Towards Ontology-Based Yellow Page Services

Mikko Laukkanen
TeliaSonera Finland
P.O. Box 970 (Teollisuuskatu 13), FIN-00051 SONERA
mikko.laukkanen@teliasonera.com

Kim Viljanen, Mikko Apiola, Petri Lindgren, and Eero Hyvönen
Helsinki Institute for Information Technology (HIIT), University of Helsinki
P.O. Box 26 (Teollisuuskatu 23), 00014 University of Helsinki, Finland
firstname.lastname@cs.helsinki.fi

Abstract

This paper discusses the possibilities of the Semantic Web technologies in both annotating services and delivering relevant services to end-users. We propose an ontology-based mechanism for both advertising and finding the services. The essential parts of the system are ontologies for describing and storing service advertisements, a semantic service finder for the end-user, and a semantic service annotation editor for service providers.

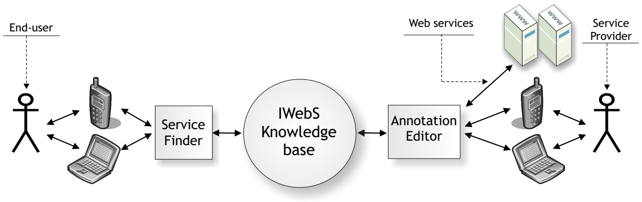

Figure 1. The general architecture of the IWebS system

1 Introduction

Yellow page directory services[1] on the Web are a widely used business concept for helping people to find companies providing services and selling products. Despite of the versatility of possibilities, it can still be difficult for the end-user to map a need to the services offered [1, 2, 3]. On the other hand, for the service provider, it may be difficult to index the service in such a way that the end-users would not miss the service. The problems with yellow page services arise in situations, where the end-user is not able to precisely state what kind of service would serve her needs.

The work presented in this paper represents the ongoing work of IWebS (Intelligent Web Services) project[2], which studies the possibilities of the Semantic Web [4] and Web Services [5] technologies in both annotating the services and delivering the relevant services to the end-users. We propose an ontology-based mechanism for both advertising and finding the services. The idea is to let the various actors in the IWebS system—in this case the end-users and the service providers—to use the terms and concepts that they are familiar with. These concepts are then mapped to the ontologies within the system. The general architecture of the IWebS system is depicted in Figure 1. The essential parts of the system are ontologies for describing and storing the service advertisements (the IWebS knowledge base), a semantic service finder for matching the services for the end-user, and a semantic service annotation editor for the service providers.

This paper is organized as follows. In Section 2 we give some background information to the problem area of mapping the end-user's need to a service. Section 3 describes a scenario for motivating the need for the IWebS system. In Section 4 we discuss how the end-users search and find services using the IWebS system. Section 5 describes the ontologies used within the IWebS system, and explains the means for annotating new services. In Section 6 we address issues that are not covered in the current version of the IWebS system. Finally, Section 7 concludes this paper.

[1] e.g., http://www.yell.co.uk
[2] http://www.cs.helsinki.fi/group/iwebs/

2 Background

Online yellow page services are a widely used service model for matching the need of a end-user with the corresponding products and services offered by companies. The business idea of yellow pages is based on helping end-users to find services as easily as possible, and to provide the advertising companies with a very targeted marketing media on end-users that are trying to find companies for a specific need.

Typical online yellow page service provides the user with the keyword-based search and hierarchical or flat-list navigation [3]. In the keyword-based search the end-user locates services by just typing in a few keywords to a search engine. The end-user does not have to figure out which categories in the yellow pages may be relevant from the viewpoint of her need. However, the end-user needs to know the relevant keywords. Also, the matched document does not necessarely prove to be relevant, for instance, if the keyword was *stamp* and the retrieved document contains the phrase "We do not sell *stamps*, but...". Also, a textual description is found only if it contains the explicit keyword. For example, one may be interested in companies dealing with *astronomy*. A telescope advertisement is not found unless it happens to mention the word *astronomy*, which may be too obvious to be mentioned.

