How to “rightsize” an Ontology: a Case of Ontology-Based Web Information Management to Improve the Service for Handicapped Persons

Ralf Klischewski
German University in Cairo, Faculty of Management Technology
Al Tagamoa Al Khames, New Cairo City, Egypt
ralf.klischewski@guc.edu.eg

Abstract

The paper describes the first steps of improving the semantic integration of Web information services for handicapped persons in the area of Hamburg, Germany. To specify the informational resources in use, an ontology is suggested building on three main concepts: (1) a partonomy of physical objects of which the attributes represent most of the relevant information, (2) a simple taxonomy of informational objects, (3) a relation between the informational objects and those physical objects they inform about. It is argued that the presentation of a domain-specific ontology based on these concepts has convinced the stakeholders involved to increase their cooperation in order to meet the informational needs of handicapped people moving around the Hamburg area. Summarizing the lessons learned it is discussed how taking into account stakeholder perspectives contributes to “rightsizing” an ontology to support Web information management.

1. Introduction

As Semantic Web applications are only at the beginning, one of the critical questions is how to convince the stakeholders involved to invest in projects aiming at an ontology-based Web information management, especially in the area of e-government (cf. [2]). Conceptualisation of the domain in focus and relating the terms and expressions in a semantically correct way seems to be a costly endeavour and a challenge for the makers of the ontology as well as for those who are intended to understand and make use of it within their context of information management and IT support. Therefore, designing the “right” ontology is one of the keys to achieving stakeholder support and project success.

This paper reports on the first phase of a project to improve the semantic integration of Web information services for handicapped persons in the area of Hamburg, Germany. It is argued that the presentation of a simple, use-oriented and easily expandable ontology for structuring all the relevant informational resources has convinced the stakeholders involved to intensify their cooperation aiming at a Web information management based on Semantic Web technologies.

The paper is organized as follows: Section 2 describes the case and the imperative to improve the information integration. Section 3 reports on the domain modelling approach and the design decisions made. Section 4 evaluates the ontology from various stakeholder perspectives and accounts for why the ontology design was governed by “rightsizing”.

2. The case: improving the Web information service for handicapped persons

The domain in focus is information on the accessibility of buildings or sights and their facilities from the point of view of handicapped persons (mainly wheel chair drivers). A cooperative effort started in 2001 to provide information with the aim of reducing barriers of mobility. The main partners of this initiative (named “Mobility for Everyone”) are the Hamburg working group for handicapped persons (LAGH), the public transport network of the Hamburg area (HVV), and the provider of the city’s web-based information service (DiBIS, accessible through www.hamburg.de).

By the end of 2003 each of these three partners operates a website based on the offline exchange of the information among the partners. The information provided covers details about how to enter a building or sight, about bathroom facilities, parking and other aspects relevant for handicapped persons when moving around. The structure and granularity of the information display is quite heterogeneous even though the focus of interest is the same. Many times, annotated icons are the primary carrier along with additional text, but some Web pages are based on structured text only.

Figure 1. Annotated icons and text to inform about accessibility of a major sports facility.

For example, the LAG provides online documents covering a whole set of buildings (e.g. museums) without any possibility to address or jump to a certain museum of interest or to scan for certain features. The HVV provides information resources for each single building (station or sight), but employs different ways of structuring the information (see figures 1 and 2). DiBIS indicates wheel chair accessibility through displaying a single icon for each building or office within the building without further details.

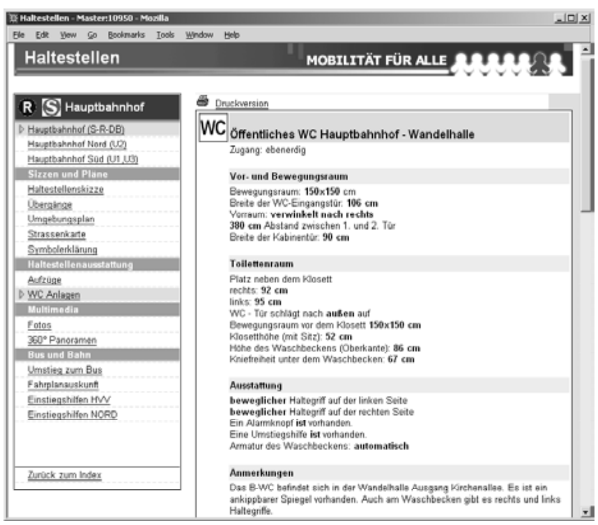

Figure 2. Structured text to inform about bathroom facilities located in the Hamburg main station.

