A Semantic View-Based Portal Utilizing Learning Object Metadata

Teppo Känsälä and Eero Hyvönen

Semantic Computing Research Group (SeCo)
Helsinki University of Technology (TKK), Laboratory of Media Technology, and
University of Helsinki, Department of Computer Science
firstname.lastname@tkk.fi
http://www.seco.tkk.fi/

Abstract. This paper presents a case-study of converting the content of an existing eLearning portal based on the Learning Object Metadata (LOM) format into a semantic portal. It is shown how the content can be provided for the end-user from different semantic perspectives (views) based on a set of ontologies whose concepts are extracted from the original metadata. A semantic recommender system based on the same ontologies and metadata enhances the system usability further. This recommender system links the learning objects of the portal with each other with short labels that explain to the end-user the reason for the links, which is especially important in the eLearning domain. Furthermore, the system links the material also with the content of another cultural semantic portal, providing interportal semantic links. The application, consisting of some 2200 video and audio clips as learning objects, is available on the web where it can be compared with the original portal.

1 Introduction

The Semantic Web enables new ways to design and implement "semantic portals" [7] with improved, content-based searching and linking capabilities [10]. In this paper we present the case-study of creating a semantic eLearning portal called Orava[1]. It is based on the content of an existing traditional portal Klaffi[2] maintained by the Finnish Broadcasting Company YLE Ltd. Klaffi contains some 2200 educational video and audio clips annotated by using the Learning Object (LO) Metadata (LOM) standard[3]. The goal of our work was to demonstrate the practical benefits of using semantic web technologies in portals, as developed within the OntoViews tool [8] and the cultural MuseumFinland portal [5], but applied to LOM and in the eLearning domain.

We start by discussing what metadata changes and ontologies were needed to migrate from the original Klaffi portal to its semantic version Orava. The implementation of Orava is then briefly described and compared with Klaffi. Finally, contributions of the system are summarized.

[1] Available at: http://demo.seco.tkk.fi/orava/
[2] http://www.yle.fi/klaffi/
[3] http://ltsc.ieee.org/wg12/

2 Creating ontologies for LOM metadata

On the Semantic Web we need content described in a machine understandable way. In our case this meant that the literal content descriptions used in the XML-structured LOM metadata of Klaffi had to be converted into resources (URIs) whose meaning is represented in terms of Semantic Web standards, such as RDF and OWL.

LOM consists essentially of a set of standardized metadata fields whose values are filled when annotating learning objects. LOM can be seen as a kind of Dublin Core (DC) application profile for eLearning applications. As in DC, there may be some constraints on how the values should be encoded, e.g. how to encode time, but in general LOM specifies only metadata field names and not the content to be used as field values.

In the Klaffi material, the following LOM fields were considered particularly important from the end-user's perspective: *description* (explaining the content textually), *keyword* (describing the content using a controlled vocabulary), *typicalAgeRange* (main target age group), *difficulty* (of the material for the target group), *format* (type of the material), and *partOf* (telling the larger context in which the learning object belongs to).

Converting such a metadata schema and metadata syntactically into RDF is in principle simple: Each LO can represented by a new RDF-resource R. Each LOM field can be represented by a property P of R, and each value in LOM metadata becomes a value V in RDF, thus forming RDF-triplets of the form $< R, P, V >$. V can be a literal value or an RDF-resource in an ontology depending on the nature of the property P. A key idea of the transformation was that each LOM metadata field was associated with a corresponding ontology. If V is an RDF-resource, then it is taken from the ontology corresponding to the field. In this way we could keep the content descriptions uniform and interoperable across the whole Klaffi database, and even with other external knowledge bases. The idea of systematically using ontological RDF-resources as field values instead of literal values is also a prerequisite for applying the view-based search paradigm [9, 2, 6] in the final user interface. Each field type of interest to the end-user can be used as an orthogonal, ontological classification view along which the LOs can be projected.

