Semantic E-government Portals – A Case Study

Teemu Sidoroff and Eero Hyvönen
teemu.sidoroff@tkk.fi, eero.hyvonen@tkk.fi

Helsinki Institute for Information Technology (HIIT)
Helsinki University of Technology, Media Technology

Abstract. This paper presents a case study on how semantic search and browsing techniques can be applied to solving the problems of content discovery, aggregation, and linking in e-government portals. At the same time, adaptability of a semantic portal tool, OntoViews, based on the multi-facet search paradigm, to different kinds of content was tested. Our conclusion is that representing content and their linking in terms of semantic web ontologies and logic rules is flexible from the system construction viewpoint, can be used to provide the end-user with useful "semantic" services, and can reduce human effort in portal maintenance.

1 Introduction

Governmental information and services on the web are typically provided by different organizations as independent web pages, databases, services etc. at different web locations. This creates severe obstacles for citizen end-users. First, *discovery* of relevant content is difficult because it requires prior knowledge of the administrative organization providing the contents. Second, the information and service needs often require *aggregation* of content from several information providers, which is difficult if heterogeneous content is provided by several independent web sites. For example, when a new baby is born, relevant information for the family is provided by health care organizations, social organizations, the church, legal administration, and others.

Portals try to ease these problems by collecting content from various organizations into a single site organized according to the clients expected information needs. Portals can be broadly classified into three major groups by their functionality: First, *service portals* collect a large set of services together into a localized miniature version of the web and are meant for wide audiences. Such portals include Yahoo! and other "start pages" provided by various Internet service providers for their customers. Second, *community portals* [1] act as the virtual socializing or meeting place of a community. The community can be evolving around the portal itself, or the portal can act as the extension of the community members' activities. Community portals frequently contain bulletin boards and other means of communication. Third, *information portals* [2] act as hubs of data on the web either by containing a large amount of information about a domain

within the portal or contain a structured collection of annotated hyperlinks to other resources.

This paper considers the problems of content discovery and aggregation in e-government information portals. As a solution approach, application of semantic web techniques[1] is studied. According to the semantic web vision, the web is changing into a web of metadata annotated using machine "understandable" vocabularies, ontologies [3, 4], and much of the semantic web content will be published using semantic portals[2] [5] based on web standards, such as RDF and OWL [6]. Semantic portals typically provide the end-user with two basic services: 1) a search engine based on the semantics of the content [7] and 2) dynamic linking between pages based on the semantic relations in the underlying knowledge base [8].

We present a case study [9] of the semantic web approach: a semantic prototype version[3] of (a part of) the Suomi.fi portal[4] containing information from various Finnish government organizations as a collection of annotated links. Suomi.fi is the largest and most used governmental citizen portal in Finland. Our case study suggests that the content discovery problem can be approached effectively by more accurate and flexible semantic-based search engines, and the aggregation problem by a dynamic semantic linking system that automatically consolidates heterogeneous content related to the same information need or problem situation. For example, heterogeneous information related to the baby-is-born situation can be accumulated based on metadata, ontologies, and logic rules. From the business viewpoint, automatic linking aggregation lessens need for human editorial work in portal maintenance. Our work also shows, that by using domain independent logic and semantics, the same portal engine can be adapted fairly easily to different kind of domain dependent content and annotation schemas.

2 The Original Portal

Suomi.fi portal is aimed to ease the discovery of public services. It offers news and information about the public administration that is relevant to the daily life of citizens. The system is maintained by an editorial team at the Ministry of Finance in Finland. The main bulk of the original Suomi.fi portal content is a large link collection that is grouped into a taxonomy of 12 main subject areas of interest, such as "Culture and hobbies", "Education and libraries", "Employment and entrepreneurship", etc. The portal is organized into a Yahoo!-like category taxonomy of topics and web pages. Figure 1 depicts the home page of the portal.

The taxonomy consists of two levels of links leading to web sites and information pages maintained by various public services, ministries, and other

[1] http://www.w3.org/2001/SW/
[2] See, e.g., http://www.ontoweb.org/.
[3] http://museosuomi.fi/suomifi/
[4] http://www.suomi.fi