At first, the initiative “Mobility for Everyone” had agreed to share documents and structured data in an offline mode and to establish links between the different information services. But it was shortly realised that further steps towards an integration of Web information are needed to address the following shortcomings:

- *Extensibility:* each of the partners runs different strategies to extend the information (e.g. to include information about a new building or a new facility inside), requiring new effort at each website.
- *Timeliness:* offline updates are shared only every few months or once a year; currently it is impossible to provide any up-to-date information pertaining to a facility, for example extracts from fault reports which technicians from facility service agents already prepare in electronic form.
- *Efficiency:* to some extent the same information is provided through different sites, not necessarily in the same quality (precision, timeliness), at present the information management has no means to avoid this inefficiency or to reach for coherency.
- *Resource identification:* at best, the current granularity allows to use URLs for identifying informational resources related to one building or sight. Except for full text search there is no way to automatically scrutinize and extract detailed information.

In October 2003, the author joined a project meeting to discuss the above problems and a possible cooperation with the University of Hamburg. It was agreed that during the winter semester a small group of students (participating in a project-based graduate course, supervised by the author) should analyse the current situation and elaborate suggestions for improvement on the basis of applying Semantic Web concepts and technologies. The results of this endeavour were presented at the next project meeting (January 28, 2004).

The following section reports mainly on the modelling effort to support the Web-based information management of the partners involved. The last section will recall the current shortcomings in order to evaluate to what extent the proposed ontology provides a solution.

3. Design of the ontology

The point of departure was the vision of the Semantic Web to enable machines finding their way around in a world of meaningful objects. The collaborative Web information management in this case was sought to be improved through semantic markup and semantic links which “allow machines to follow links and facilitate the integration of data from many different sources” [1]. It was agreed that the basis for this improvement must be an ontology through which all relevant informational resources can be identified, semantically structured and marked up with data making reference to this ontology.

Fortunately, the domain in focus (mobility information for handicapped people) seemed to be not too complex. For an experienced ontology designer it might even appear as a simple task. However, the task was not to produce a high-quality ontology as such, but to suggest a solution which fits the given context in which no experience or knowledge with Semantic Web concepts and technologies existed. Therefore, a number of questions were debated at the beginning of and throughout the ontology modelling: What is the (best) starting point for the ontology design? What is the most efficient way to approach? What about the existing conflicting concepts and visualizations? What is the appropriate degree of ontology formalization (cf. [5] for the range of choice)? How to make sure that the use of the ontology will overcome current shortcomings?

The following subsections trace the approach to the domain modelling and the design decisions which led to the result presented. For the following reasons the tool SemTalk (version 1.2.5; see www.semtalk.com) was chosen support the domain modelling and ontology design: free license for scientific use, easy to install (plug-in for Visio®), easy to use (training is required for the concepts, but not for the tool handling), interfaces to export models to RDF and OWL as well as to HTML (for external view). Within the context of this project phase the tool served well and its limitations (e.g. lack of multi-user environment, insufficient handling of model complexity, limitations to a subset of RDF and OWL) did not hamper the work at this time.

3.1. Getting started: analysing informational resources

The HVV runs an internal database to support its information service “Mobility for Everyone”. At the beginning of the ontology design, the structure of this database was analysed to identify and find out about the informational elements and their interrelation. At the same time, the text and the icons of the online presentations were scrutinized for the same purpose.

First ad hoc efforts of the student modellers to arrange those elements pointed in quite different directions. Obviously, choices had to be made which would preform the modelling path. It was decided to concentrate on bathroom facilities as a detailed focus to clarify the issues. The informational elements at this level covered, for example, the breadth of the inner bathroom door, the height of the washstand, the availability of handholds near the closet (e.g. left or right, fixed or movable), sometimes along with further comments (figures 1 and 2 display examples in German language). Even at this level the identification of informational elements was not trivial. For example, it remained an issue of debate whether the existence of a certain type of handhold or the attributes of any existing handhold should be regarded as the most elementary piece of information.

3.2. Separating schema and instances

Despite all the differences (see section 2) it was obvious that in all cases the information provided was centred around a certain building or sight. Therefore, the ontology must function as a schema (i.e. as machine readable instructions) for how to generate, annotate, process and transfer a set of data which represents information about the mobility conditions of these physical objects. In consequence, the SemTalk class diagram was used to model all classes and their attributes and relations which prescribe and explain all elements of the instance diagrams, i.e. the representations of the mobility information pertaining to a certain building such as a museum, the “Alsterdorfer Sporthalle” or the Hamburg main station (cf. figures 1-2).

3.3. Defining elements, relations and attributes

The choice of the elements and their attributes and relations was the most difficult part in the ontology modelling process. The main question was whether to primarily focus (a) on the informational resources as elements in their own right, or (b) on the physical objects they intend to inform about. There are valid arguments for each choice: Ad (a), the informational resources are the domain in focus, and concepts and structures are already existing; whereas there is no explicit model of any building or other physical object. Ad (b), most of the information provided is inherently connected to the physical objects and the structure of the informational resources provided implicitly presupposes a conceptualisation of buildings.