The main problem in transformation was that ontologies for LOM field values were not available, but had to be created. In our case the fields *difficulty*, *typicalAgeRange*, and *format* had a limited range of possible values and were easily ontologized by hand. We only needed to create small ontologies based on the terms used in field values, describing how the literal values in the original LOM-metadata are mapped onto the corresponding ontology resources organized into taxonomies. The views for the user interface could be projected later straight forward from the ontologies, too.

The *keyword* field in LOM was more challenging as it can contain virtually anything. Creating a proper ontology and mapping the values of the keyword field onto this ontology would have been an exhausting and resource intensive process. To create a meaningful ontology one needs to have a clear understanding what the ontology is used for in order to model problem domain. In the case of the keyword field, the purpose is to describe the content subject of the LO by a few keywords. We noted that most video and audio clips have some school subject, such as "mathematics" or "history", as a keyword. This observation and the fact that school subject is an import search criteria for most students, gave us a reason to believe that it's reasonable to create an additional ontology for school subjects. The new ontology is used as a new view in the Orava-portal user interface, where the user can constrain search by selecting school subjects of interest to her.

In addition to the *school subject* ontology a *theme ontology* was created based on the values found in the *keyword* metadata field. In practice, we extracted most frequent keywords used in the metadata, and created the theme ontology by using concepts underlying the frequently used keywords. In this way we could guarantee that all relevant concepts used in the metadata were included in the ontology. If we had created a general ontology without considering the actual keywords, much work would have been wasted, because usually only concepts that are really used in the metadata are of value in views of the user interface, and are needed in creating the semantic recommendation links. If annotation frequencies were not considered, there would be the danger that the LOs would be unevenly distributed over a general ontology. i.e., there would be ontological concepts that relate only to one video or ontological concepts that relate to almost all of the videos. Such categories and unevenly distributed material would not be very useful in view-based search.

3 Semantic search, browsing, and autocompletion

The Orava portal was implemented using the OntoViews tool[4] [8]. OntoViews consists of the following components: 1) Ontodella [12] is a logic server that provides semantic linking service and facilities for the projection of content objects, in our case LOs, to views based on the underlying ontologies. 2) Ontogator is a view-based multi-facet search engine. 3) OntoViewsC is a system built on top of the Apache Cocoon[5] for transforming the XML and RDF data provided by Ontodella and Ontogator into the final user interface.

Orava portal (and OntoViews) is based on the *view-based search* paradigm [9, 1, 13] that has been generalized for the Semantic Web in [3, 5]. The user interface in view-based search has several views (facets), offering different perspectives into the content. In our case, a view (facet) is a category hierarchy, such as "Theme", where each category represents an ontology concept. Other "semantic" features of Orava are semantic recommendations and semantic autocompletion, that is able to automatically complete user written textual words into matching ontological categories for semantic search [4].

[4] Available at http://www.seco.tkk.fi/projects/semweb/dist.php
[5] http://cocoon.apache.org/

3.1 Semantic search

The views are created by projecting view hierarchies from the ontologies, and by mapping RDF LOs onto them by using logic rules written in SWI Prolog[6]. The rules are defined freely based on the underlying annotation schema, and can be adapted to different domains and annotations of varying complexity. In our case, the subsumption hierarchies of the ontologies created provide directly natural taxonomies that could be shown to the end-user as views. This might not be the case with complex ontologies, but the separation of ontologies from user viewable category hierachies enable flexible development in such cases, too.

The following views were created for Orava: 1) "Oppiaine" (School subject) 2) "Teema" (Theme of the LO) 3) "Kohderyhmä" (Target audience by education level). 4) "Vaikeustaso" (Difficulty level) 5) "Mediatyyppi" (Media type) 6) "Ohjelmasarja" (Program series)

Figure 1 presents the Orava user interface depicting one video LO. The navigational search views are seen on the left. A category is selected by clicking on a subcategory link listed on the view. This initiates view-based search and shows to the user the next level of choices available on the view. The numbers in parentheses show pro-actively the number of hits if the link is selected next for constraining the search. Categories that produce no results are hidden on the user interface, which eliminates dead-end situations during searching.