In the case of hierarchical or flat-list navigation a typical yellow page service provider maintains a list or a hierarchy of product and service categories, such as "Electronic equipment" or "Car Rental". All advertisements are then placed under one or several categories to help the user to find the services. Based on the categorization, the user can navigate to the category that best fits the user's intentions. However, from the viewpoint of the user's need, indexing a company's advertisements according to a business category, such as the aforementioned product and service, is not very useful unless the category unambiguously implies the services offered to the user. For example, a camera repair service can potentially be offered by an importer company, appliances shop, camera shop, photo shop, or an optician. What kind of companies offer repair services depends for instance on the service business at hand, on the thing and brand being repaired, and on local practices. In order to enhance the search capabilities of yellow pages the advertising companies should more clearly state what services they actually offer.

The IWebS project has been launched to investigate the possibilities the Semantic Web and the Web Service technologies offer for creating more effective methods for finding real-world services. The goal of IWebS is to create an intelligent yellow pages service, where the semantically annotated services cover both static and dynamic advertisements, whose availability to the end-user depends on the context in which the user and the services are.

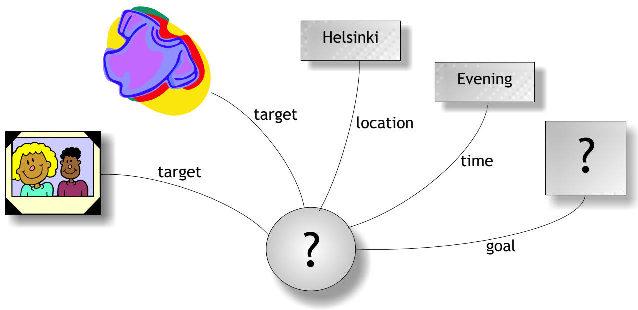

Figure 2. A query to the IWebS system.

3 Usage Scenario

In the scenario an end-user named Cathy is attending a conference, where she got the chance to get in the same picture with a famous invited speaker. At the end of the conference day Cathy wants to get the digital picture printed on a t-shirt, which she will keep as a souvenir from the conference.

Cathy has the evening off, so she would like to get the picture printed right a way. However, Cathy does not have any idea of which service provider (e.g., a shop or a print house) could do the job for her. Therefore, she activates her smartphone, and presents the problem to the IWebS system. Cathy speficies that the related objects for the query are the picture and a t-shirt. The service needs to be available within the Helsinki downtown area, and it has to be open at the evening. Cathy is not familiar of what the procedure (i.e., the goal) of getting the picture onto a t-shirt is called, thus, she leaves that for the IWebS system to find out. The information Cathy specifies in her query is visualized in Figure 2.

The IWebS system processed the query, matches the service providers, and returns a list of relevant service providers to Cathy together with information about the services themselves, their opening times, and directions on how to get there. In this case the list contains the following service providers:

- A nearby shop selling t-shirts and a print house. In this case Cathy needs to first buy the t-shirt, and then let the print house print the picture on it.

- A photo shop and a specialized stall on a market. The market stall happens to be available in Helsinki on that week, and it sells t-shirts with any picture printed on them. But before this, Cathy needs to go to the photo shop to have the picture developed.

- An art shop, a nearby shop selling t-shirts, and a photo shop. The art shop sells the required equipment for printing the picture on a t-shirt by oneself. To do that, Cathy needs to buy the t-shirt and develop the picture.

From these, Cathy chooses the market stall, because after the printing, she has some time to wander at the market.

The scenario so far has assumed that there indeed exists services annotated in the IWebS system. We will now extend the scenario to show how the dynamically available services, such as the moving market stall, are added and updated to the system.

A moving market stall holder Mick is planning to visit Helsinki this week to do some business. He has added an advertisement to the IWebS system earlier, and likes to update his service offering definitions. He activates his Palm device, and changes all the services he offers (selling t-shirts, souvenirs, and hot dogs) to be available in Helsinki, and provides the exact location and opening time information. He also adds a new service—printing pictures on t-shirts—by annotating it to be related to terms ”print”, ”picture” and ”t-shirt”. In addition, the IWebS system suggests that the new service should relate to ”Personal appearance” and ”Refresh / entertainment” based on other similar annotations. Mick agrees, and commits the changes to the IWebS system.