The case is somewhat comparable with the ontology-based annotation of photos (cf. [3]): On one hand a structure of the information about the photo (photo and medium features) is required; on the other hand there is a need to model the physical world to describe the subject matter which we can see on the photo. Similarly, it was decided to go for a hybrid ontology defining informational objects informing about physical objects which have a number of relevant properties. On the class level, the informational object has subclasses organized in a simple taxonomy. All subclasses of the physical object are organized in a partonomy with “building” as the root element (see figure 3 for part of the class diagram). Each element of the partonomy may have one level of subclasses (e.g. building’s entrance is of the type escalator, elevator, staircase, or ramp).

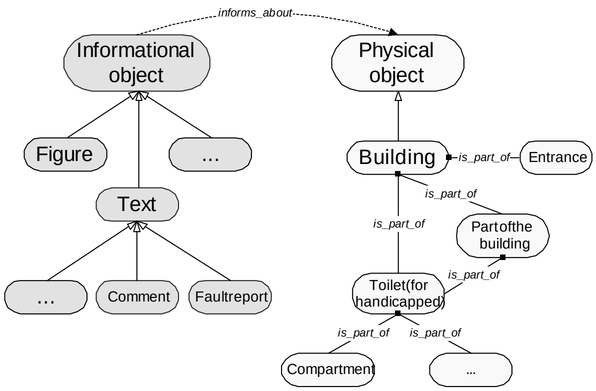

Figure 3. Class diagram (clipping) of the ontology

Each of the object subclasses is assigned a number of attributes which, as far as defined on the instance level, represent most of the relevant information for handicapped persons. For each attribute the type of values (integer, Boolean or text) is prescribed, sometimes even the range (e.g. the minimum and maximum of the heights for underriding). The modelling has been done in German only (see e.g. figure 4).

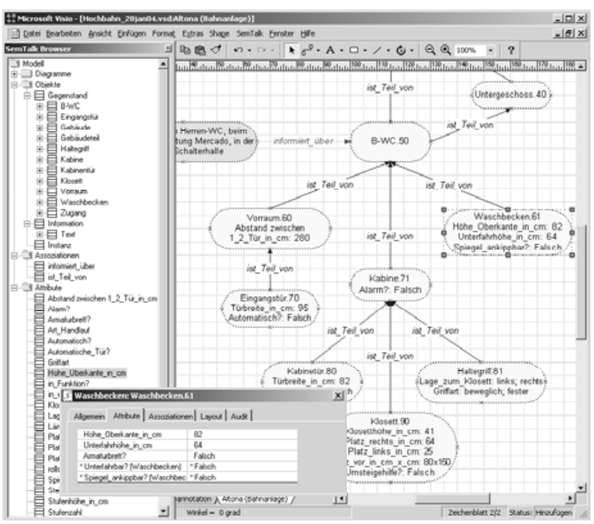

Figure 4. Instance diagram (clipping) for information related to the toilet facility at the Altona station

3.4. Expanding to formal logic

The main goal of this first phase of the project was to demonstrate the power of ontology use compared to the current state of information management. Therefore, the modelling of elements and their relations and attributes has been elaborated as far as all available information related to a building could be captured through an instance of the class model. The ontology was supposed to be adequate, simple, understandable as well as semantically correct and expandable. However, in this first phase of the project the design did not aim for formal rigor. The ontology serves as a schema for the instances (a set of data representing the mobility information related to a building) which must be modelled through the use the prescribed ontology elements, relations and attributes. However, not all of these prescriptions must be instantiated for every set of data, thus leaving the instance with predefined default values (e.g. see figure 4: the small window shows the defined and undefined (*) values of the washbasin's attributes). In consequence, the application of formal theory is still limited. For example, Smith and Rosse [4] argue that the *part_of* relation between the classes A and B is formally valid only if the relation is valid for all their instances, respectively. Up to now, this is not assured through the case ontology (e.g. a toilet may be part of the whole building or of one of its parts such as first floor or basement).

Another step up the ladder towards formal rigor is to include cardinalities (which is well supported through SemTalk) for all relations to formally restrict the structure of the data sets and to ensure the semantic correctness of the models (e.g. there is exactly one closet in each toilet compartment).

Envisioning future applications, the ontology-based information management should allow to generate automatic warnings such as "toilet at first floor may not be accessible due to elevator failure". The ontology should support this through derivation rules which enable a monitoring agent to analyse the likely impact on mobility when e.g. a fault report is associated to one part of an instantiated building. However, to achieve this the current ontology must be further developed (for which support through SemTalk is rather limited).