A benefit of the view-based search paradigm is that user can select categories of different views *simultaneously*. Each selected category works as search constrain so that only video and audio clips belonging to all selected categories are shown. For example, video series "Abitreenit" consists of programs targeted to students taking their matriculation exam. A student wanting to practice his math skills for the matriculation exam can easily select categories "Abitreenit" and "Matematiikka" (mathematics).

3.2 Semantic recommendations

Semantic recommendations are links to related video and audio clips that might be of interest to the user when viewing a LO page. Orava contains two kind of recommendation links: 1) links to other video and audio clips inside Orava and 2) inter-portal links to the pages of the semantic portal MuseumFinland[7], that is based on an RDF(S) repository of cultural artifacts from Finnish museums and sites. A goal of developing Orava was to demonstrate how separate portals can share knowledge based on ontologies, ontology mappings, and semantic web technologies.

In figure 1 the user is viewing information about a video that is shown in the upper part of the middle section. Below this metadata, there are the semantical links to other video and audio clips, and on the bottom the inter-portal links to MuseumFinland, indicated by a prefix. The titles in bold font above the links

[6] http://www.swi-prolog.org
[7] http://www.museosuomi.fi/

explain shortly the reason why these particular links were recommended. In this case, the video contains a speech given by President of Finland during the Olympic Games in Helsinki in 1952. The recommendation titles tell to the user “Links to videos of the same subject Olympic Games” in Orava, and “Collection artifacts related to sports in MuseumFinland”.

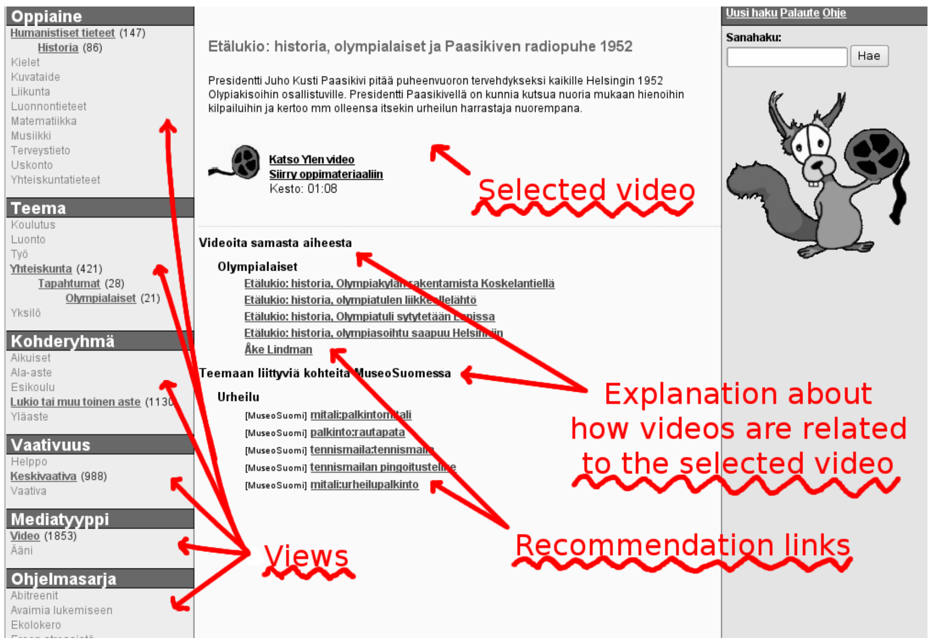

Fig. 1. The Orava user interface.