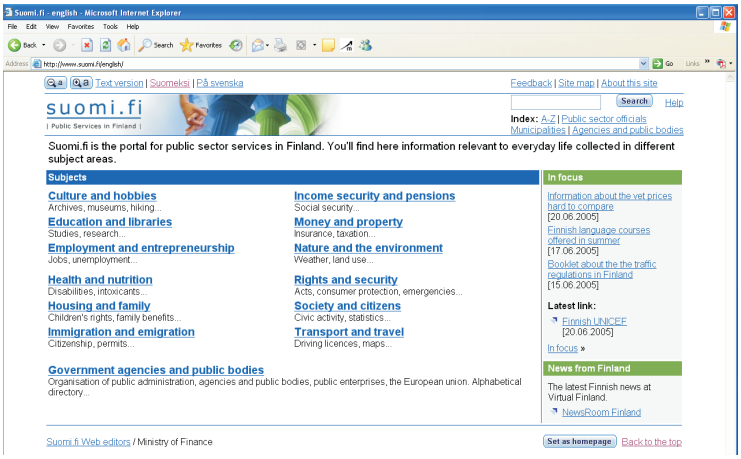

Fig. 1. Suomi.fi portal with first level categories of the topic taxonomy shown.

governmental agencies. The links are annotated with a brief textual description of their content and with a set of keywords to help finding links with keyword search. By selecting a category, information, and links related to it are shown as well as a set of subcategory links to browse the directory further. In addition to browsing, a keyword based search engine focusing on portal contents and categories is provided. Suomi.fi is an example of the single facet directory architecture widely used on the web. We will call this architecture the *Yahoo! approach*. This approach has been applied also on the semantic web, e.g., in the Open Directory Project Dmoz[5].

We decided to try to enhance the portal in two ways. First, in order to enhance content discovery, the single-facet paradigm of the Yahoo! approach was generalized into multi-facet search based on ontologies, as suggested in [10]. Second, in order to solve the content aggregation problem, a semantic link recommendation system was devised grouping relevant subject topics together with natural language explanations. These ideas originated from our earlier work on developing the semantic portal MuseumFinland[6] [11] for publishing cultural collections on the Semantic Web.

In this work, a generic semantic portal tool OntoViews[7] [12] was created, and a goal in our case study was to test how well this tool could be adapted to a quite different domain. OntoViews consists of three components: 1) Ontogator is a multi-facet search engine based on ontologies. 2) Ontodella is a logic server that stores the RDF data that is to be presented in the portal and provides, e.g., dynamic semantic linking services. 3) OntoViews-C: a system on top of

[5] http://dmoz.org
[6] Available with an English tutorial at: http://museosuomi.fi
[7] The software is available at http://www.cs.helsinki.fi/group/seco/museums/dist/

the Apache Cocoon pipeline framework[8] for transforming the RDF data into browsable web pages.

In the following we present why and how the semantic search, link aggregation, and interface components of the semantic Suomi.fi portal SW-Suomi.fi were developed.

3 Suomi.fi on the Semantic Web

The problem in using a single classification taxonomy is that content can be found through only one mental view. This hinders content discovery when the user has a different mindset or view of the world than the creator of the taxonomy [13], or wants to view contents simultaneously through multiple views. For example, the user may be interested in a certain category of services in a given location at which the service is provided.

In multi-facet search [14, 15], the content can be classified and viewed along several orthogonal views simultaneously. Each category is related to a set of search objects (in our case annotated links to web pages) that we will call its *projection*. The *extension* E of a category is the union of its projection P and the extensions of its subcategories S_i: $E = P \cup S_1 \cup S_2 \cup ... \cup S_n$. A search query in multi-facet view-based search is formulated by selecting categories of interest from the different facets, typically one selection from a facet. The answer to the query is simply the intersection of the extensions E_i of the selected categories: $A = \cap\{E_i\}$.

OntoViews supports the multi-facet search paradigm. The information items of the semantic Suomi.fi portal were described by using a set of ontologies, which were projected into taxonomies for the user interface. The projected taxonomies describe the information content of the portal from different view points. Ontologies are used as the basis of the navigational structure, and OntoViews provides a mechanism for automatically projecting the views from the underlying knowledge base that consists of the ontologies and metadata annotated using them.

The ontologies, defined in OWL, describe the following views for the portal's information items: 1) **Topics** is based on the link taxonomy of the original portal. The topics range from *museums* to *taxation*. 2) **Location**, the sphere of interest for the page. The location ontology consists of countries, municipalities, cities, and information about their partonomy relations. 3) **Target audience**, the ontology dissects the users of the portal into different groups, such as students, employers and organizations. 4) **Life events**, the ontology covers things and concepts that are not strictly bound to any single target audience, such as *marriage*, *moving*, *getting ill*. Some of the things and concepts have relations to specific target audiences as well, e.g. the life event *starting studies* is related to the target audience *students*. 5) **Organizational structure**, the ontology describes the organizations and their relations. The organizations represented in the ontology are responsible for the links that are present in the portal. 6) **Page**

[8] http://cocoon.apache.org/

content, the ontology contains concepts that describe different types of page content. These include concepts such as *authority contact*, *study material* and *discussion forums*.