4 Searching for the Services

In a general user driven information retrieval system the user's input can be collected implicitly (user's context and profile), explicitly by keywords typed by the user, or explicitly by navigation-based input. In our case, based on the input from the user, the system must be able to present the user's problem (e.g., the Cathy's problem in Figure 2) in such a format that the problem can be solved by the available services.

OntoSeek [1] provides the user with a natural language interface where the user can describe her problem using arbitary natural language terms and describe relations between them as lexical conceptual graphs which resembles the figure 2. OntoSeek uses ontologies such as the WordNet[6] for expanding the queries with, e.g., synonymes, which helps to match the queries with the natural language advertisements. The YPA system provides the user with a natural language search to yellow page advertisements [2]. The system uses natural language processing and information retrieval technologies for searching the semi-structured advertisements. With YPA the user can make questions like "I need to get my camera repaired!" which are answered based on the advertisements and the world model (the WordNet[6]).

The strength of both the OntoSeek and YPA system is that any collection of natural language advertisements can be queried by the systems. This is also a weakness, since the natural language understanding can be difficult and error-prone. The OntoSeek uses lexical conceptual graphs to present the queries, but since the vocabulary and the relations are unconstrained, the graphs can not be validated and hence the queries can be unsound [1].

In our work, we propose using restricted terms and relations, described by ontologies, for making the queries and for describing the available services. If the right terms and relations are found, this helps both the end-user and the advertisers in describing their needs and offerings. In addition, using ontologies the user interfaces can be built in such a way that they help and direct the user's action towards semantically sound results, such as in the MuseumFinland [7] system.

As a first step we have tested the Museum Finland framework [7] in the IWebS context. The Museum Finland's user interface is based on the idea of a view-based search [8] where the user can make multiple selections from different views on the underlying content, presented in RDF(S) The views can be presented as tree-structured categorizations. The user can make queries to the underlying content by making selections (restrictions) using one or several views. The result of the query is those resources that matches all the restrictions. In Museum Finland the view-based search has been extended by keyword-based search, which provides an additional way to define restrictions to the query.

The outcome of the test was that the Museum Finland-based IWebS system made it possible to do view-based search on the advertises imported to the underlying knowledge base. In this first step we used only two views (service categorization and location), of which the service categorization is too difficult for end-users to use. In the following, we will describe more user-friendly ways of representing the query.

5 Service Ontologies and Annotation

Service metadata is needed for ontology-based search to function efficiently. We argue that creating semantically correct metadata is a fundamental problem of the Semantic Web. The creation of metadata can be done either automatically or manually. Automatic annotation usually means processing large amounts of existing data using natural language processing and data mining techniques. Depending on the data, the automatic annotation can be too complicated task for computers, and some help from a human user is required. Within the IWebS system, the most relevant problems associate to manual annotation; how to get the best possible annotation with a minimal effort from the user, how to automatize the process to its full potential, and how to validate the annotation.

5.1 Describing the Services using Ontologies

Traditional yellow page services are classified from one point of view as described in Section 2. We are interested in describing the services as processes, which have goals, targets (e.g., t-shirt in Cathy's problem) and take place in time and location. The services in the IWebS system are described using a set of ontologies, which are goal, target, service provider, service offering, Standard Industrial Classification (TOL) [9], Classification of Individual Consumption by Purpose (COICOP) [10], time, and location. The goal and target ontologies are targeted for specifying the end-users' needs. The service provider, service offering, and the classification ontologies are used for describing the service offerings.

The goal ontology consists of abstract concepts, which express activities such as *Alter*, *Copy*, *Create*, *Erase*, and *Move*. The terms in the goal ontology imply the abstract meaning of several domain specific terms, and aim at giving the user with means to query the services by a common sense. Thus, the user does not have to know any domain specific terms, when querying services from the goal viewpoint. Very similar terms can be found from the existing Process ontology in Standard Upper Merged Ontology (SUMO) [11], which could perhaps be translated to Finnish and used as a goal ontology in the IWebS system.