4. Lessons learned

The presentation of the above ontology was a success in the respect that the stakeholders involved (more than ten webmasters and content managers from HVV, DiBIS, and LAG) became enthused about the new options for their cooperative Web information management. To explain this positive reaction the ontology is evaluated from the perspectives which are mainly relevant for the stakeholders.

4.1. Perspective information management

Discussing the options for improving the cooperative Web information management, the stakeholders involved had realized significant problems and limitations with respect to extensibility, timeliness, efficiency and resource identification (see section 2). The use of the ontology could bring about a quantum leap in all of those areas: It supports the whole life cycle of the information as a schema for the data sets to be collected, transferred, processed, and presented. Thus, semantic interoperability is ensured between all partners when including information about new buildings and sights. Since the ontology allows a fine-grained structure and annotation, the informational resources can be identified on the elementary level, and the machine readability enables automatic addressing and retrieval. With the option for a virtual information space, current redundancies can be avoided and the exchange of information between websites can be organized real-time according to the profile of each information provider. Furthermore, new partners can be enrolled in the network (to provide e.g. up-to-date fault reports or information about new buildings) as long as they comply with the schema or a schema transformation is possible.

4.2. Perspective IT support

For practitioners, the enthusiasm about new visions is easily spoiled when learning about the work effort and/or financial investment required to make the dreams come true. In this case, the webmasters were pleased to understand that (a) the SemTalk provides RDF export of the class model (as well as an OWL export of possibly more advanced versions), and (b) the RDF-based instances already include markup for all relevant data which can be embedded also in regular HTML pages. This way there is only little effort required for sharing of the ontology between any partners involved (accessible even for machine agents) as well as for retrieving, processing, and presenting the data (most of it can be done on the basis of scripts, e.g. for export from an internal database to the Web server). All in all, the initial infrastructure investments seem rather small, and each partner can advance step by step according to the specific needs and (financial) abilities.

4.3. Perspective ontology design

To specify the content of informational resources, the ontology suggested builds on three main concepts: (1) a partonomy of a building to be composed of physical objects of which the attributes represent most of the relevant information, (2) a simple taxonomy of informational objects, (3) the relation "informs_about" for relating the informational objects to the physical objects at the instance level.

At the presentation of this ontology it was made clear that the university can only initiate the semantic modelling and that the domain experts are to be found in the stakeholder organizations themselves. Hence, all further design of the ontology must rely on the cooperation of these experts and their collective understanding of the concepts to be modelled and applied. The "experts" attending the presentation accepted the basic layout of the ontology as reasonable and agreed that detailed changes and/or extensions could easily be effected at both the partonomy and the taxonomy.

4.4. Outlook: service improvement?

The presentation of the above ontology as the key to semantic integration has convinced the stakeholders involved to intensify their cooperation in order to meet the informational needs of handicapped people moving around the Hamburg area. The ontology-based approach was accepted, at least in principle, because the domain modelling and the design of the ontology was led by the analysis of the stakeholders' problems, the analysis of existing informational resources and the foreseen capabilities to comprehend and realize the vision of the ontology usage. Thus, in retrospect, the ontology design was governed by the aim of "right-sizing" rather than accomplishing a set of given formal criteria. However, further developments of the ontology easily allow for expansion and formal upgrading.

In the end, do handicapped people in the Hamburg area now profit from an ontology-based Web information service? Well, not yet. The initiative "Mobility for Everyone" is a low priority project without any budget. Project meetings are scarce, and progress mainly depends on options opened up through external factors. However, the protagonists are now highly motivated and are likely to seize every chance when participating in high-priority Web information projects within their organizations.

5. References

[1] T. Berners-Lee and E. Miller, "The Semantic Web lifts off." *ERCIM News*, No. 51, October 2002, pp. 9-11

[2] R. Klischewski, "Semantic Web for e-Government." *Proceed. EGOV 2003*, Springer, Berlin, 2003, pp. 288-295

[3] A.Th. Schreiber, B. Dubbeldam, J. Wielemaker, and B. Wielinga, "Ontology-Based Photo Annotation", *IEEE Intelligent Systems*, May/June 2001; http://www.cs.vu.nl/~guus/papers/Schreiber01a.pdf

[4] B. Smith and C. Rosse, "The Role of Foundational Relations in the Alignment of Biomedical Ontologies", to be presented at *Medinfo*, San Francisco, Sept. 7-11, 2004

[5] B. Smith and C. Welty, "Ontology: Towards a new synthesis", in: Welty, C., Smith, B. (eds.): *Formal Ontology in Information Systems*. Ongunquit, Maine; ACM Press, 2001, pp. iii-x.