The following semantic recommendation link rules were implemented: 1) Link videos that share a common theme and the target audience. 2) Link videos based on the LO series relation, such as the ”next part of series”, ”previous part of series” and ”same part of series”. 3) Link videos to artifacts in MuseumFinland if they share a common theme. 4) Link videos to artifacts in MuseumFinland if they share the “place of manufacture” or “place of usage”.

A problem encountered in creating the inter-portal links between Orava and MuseumFinland was that the systems use different ontologies as they focus on different domains. It was therefore necessary to map Orava’s ontological concepts to MuseumFinland’s concepts. The mapping was achieved by using a string matching for matching concepts based on the labels of the concepts, and by checking the validity of the mappings manually. This has the unfortunate consequence that when a matching concept isn’t found, which is highly probable when ontologies are of different domain, one cannot create inter-portal links.

3.3 Semantic autocompletion

The original Klaffi portal makes use of free text search, as customary on the current web. Orava provides the user with semantic view-based search but research has shown [1] that in many cases text based search is also needed. Semantic autocompletion [4] is an approach to combine the benefits of the both approaches.

Figure 2 shows an example of search with the on-the-fly autocompletion feature of Orava, where the system tries to complete input text to semantically matching categories in the views. Here we see that as soon as user has typed in the string "str", the portal immediately shows matching categories and their parent categories in the form of category-trees under the search box. In this way, the user gets immediate feedback allowing her to discover topics at the time of formulating a search query. In addition, a few links for actual videos matching the search are shown allowing user to similarly discover actual video and audio clips.

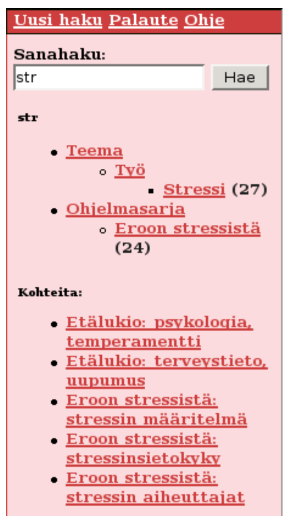

Fig. 2. The upper-right corner of the Orava user interface, where user has written "str" in the search box. Semantic autocompletion shows matching categories and videos as links for the user to select.

Ordinary free text search might give us false hits. For example if one selects the keyword "venäjä" in Finnish, this can mean either Russia as a country or Russian as a language. As a result, the hit list contains results like history of Russia or culture of Russia even if the user wanted results for the Russian language. The semantic autocompletion of Orava can help in disambiguating the meanings by listing the choices to the end user. In this case, the country and and language are found in different views which disambiguate the meanings, and the user can select the right interpretation and initiate the search. At the same time, Orava is able to do free text search based on the textual field values, if ontological resources in the LOM metadata are not available, and in this way combine the both search paradigms. All potentially matching items can be shown on the fly in addition to providing the user with the possibility of disambiguating meanings.

4 Discussion

According to [11] a good search scheme involves: 1) Narrowing and broadening of the search query. 2) Selection of vocabulary appropriately. 3) Ability to modify the search query. The semantic view-based search paradigm seems to support these demands. It allows user to narrow and broaden interesting topics as seen in view hierarchies, and modify the query. The views also give the user the search vocabulary, and she immediately gets a general idea of the portal content along the views, which makes query formulation much easier. The user might focus, for example, on the topic "Nature", select the topic, notice that it has interesting subtopics like "Protection of Nature", refine the search further, and then go on refining the search along another view, such as "Target audience". Furthermore, the number of hits counted for each category selection guides search towards result sets of appropriate size. In the original Klaffi, for example, this kind of flexibility is impossible, because there are only few predefined topic links that result into free text search, and the only possibility to further refine the search is to make a completely new free text search. Once the user has found a video or audio clip of interest, following semantic recommendation links allows her to discover other possibly interesting video and audio clips. This feature, based on ontological reasoning, is completely missing from the original Klaffi portal.