The product and the "fields of life" ontologies are used for describing the targets of the service offerings. The product viewpoint is defined by the COICOP. The top level of the fields of life ontology consists of *Home*, *Work*, *Health*, *Education*, *Refresh/Entertainment*, *Personal Appearance*, *Capital*, *Food and Supplies* and *Social Interactivities*.

An instance of a class in the service provider ontology can be anything that is able to provide one or more service. The services in turn are modeled as service offerings. For instance, a barber shop is a service provider, which has two service offerings: making haircuts and selling hair lacquers.

All ontologies are presented in OWL-format. The time ontology was created by our own, and it is used to present for instance the opening times. The location ontology was imported from the Museum Finland project [7].

The ontologies are bound together using properties. Figure 3 depicts these bindings. The service provider has one or more service offerings, and the service provider is located at some location. The service offering has also a location, which can be different from the one of its service provider. Furthermore, the service offering is classified using the TOL. Finally, the service offering may have a goal, and it is targeted at either an instance in COICOP or in the "field of life" ontology.

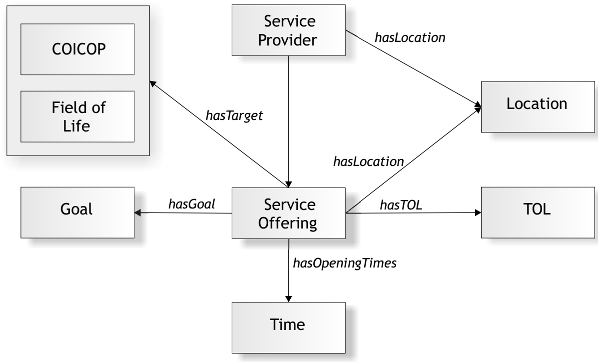

Figure 3. The ontologies and their relationships within the IWebS system

5.2 Using an Annotation Editor for Creating Service Annotations

There exists a range of annotation editors such as the Annotea [12], the SHOE Knowledge Annotator [13], the AeroDAML [14], the MnM [15], and the OntoMat [16].

In Annotea [12], the annotation means attaching web pages with users "comments" such as advices, change suggestions or opinions about the page. Although the annotation RDF-schema in Annotea can be extended by users, it does not support the use of multiple ontologies, as needed in, e.g., the IWebS system. Annotea is based on a document-centric approach, where the users are browsing documents and examining annotations related to them. The annotations are not intended for helping find the data.

SHOE Annotator [13] aims at annotating Web pages by linking them to ontologies using its own SHOE language. The annotations are then collected to a server and used for finding the pages easier. SHOE does not support RDF. AeroDAML is a Web service, which automatically annotates a given Web page using a given ontology with the help of WordNet [14].

The MnM [15] and the OntoMat [16] are aimed at solving problems of automatic annotation. They include features such as extraction of text phrases from documents—automatically and semi-automatically—using techniques such as natural language processing.

All of the editors mentioned above are hard to use for persons with minimal skills in computer usage, and who are not familiar with ontological concepts nor the problems of annotation. In our case, the editor should be easy to use for users interested in describing their services but not interested in technical details about annotating.

Our goal is to develop a user friendly annotation editor, which guides the annotator to make correct annotations based on the ontologies. One possibility is to provide the

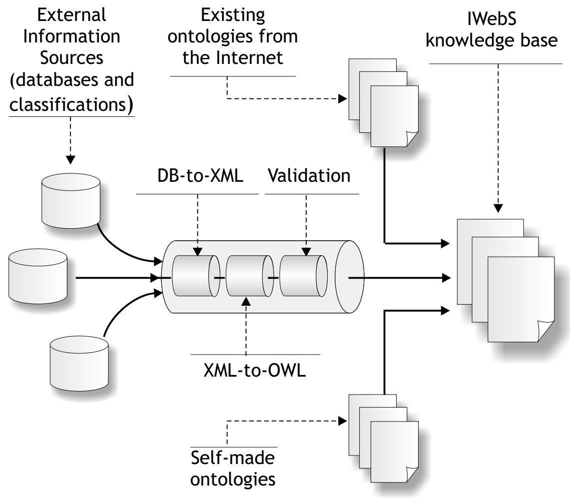

Figure 4. Importing instance data into IWebS

annotator with a multi-view-based user interface, which restricts the choices in the annotation during the annotation process. By a multi-view-based user interface we mean an interface similar to the search interface described in Section 4. For example, the annotator can start the annotation by classifying the service to some inland location in the location ontology. Then, the other ontologies will be restricted so that the system guides the annotator to a reasonable annotation. In this situation, a service classification ontology would be restricted so that it would not be possible to annotate the service to "waterborne-traffic"-class, since the location is (based on the ontological knowledge) far away from water.