The *Topics* view contains the whole subject topic taxonomy of Suomi.fi (cf. figure 1). The other views provide new ways for classifying the content as described above. A major idea in our approach is that new kinds of views for view-based search can be created easily, and the content be projected on them, based on end-user needs. For example, the new view *Life events* presents the content classified along the categories depicted in the web page of figure 2. The numbers in parentheses tells how many hits will be found, if a category link is selected in multi-facet search. In this way the user never ends up in empty result set.

Fig. 2. The Life events facet view of SW-Suomi.fi, projected from an OWL ontology.

The portal contains two types of information items. First, the *simple information page* presents a hyperlink to a web page that is displayed with relevant information about it, such as the organization responsible of that page and a brief description. Second, the *compound pages* tie together several simple information pages into a list that is displayed as a single resource. The portal also contains simple information pages that do not contain any hyperlinks to other web resources, but simply contain a piece of information that can be used as building blocks in compound pages. Compound pages make it possible to tightly couple together resources that form a clearly defined topic collection or narrative. For example, a compound page may collect together resources that are of interest w.r.t. *motherhood*. By collecting compound pages it is possible to aggregate together information from different sources and present it in a clear way.

In OntoViews, the Ontodella also acts as the dynamic linking engine of the portal. It can be used for accumulating content and providing links to semantically related web pages. The links are generated dynamically by the engine based on explicit logic rules in SWI-Prolog[9]. Using the rules it is possible to link together any of the information items of the portal that satisfy some arbitrarily complex rule. Usually this means that there exists some resource in the ontology to which the linked items are connected through similar properties. The recommendation rule system can also be used to provide navigational links that provide structural form to the portal.

The user interface of SW-Suomi.fi, created by OntoViews-C pipelines, is in many ways similar to the Flamenco system [13]. During searching and browsing, the left hand side contains the browsable facets, while the center of the screen is reserved for the information items that match the search criteria.

The faceted user interface makes it possible to narrow down the search result by simultaneously selecting categories from different facets. For example, it is possible to narrow down the search to include only items that contain contact information of different authorities, such as police or tax offices in the city of Helsinki by selecting the *authority contact* category from the content type facet and *Helsinki* from the location facet.

The content of the portal is fully described by ontologies and metadata conforming to them. The information items to be searched for are instances of the `Webpage` class of the portal ontology. Their content is described in terms of ontology resources. When new content is added to the portal, it is annotated by using a set of ontology resources, and when the portal data is reloaded, the new content is added to the browsable content of the portal.

The portal ontologies can be managed with a ontology editor, such as Protégé [16], which was used for developing the portal. However, for the portal maintenance a simple ontology editor is not the most easy to use solution. Once the portal user interface and the taxonomies used for different facets are set, most of the maintenance consists of adding and removing content from the portal. This only involves creation and deletion of `Webpage` instances, and could be achieved with a specialized content management tool for this purpose.

4 Discussion

The goals of the SW-Suomi.fi case study was to enhance content discovery and aggregation in a Dmoz.org- or Yahoo!-like information portal, and test the adaptability of the OntoViews framework for the task. Several lessons were learned:

Facets can be used for helping the user in information retrieval in many ways. First, the facet hierarchies give the user an overview of what kind of information there is in the repository. Second, the hierarchies can guide the user in formulating the query in terms of appropriate concepts. Third, the hierarchies do not suffer from the problems of homonymous query terms. Fourth, the facets

[9] http://www.swi-prolog.org

can be used as a navigational aid when browsing the database content [15]. Fifth, the number of hits in every category that can be selected next can be computed *beforehand* and be shown to the user [14]. In this way, the user can be hindered from making a selection leading to an empty result set—a recurring problem in information retrieval systems—and is guided toward selections that are likely to constrain (or relax) the search appropriately.

Of course, view-based search is not a panacea for information retrieval. According to user tests [17], keyword search, supported also in SW-Suomi.fi, is usually preferred if the user is capable of expressing her information need terms of accurate keywords. Obviously, multi-facet search is also more difficult to understand to citizens than Yahoo!-like single facet search.

SW-Suomi.fi provides the user with a dynamic semantic browsing facility, too. We think that the idea on linking and accumulating semantic content with logic rules is important in many ways. Human editing and maintenance effort can be reduced, because linking is created and enhanced automatically by OntoViews based on metadata and ontologies. The same semantic content can be reused for different applications by just changing the rules and ontologies, without need to modify the actual content. Furthermore, new kinds of content can be incorporated flexibly into the system.

