responsible for the caching, adapted to the responding data source, e.g. for the caching of POIs in a certain area of town, a tiling cache mechanism is used. Furthermore, for the hotel data, only the frequently updated

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 3 / ISSUE 2 / AUGUST 2008 / 82-96
© 2008 Universidad de Talca - Chile
This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762008000100008

information like rates and availability are obtained from a web service, whereas all static information is kept in the local database.

4.2 Evaluators

The core of the semantic hotel evaluation framework is composed of a set of evaluators that represent a function operating on a hotel resource and the engine to compute these evaluators. This evaluator function is parameterized with information from the customer profile, e.g. what is the upper price limit for a hotel or what kind of spa facilities are preferred by the customer. The function utilizes the hotel and additional information to compute a score for that hotel according to the customer profile. The score represents the degree of matching between the customer's wishes and specific constraints (e.g. price, location, facilities) and a hotel. The resulting value of this evaluation function is either of a Boolean type, which means that the evaluator is a constraint, or a floating point value ranging between 0 and 1 – such evaluators are called *raters*. The evaluation engine aggregates the results of the basic evaluators defined by the customer profile and computes an overall score for every hotel, while keeping track of the single evaluator scores for later reviewing and new weightings. A visual description of the semantic engine is offered in Figure 5.

4.3 Constraints

Constraints limit the available resources (hotels) and, thus, model hard conditions for the selection of a hotel. A constraint utilized in the given scenario would be a PriceConstraint parameterized with the maximal rate that Ms. B is poised to pay for a hotel room – PriceConstraint will yield 1 if an available hotel room costs less than the maximum rate. Further constraints would restrict the available hotels to those of a certain hotel chain or those providing a certain facility like a sauna.

4.4 Raters

A rater is used to evaluate a hotel against a specific part of the customer profile, e.g. in our scenario a semantic rater for hotel features is applied. This rater is parameterized using a part of the customer profile like "wireless internet access in room" or "Finnish sauna" and attributed the appropriate weights (e.g. "business features 20 %, spa features 80 %"). The semantic rater will use the hotel's information about the features offered and compute a matching score between 0 and 1 for the hotel.

4.5 Accumulators

To aggregate the results of several evaluators we apply accumulators. An accumulator is a function that obtains the scores of a set of evaluators and computes an aggregated score using a specific function, e.g. the average or weighted average. In a tree-like manner, the accumulator results are themselves aggregated using other accumulators until, finally, the overall score for a resource is computed. The information regarding which evaluator results is to be accumulated in which way is recorded in the customer profile. The most simple aggregator takes the score outputs of a list of evaluators, each ranking the resource in a certain way, and computes the (weighted) mean value. This then denotes the accumulated score.

The accumulators, connected via their respective in- and outputs, form a tree-like structure at which's top the overall score is computed.

4.6 Generic Evaluators

To support domain experts in quick testing and development of their own evaluators, the framework offers two kinds of generic constraints w.r.t. the raters. These generic evaluators may be scripted at runtime with either Java or Prolog program code, while giving the hard-coded evaluator access to all libraries and data sources. Thus, domain experts with some programming experience may test new evaluators first, using a generic evaluator.

4.7 Semantic Matching

Semantic matching is a technique which combines semantic annotations using controlled vocabularies with background knowledge about a particular application domain. Within the framework, the domain specific knowledge is represented by ontologies (cf. Sec. 4.2) containing formal definitions of tourism-related concepts as well as specifications of relationships between these concepts (e.g. taxonomic classifications of hotel amenities and POIs). Having all this background knowledge provided in a machine-understandable format makes it possible to compare hotel features with user preferences based on semantic similarity [18] instead of merely relying on the containment of keywords like most of the contemporary search engines do. The similarity between two atomic concepts c_1 *and* c_2 is determined by the distance $d_c(c_1, c_2)$ between them, which reflects their respective positions in the underlying concept hierarchy [21], [23]. Consequently, concept similarity is formally defined as:

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 3 / ISSUE 2 / AUGUST 2008 / 82-96
© 2008 Universidad de Talca - Chile
This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762008000100008

$$sim_c(c_1, c_2) = 1 - d_c(c_1, c_2)$$

Every concept in the taxonomy is assigned a milestone value [23] and calculated with the following formula:

$$\text{milestone}(n) = 1/ (2k^{l(n)})$$

where k is a factor greater than 1 and indicates the rate at which milestone values decrease along the hierarchy and can be assigned different values depending on the taxonomy depth; l(n) is the hierarchy level of the concept n in a given taxonomy. Since the distance between two concepts $d_c(c_1, c_2)$ represents he path from one concept to the other via the closest common parent (ccp), the instance is calculated as follows:

$$d_c(c_1, c_2) = d_c(c_1, ccp) + d_c(c_2, ccp)$$
$$d_c(c, ccp) = milestone(ccp) + milestone(c)$$

This approach implies two assumptions: First, the distance between siblings is greater than the distance between parent and child. Consequently, Finnish sauna and Turkish steam bath, both being sub-concepts of the concept sauna, are less similar than Finnish sauna compared with sauna. Secondly, the semantic difference between upper level sibling concepts is greater than between sibling concepts on lower hierarchy levels. In other words, two general concepts, like sauna and swimming pool are less similar than two specialized ones like Finnish sauna and Turkish steam bath.

The approach presented in this section, allows the framework to evaluate fuzzy queries like “I prefer hotels with Finnish sauna”. Since the project is not situated in the natural-language-processing area, such queries must be formulated using one of the query languages (RDQL, RuleML etc.) of the Semantic Web.

In this case, all hotels offering this amenity would of course get the highest ranking. Additionally, each hotel having for example Turkish steam bath would also be awarded higher rankings to hotels which do not offer such spa facilities. Moreover, semantic matching can also be applied to evaluate quantitative queries, for example “I prefer hotels with lots of Chinese restaurants in their vicinity”. The result of this query would also present at a slightly lower ranking, hotels that have, for instance, Thai restaurants nearby as result of semantic similarity between Chinese and Thai cuisine.
The evaluators performing semantic matching are domain independent and can be customized with respect to the underlying ontology (cf. [15] for more detailed description of the applied Semantic Matching Framework – SemMF (Site 11) and [11] where the semantic matching approach has been described in the context of the human resource domain on the basis of a particular example). This flexibility makes it possible to utilize them in a wide range of applications.

4.8 Search Results

While the engine is matching a profile to a set of available hotels, computing an overall score and evaluating the single evaluators, all information leading to a hotel ranking is stored together with the respective sub-scores. In doing so the customer may be offered further details in response to the essential question: Why is this hotel rated with this score? Furthermore, the framework allows re-weighting of the aggregated scores (“What sort of ranking can be expected if business features are not that important to me?”) or even re-arrangement of the customer profile excluding parts of the evaluation (“What will happen if I neglect the requirement of Japanese restaurants near the hotel?”).

5 Experiments and Results

5.1 Testing System

Since in the *Reisewissen* framework we can distinguish between several types of users, the development of the user interfaces for interaction with the system must also consider various needs, different complexities and previous knowledge of particular users. The end user or customer will be served by a low level web-based interface for hotel search. This interface may be enhanced by “expert options” which allow the customers to modify their individual profile, e.g. add special sub-profiles or adjust parameters for the existing profile. The design of these interfaces is beyond the scope of the project. A domain expert, e.g. a travel manager for a large company or someone with expert knowledge in the travel business is presented a more sophisticated user interface, which allows for the construction of complete (sub-) profiles and their testing. These profiles may then be made available to the customers. One of the goals of the *Reisewissen* project is the implementation of a development and testing environment for domain dependent evaluators and profiles. Using this environment, a domain expert is able to test the available data sources and evaluate meaningful rules, evaluators and parameterizations.

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 3 / ISSUE 2 / AUGUST 2008 / 82-96
© 2008 Universidad de Talca - Chile
This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762008000100008

A domain expert may use the workbench to create a complete customer profile, match this against a number of resources and then check the results for plausibility. Thus, pre-defined profiles like “traveling businesswoman” or “interested in spa” or those containing company guidelines may be developed and provided to potential customers, who can then choose one profile suitable w.r.t their preferences from the existing profiles, combine two or more or create their own profiles. A dialog window (cf. Figure 6) presents the evaluator which is actually worked on together with its possible parameters.