Based on the ontologies, annotation recommendations could be created suggesting services that the annotator would offer. Recommendations could be created based on ontological rules and existing annotations. For example, a user annotating her service as a barber shop could be asked, if her shop also sells hair lacquers or other stuff.

After the initial service annotation, the updates to the annotation can be done either by a human end-user (i.e., the service provider), or by a legacy system of the service provider. For instance, for a small flower shop owner it is easier to use a Web-based tool to edit the annotation for the shop. However, medium or large companies, such as restaurants or gas station chains, could integrate their legacy systems to automatically keep the annotation up-to-date. This can be done by using the Web service interface to the annotation editor (see Figure 1).

5.3 Importing Instance Data

For importing the instance data from existing databases or other information sources we have built a publication pipe (See Figure 4), with which we have translated service provider and service offering data stored in a legacy database into corresponding OWL instances. The service provider annotations cover over 200 000 advertisements representing service providers all across Finland.

The publication pipe operates in three subsequent phases. In the first phase the data is encoded into XML. The second phase translates the XML-encoded data into OWL language. This phase is the most important, and requires data-specific translators. The output of this phase is the OWL ontology (classes and properties) and the actual instances representing the original data. In the third phase the generated OWL ontology is validated. If the final validation phase passes, the data is correctly transformed into OWL, and is usable by the IWebS system.

6 Future Work

The current version of the IWebS system provides both keyword and navigation-based user interface for querying services. In the future we are improving the query interface so that the end-user does not have to know explicitly what she is looking for. We are aiming at a solution, where the end-user only needs to express her problem to the IWebS system, which in turn infers what kind of services could solve the problem.

We are also interested in dynamic content, whose availability to the end-user depends on the contexts where both the end-user and the service provider are. The service providers are given the possibility to update their service profile on the fly. The update should be done either by hand using the annotation editor, or automatically using Web services, which integrate the service providers' legacy systems into the IWebS system (see Figure 1). In doing so, for instance a barber shop could advertise a happy hour with discounted prices in a ad hoc manner.

The IWebS system is intentend to be available both for stationary (desktop) and mobile users. Currently only the former case is supported. The mobile devices range from low-end mobile phones to high-end personal digital assistants (PDA) and smartphones. Thus, the user interfaces for the IWebS system needs to range from mobile phones to a full-blown Web (or XForms [17]) browser.

Finally, since the quality of the data in yellow page services is higher than the data in the Web and, on the other hand, the companies public Web pages contain typically more information than the advertisements in the yellow pages, the yellow page data could perhaps be used as a bootstrap data for a domain specific internet search engine that would index advertisers Web pages. This would combine the benefits of the closed, high quality service advertisement registry with the greater variety of information published on the Web by the advertisers.

7 Conclusion

In this paper we introduced the work being done in the IWebS project, which studies the possibilities of the Semantic Web and Web Services technologies in both annotating the services and delivering the relevant services to the end-users. The IWebS system differs from other online yellow page services in that it utilizes ontologies in both queries and service annotations. The baseline idea is to let the end-user and the service provider to use the terms and concepts that they are familiar with. These concepts are mapped to the ontologies within the system. The essential parts of the system are ontologies for describing and storing the service advertisements, semantic service finder for matching the services for the end-user, and semantic service annotation editor for the service providers.

The current prototype of the IWebS system is based on the Museum Finland framework [7]. The Museum Finland's user interface is based on the idea of view-based search [8] where the user can make multiple selections from different views on the underlying content. The view-based search has been extended by keyword-based search, which provides an additional way to define restrictions to the query.

Acknowledgements

This work was funded by the National Technology Agency Tekes, Fonecta, TeliaSonera, and TietoEnator.