OntoViews turned out to be adaptable to new domains. During the case study, the tool was simultaneously applied to the Dmoz.org RDF-repository, to creating another semantic portal Orava[10] whose content comes the Klaffi e-learning and video portal of the National Broadcasting Company YLE (some 2200 video clips and web pages), and the MuseumFinland RDF(S) knowledge base. Still another case study, CultureSampo[11] extending MuseumFinland content to all kinds of semantic cultural content is underway. A practical shortcoming of the OntoViews tool was that projecting content into views cannot be done incrementally but is done from the scratch after any change to the underlying knowledge base.

Acnowledgements

This work was supported by the National Technology Agence Tekes, the Finnish Ministry of Finance, Fonecta Ltd, TietoEnator Corp., TeliaSonera Corp., and Stakes. Thanks to Pirkko Romakkaniemi, Eetu Mäkelä, Samppa Saarela, Kim Viljanen, Petri Lindgren, and Mikko Laukkanen for discussions.

References

1. Staab, S., others, J.: Semantic Community Web Portals. In: Proceedings of the 9th International World Wide Web Conference, Amsterdam, The Netherlands, Elsevier (2000)

[10] Available at: http://museosuomi.fi/orava/

[11] http://www.cs.helsinki.fi/group/seco/ontologies/kulttuurisampo/

2. Reynolds, D., Shabajee, P., Cayzer, S.: Semantic Information Portals. In: Proceedings of the 13th International World Wide Web Conference on Alternate track papers & posters, New York, NY, USA, ACM Press (2004)
3. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. Scientific American **284** (2001) 34–43
4. Fensel, D.: Ontologies: Silver bullet for knowledge management and electronic commerce (2nd Edition). Springer-Verlag (2004)
5. Maedche, A., Staab, S., Stojanovic, N., Struder, R., Sure, Y.: Semantic portal — the SEAL approach. In Fensel, D., Hendler, J., Lieberman, H., Wahlster, W., eds.: Spinning the Semantic Web. MIT Press, Cambridge, MA (2003)
6. : OWL Web Ontology Language Overview (2004) http://www.w3.org/TR/owl-features/.
7. Decker, S., Erdmann, M., Fensel, D., Studer, R.: Ontobroker: Ontology based access to distributed and semi-structured unformation. In: DS-8. (1999) 351–369 citeseer.nj.nec.com/article/decker98ontobroker.html.
8. Goble, C., Bechhofer, S., Carr, L., Roure, D.D., Hall, W.: Conceptual open hypermedia = the semantic web? In: Proceedings of the WWW2001, Semantic Web Workshop, Hongkong. (2001)
9. Sidoroff, T.: Semantic web portals (in Finnish, Semanttiset portaalit). Master's thesis, University of Helsinki, Dept. of Computer Science (2005) http://www.cs.helsinki.fi/group/seco/publications.
10. Hyvönen, E., Saarela, S., Viljanen, K.: Application of ontology-based techniques to view-based semantic search and browsing. In: The semantic web: research and applications. First European Semantic Web Symposium, ESWS 2004, Heraklion, Greece. (2004) 92–106.
11. Hyvönen, E., Junnila, M., Kettula, S., Mäkelä, E., Saarela, S., Salminen, M., Syreeni, A., Valo, A., Viljanen, K.: Finnish Museums on the Semantic Web. User's perspective on MuseumFinland. In: Proceedings of Museums and the Web 2004 (MW2004), Seleted Papers, Arlington, Virginia, USA. (2004) http://www.archimuse.com/mw2004/papers/hyvonen/hyvonen.html.
12. Mäkelä, E., Hyvönen, E., Saarela, S., Viljanen, K.: OntoViews - A Tool for Creating Semantic Web Portals. In: Proceedings of the Third Internation Semantic Web Conference (ISWC2004), Hiroshima, Japan, Springer Verlag (2004)
13. Yee, K.P., et al.: Faceted Metadata for Image Search and Browsing. In: Proceedings of the conference on Human factors in computing syst ems, Ft. Lauderdale, Florida, USA, ACM Press (2003)
14. Pollitt, A.S.: The key role of classification and indexing in view-based searching. Technical report, University of Huddersfield, UK (1998) http://www.ifla.org/IV/ifla63/63polst.pdf.
15. Hearst, M., Elliott, A., English, J., Sinha, R., Swearingen, K., Lee, K.P.: Finding the flow in web site search. CACM **45** (2002) 42–49
16. Knublauch, H., Fergerson, R.W., Noy, N.F., Musen, M.A.: The Protégé OWL Plugin: An Open Development Environment for Semantic Web Applications. In: Proceedings of the Third Internation Semantic Web Conference (ISWC2004), Hiroshima, Japan, Springer Verlag (2004)
17. English, J., Hearst, M., Sinha, R., Swearingen, K., Lee, K.P.: Flexible search and navigation using faceted metadata. Technical report, University of Berkeley, School of Information Management and Systems (2003)