From the engineering viewpoint, the clear distinction of data, logic, and presentation in Orava allows good maintainability of the system. By simply changing the logic rules in one part of the system one can change what semantic recommendation links the user sees. No changes in the content are needed. In the same way, by only altering view projection rules one can change the category hierarchies shown to the user. A benefit of the semantic web approach used is that the recommendation links and views with category hierarchies can be updated automatically when new video and audio clips are added or ontologies modified. A shortcoming of the current implementation is its inability to do this on the fly, but this limitation is due to the used tools to build the portal and are not inherent in the general approach.

Acknowledgements Orava was originally implemented by Juhani Bergström, Peter von Etter, Teppo Känsälä, Olli Lyytinen, Jessika Penttinen, and Mikko Waris. Teemu Sidoroff and Hannu Räisänen contributed by guiding the implementation work. Our work was partly funded by the Finnish Funding Agency for Technology and Innovation (Tekes)

References

1. J. English, M. Hearst, R. Sinha, K. Swearingen, and K.-P. Lee. Flexible search and navigation using faceted metadata. Technical report, University of Berkeley, School of Information Management and Systems, 2003. Submitted for publication.
2. M. Hearst, A. Elliott, J. English, R. Sinha, K. Swearingen, and K.-P. Lee. Finding the flow in web site search. *CACM*, 45(9):42–49, 2002.
3. E. Hyvönen, S. Saarela, and K. Viljanen. Application of ontology-based techniques to view-based semantic search and browsing. In *Proceedings of the First European Semantic Web Symposium, May 10-12, Heraklion, Greece*. Springer–Verlag, Berlin, 2004.
4. Eero Hyvönen and Eetu Mäkelä. Semantic autocompletion. In *Proceedings of the 1st Asian Semantic Web Conference*, 2006. To be published.
5. Eero Hyvönen, Eetu Mäkelä, Mirva Salminen, Arttu Valo, Kim Viljanen, Samppa Saarela, Miikka Junnila, and Suvi Kettula. Museumfinland – finnish museums on the semantic web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 3(2–3):224–241, Oct 2005.
6. Eero Hyvönen, Samppa Saarela, and Kim Viljanen. Application of ontology techniques to view-based semantic search and browsing. In *The Semantic Web: Research and Applications. Proceedings of the First European Semantic Web Symposium (ESWS 2004)*, 2004.
7. A. Maedche, S. Staab, N. Stojanovic, R. Struder, and Y. Sure. Semantic portal — the SEAL approach. In D. Fensel, J. Hendler, H. Lieberman, and W. Wahlster, editors, *Spinning the Semantic Web*. MIT Press, Cambridge, MA, 2003.
8. E. Mäkelä, E. Hyvönen, S. Saarela, and K. Viljanen. OntoViews - A Tool for Creating Semantic Web Portals. In *Proceedings of the Third Internation Semantic Web Conference*. Springer Verlag, November 2004.
9. A. S. Pollitt. 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.
10. Dave Reynolds, Paul Shabajee, and Steve Cayzer. Semantic Information Portals. In *Proceedings of the 13th International World Wide Web Conference on Alternate track papers & posters*. ACM Press, May 2004.
11. AG Sutcliffe, M. Ennis, and SJ Watkinson. Empirical studies of end-user information searching. *Journal of the American Society for Information Science*, 51(13):1211–1231, 2000.
12. Kim Viljanen, Teppo Känsälä, Eero Hyvönen, and Eetu Mäkelä. Ontodella - a projection and linking service for semantic web applications. In *Proceedings of the 17th International Conference on Database and Expert Systems Applications (DEXA 2006), Krakow, Poland*. IEEE, September 4-8 2006. To be published.
13. Ka-Ping Yee, Kirsten Swearingen, Kevin Li, and Marti Hearst. Faceted metadata for image search and browsing. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 401–408, New York, NY, USA, 2003. ACM Press.