References

[1] N. Guarino, C. Masolo, and G. Vetere, "OntoSeek: Content-Based Access to the Web," *IEEE Intelligent Systems*, pp. 70–80, May/June 1999.

[2] A. De Roeck, U. Kruschwitz, P. Neal, P. Scott, S. Steel, R. Turner, and N. Webb, "YPA - an intelligent directory enquiry assistant," *BT Technology Journal*, vol. 16, no. 3, pp. 145–155, 1998. [Online]. Available: citeseer.ist.psu.edu/roeck98ypa.html

[3] E. Hyvönen, K. Viljanen, and A. Hätinen, "Yellow Pages on the Semantic Web," in *Towards the Semantic Web and Web Services, the Proceedings of XML Finland 2002 Conference*, 2002, pp. 3–14.

[4] T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," *Scientific American*, vol. 284, no. 5, pp. 34–43, May 2001.

[5] D. Booth, H. Haas, F. McCabe, M. Champion, C. Ferris, E. Newcomer, and D. Orchard, "Web Services Architecture," Aug. 2003, W3C Working Draft 8, available at: http://www.w3.org/TR/2003/WD-ws-arch-20030808/.

[6] C. Fellbaum, Ed., *WordNet: An Electronic Lexical Database*. The MIT Press, May 1998, iSBN 0-262-06197-X.

[7] E. Hyvönen, M. Junnila, S. Kettula, E. Mäkelä, S. Saarela, M. Salminen, A. Syreeni, A. Valo, and K. Viljanen, "Finnish Museums on the Semantic Web. User's Perspective on MuseumFinland," in *Museums and the Web 2004 (MW2004)*, Arlington, Virginia, USA, Mar. 2004.

[8] A. S. Pollitt, "The Key Role of Classification and Indexing in View-Based Searching," University of Huddersfield, UK, Tech. Rep., 1998, available at: http://www.ifla.org/IV/ifla63/63polst.pdf.

[9] Statistics Finland, *Standard Industrial Classification TOL 2002*. Helsinki: Valopaino, 2002, ISBN 952-467-097-6, Available at: http://www.stat.fi/tk/tt/luokitukset/index_talous_keh_en.html.

[10] United Nations, Statistics Division, *Classification of Individual Consumption by Purpose (COICOP)*, New York, USA, 1999, Available at: http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=5&Lg=1.

[11] I. Niles and A. Pease, "Towards a Standard Upper Ontology," in *The Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001)*, 2001.

[12] J. Kahan, M. Koivunen, E. Prud'Hommeaux, and R.Swick, "Annotea: Open RDF Infrastructure for Shared Web Annotations," in *The Proceedings of the WWW10 International Conference*, 2001.

[13] J. Heflin and J. Hendler, "A Portrait of the Semantic Web in Action," *IEEE Intelligent Systems*, vol. 16, no. 2, 2001.

[14] P. Kogut and W. Holmes, "AeroDAML: Applying Information Extraction to Generate DAML Annotations from Web Pages," in *The First International Conference on Knowledge Capture (K-CAP 2001). Workshop on Knowledge Markup and Semantic Annotation*, Victoria, B.C., Canada, Oct. 2001.

[15] M. Vargas-Vera, E. Motta, J. Domingue, M. Lanzoni, A. Stutt, and F. Ciravegna, “MnM: Ontology Driven Semi-Automatic and Automatic Support for Semantic Markup,” in *The Proceedings of the 13th International Conference on Knowledge Engineering and Management (EKAW 2002)* , A. Gomez-Perez, Ed. Springer Verlag, 2002.

[16] S. Handschuh, S. Staab, and A. Maedche, “CREAM—Creating Relational Metadata with a Component-Based, Ontology-Driven Annotation Framework,” in *The First International Conference on Knowledge Capture (K-CAP 2001)*, Victoria, B.C., Canada, 2001.

[17] M. Dubinko, J. Leigh L. Klotz, R. Merrick, and T. V. Raman, “XForms 1.0,” Oct. 2003, W3C Recommendation, available at: http://www.w3.org/TR/2003/REC-xforms-20031014/.