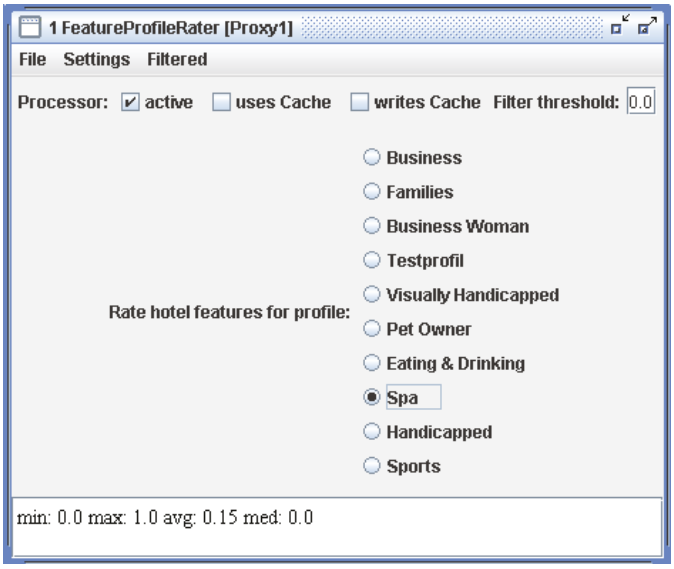


Figure 6: Processor Dialog Window

5.2 Test Cases and Results

Test cases were run using actual data from a set of about 500 hotels located in London. For the test runs, several customer profiles were created (business woman, pet owner, representative of a large IT company etc.), each aggregating a list of different hotel preferences such as a price range, locational preferences, hotel amenities and the like. The test profiles ran through the evaluation process, using different weights for the single preferences and comparing the results. The results showed that a complex semantic evaluation delivers better and more meaningful results than a simple ranking by price: basically the possibility of adjusting the results using different weight models proved valuable for customers. However, some problems with the ranking efficiency has been noticed when profiles with a lot of specialized preferences were used. These issues have been addressed using more (and more intelligent) caching methods.

6 Conclusion

In this article we presented the *Reisewissen* framework for the search and evaluation of hotels in large databases. The evaluation of the hotels is conducted on the basis of the information derived from several heterogeneous and distributed sources and under the application of the Semantic Web technologies which, in turn, are used to aggregate and utilize the given information. Customers are represented by profiles describing their particular requirements w.r.t. the preferred accommodation. The evaluation framework uses all available information to search and rank hotels in various contexts (location, features, price etc.). The rankings in the individual categories are aggregated and an overall match for each hotel reflecting the match between hotel and customer requirements is computed. The implemented framework makes use of semantic matching that allows for the integration of the semantic “distance” of concepts into the hotel evaluation.

Since the problem of the numerical evaluation is not necessarily restricted to the hotel domain the framework is divided into a domain dependent and a domain independent part, thus enabling an easy extension to other domains. Since one of the goals in the *Reisewissen* project is not just the design and implementation of a semantic hotel search and rating engine but mainly the evaluation of the usefulness and efficiency of semantic technologies in the tourism, the conclusion considers the questions whether these technologies have indeed proved their usefulness and what are the challenges waiting to be addressed in the future.

6.1 Lessons Learned

The main goals in the *Reisewissen* project are the integration of heterogeneous data sources and the (as far as possible) efficient and high-performance application of Semantic Web technologies in the tourism domain. Even

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762008000100008

though, open interoperability specifications are not new in e-tourism and especially the XML schemas of the Open Travel Alliance (OTA) are widespread in this field, the existing data exchange formats are not expressive enough to guarantee automatic exchange and processing of information to develop dynamic applications. Currently, tourism information systems provide good use cases for studying the potential brought by semantics and ontologies in order to solve the integration and interoperability problems they have been experiencing over the years. As stated in [4], Semantic Web technologies have a huge potential for e-tourism, e.g. w.r.t. semantically enriched information searching, integration and interpretability, personalized and context-aware recommendations and internationalization. More complex e-tourism services – beyond information services – benefit from Semantic Web Services technology, e.g. with respect to service discovery, composition and mediation and maintainability through abstraction. Nevertheless, while evaluating the usage of Semantic Web technologies in e-tourism, we have detected some, in our opinion, domain independent problems and challenges, and the requirements analysis of the hotel recommendation engine has generated constraints which must be taken into account. First of all, it must be ensured that the engine is flexible with regard to later adaptation to the production system and efficient regarding the end user querying process. Furthermore it must allow for an easy integration of information sources and provide means to generate new information from appropriately formalized expert knowledge. The added value of Semantic Web technologies lies in ontology-based applications like semantic matching and similarity searches as well as usage of already published RDF metadata. When it comes to querying at runtime and numerical evaluation, those technologies have proved to be less efficient [8]:

- **Querying Efficiency**: Originally, the idea in our project was to transform any available data into RDF triples, store it and use a purely Semantic Web-based approach to evaluate this data. The evaluation of the first prototypical implementation (pure-RDF) which operated on a single triple store was discovered to have some drawbacks, especially considering the performance of queries on the aggregated RDF data and the instability of the evaluated data stores (Jena, Sesame). Another problem was that the caching of already transformed data was rather complicated.

- **Numerical Evaluation**: Another issue with Semantic Web technologies is the problem that more complex numerical queries (like distance/radius queries in geo-coded information) require the use of non-Semantic Web languages like Java and Prolog. Technologies like Semantic Web Rule Language (SWRL) (Site 12) and OWL-E [17] proved to be unstable and inefficient. This caused additional effort when the data from RDF format had to be re-transformed into native language data structures. Especially at query time such complex evaluations should work directly on the original data.

6.2 Outlook

The *Reisewissen* Semantic Web-based approach has been prototypically tested within a hotel booking system. Its application may be extended into travel planning by incorporating more semantic relations such as conceptual descriptions of additional parts of a trip like booking flights, trains or buses. Again, the detailed preferences of a customer must be taken into account and knowledge about specific constraints (e.g. connection flights to the US might require a longer check-in time) can be automatically used to infer the “best” route. In a further perspective, a “travel companion” could infer necessary changes of the travel plan by considering online information such as flight delays and proposing a “plan B”.

Acknowledgments

This work is part of the *Reisewissen* project, a co-operation between *Freie Universität Berlin* and *ehotel AG*. The project was funded by the Investitionsbank Berlin (IBB) and the European Regional Development Fund (ERDF). We thank Jörg Garbers who has made immense contributions to the system design and its actual implementation.

Websites List

Site 1: World Wide Web Consortium (W3C)
http://www.w3.org/

Site 2: Knowledge Web EU Network of Excellence
http://knowledgeweb.semanticweb.org

Site 3: Knowledge Web Industry Board
http://knowledgeweb.semanticweb.org/o2i/

Site 4: Project SATINE
http://www.srdc.metu.edu.tr/webpage/projects/satine/index.html

Site 5: Project Reisewissen
http://reisewissen.ag-nbi.de/en

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762008000100008

Site 6: ehotel AG
http://www.ehotel.de

Site 7: Open Travel Alliance (OTA)
http://www.opentravel.org

Site 8: Open Travel Guides
http://openguides.org

Site 9: Resource Description Framework (RDF)
http://www.w3.org/RDF/

Site 10: Jena Framework
http://jena.sourceforge.net

Site 11: Semantic Matching Framework - SemMF
http://semmf.ag-nbi.de/doc/index.html

Site 12: Semantic Web Rule Language (SWRL)
http://www.w3.org/Submission/SWRL

References

[1] A. Ashri, W. Leewattankit, A. Min Tjoa, A Framework for Integrating Heterogeneous Tourism Information Sources, in Proceedings Information and Communication Technologies in Tourism Conference (ENTER2001), Montreal, Canada, 2001.
[2] T. Berners-Lee, J. Hendler, O. Lassila, The SemanticWeb; Scientific American, pp. 34-43, 2001.
[3] C. Bizer, A. Seaborne, D2RQ – Treating Non-RDF Databases as Virtual RDF Graphs, in Proceedings 3rd International Semantic Web Conference (ISWC2004), 2004.
[4] J. Cardoso, A. Sheth, Semantic Web Services, Processes and Applications: IDEA Group Inc.; 2006.
[5] A. Dogac, Y. Kabak, G. Laleci, S. Sinir, A. Yildiz, S. Kirbas, Y. Gurcan, Semantically enriched web services for the travel industry. SIGMOD Rec., vol. 33, no. 3, 21–27, 2004.
[6] M. Flügge, D. Tourtchaninova, Ontology-derived Activity Components for Composing Travel Web Services, in Proceedings International Workshop on Semantic WebTechnologies in Electronic Business (SWEB2004) co-located with Berliner XMLTage 2004, 2004.
[7] O. Fodor, H. Werthner, Harmonise: A Step Toward an Interoperable E-Tourism Marketplace, International Journal of Electronic Commerce, vol. 9, no. 2, 2005.
[8] J. Garbers, M. Niemann, M. Mochol, A Personalized Hotel Selection Engine. Poster at the 3rd European Semantic Web Conference (ESWC 2006), 2006
[9] T.R. Gruber, Toward principles for the design of ontologies used for knowledge sharing. International Journal Human-Computer Studies, vol. 43, no. 5-6, pp. 907-928, 1995.
[10] M. Haller, B. Pröll, W. Retschitzegger, A.M. Tjoa, R.R. Wagner, Integrating Heterogeneous Tourism Information in TIScover - The MIRO-Web Approach. Proceedings Information and Communication Technologies in Tourism, ENTER 2000, Barcelona, April 26-28, 2000
[11] R. Heese, M. Mochol, R. Oldakowski, Semantic Web Technologies in the Recruitment Domain, Competencies in Organizational E-Learning: Concepts and Tools (M.-A. Sicilia, Eds.), pp. 299-318, 2006.
[12] A. Maedche, S. Staab, Applying Semantic Web Technologies for Tourism Information Systems, in Proceedings 9th International Conference for Information and Communication Technologies in Tourism (ENTER2002), 2002.
[13] M. Missikoff, Harmonise: An Ontology-Based Approach for Semantic Interoperability [Online], 2002, Available: http://www.ercim.org/publication/Ercim_News/enw51/missikoff.html.
[14] L. Nixon, M. Mochol, Prototypical Business Use Cases, Deliverable 1.1.2 in the Knowledge Web EU Network of Excellence, 2005.
[15] R. Oldakowski, C. Bizer, SemMF – A Framework for Calculating Semantic Similarity of Objects presented as RDF Graphs, in Proceedings (Poster) 4th International Semantic Web Conference (ISWC2005), 2005.
[16] R. Olshavsky, Bridging the gap with requirements definition. Cooper humanizing technology [Online], 2002, Available: http://www.cooper.com/newsletters/2002_07/requirements_definition.htm.
[17] J.P. Pan, I. Horrocks, OWL-E: Extending OWL with Expressive Datatype Expressions [Online]. IMG Technical Report, IMG/2004/KR-SW-01/v1.0, 2004, Available: http://dl-web.man.ac.uk/Doc/IMGTR-OWL-E.pdf.
[18] J. Poole, J. A. Campbell, A Novel Algorithm for Matching Conceptual and Related Graphs; Conceptual Structures: Applications, Implementation and Theory, pp. 293-307, 1995.
[19] K. Prantner, K. Siorpaes, D. Bachlechner, OnTour: Semantic Web Search Assistant [Online], 2005, Available: http://e-tourism.deri.at/documents/OnTour%20Presentation.pdf.
[20] D. Quan, D. Karger, How to Make a Semantic Web Browser; in Proceedings 13th International World Wide Web Conference (WWW2004), 2004.
[21] J. F. Sowa, Conceptual structures: Information processing in mind and machine, Addison-Wesley, 1984.

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762008000100008

[22] The European e-Business Report: A portrait of e-business in 10 sectors of the EU economy, 4th Synthesis Report of the e-Business W@tch 2006/2007, European Commission - Enterprise & Industry Directorate General [Online], Available: http://www.ebusiness-watch.org/key_reports/documents/EBR06.pdf.
[23] J. Zhong, H. Zhu, J. Li, Y. Yu, Conceptual Graph Matching for Semantic Search, Lecture Notes in Computer Science, no. 2393, pp. 92, 2002.