OntoWeb Portal: Content Upkeep

Peter Spyns, Andriy Lisovoy & Robert Meersman
Vrije Universiteit Brussel, STAR Lab, Pleinlaan 2, B-1050 Brussel, Belgium
{Firstname.Lastname}@vub.ac.be

Jens Hartmann, York Sure & Rudi Studer
Institute AIFB, University of Karlsruhe, 76128 Karlsruhe, Germany
{lastname}@aifb.uni-karlsruhe.de

Identifier: 6.5
Class: Deliverable (Demonstrator)
Version: 1.1
Version date: 15-06-04
Status: final/draft
Distribution: Public
Responsible Partner: AIFB, VUB
OntoWeb Consortium

This document is part of a research project funded by the IST Programme of the Commission of the European Communities as project number IST-2000-29243.

Leopold Franzens Universität Innsbruck
Institut für Informatik (IFI)
Next Web Generation Research Group
Technikerstraße 25 6020 Innsbruck, Austria
Tel: +43-512-5076485/8
Fax: +43-512-5079872
Contactperson: Dieter Fensel
E-mail: dieter.fensel@uibk.ac.at
## Revision Information

<table>
<thead>
<tr>
<th>Revision date</th>
<th>Version</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>09/06/2004</td>
<td>1.0</td>
<td>First skeleton and VUB draft</td>
</tr>
<tr>
<td>15/06/2004</td>
<td>1.1</td>
<td>Additions of AIFB and final version</td>
</tr>
</tbody>
</table>
Abstract

This report describes deliverable 6.5: content upkeep. The main goal of the portal is to provide the OntoWeb community members with an integrated access to all kinds of information related to the OntoWeb network. The final version of the OntoWeb portal supports both (conceptual) browsing and querying of the information stored in the portal, as well as content provision and access.

The user is then able to browse and query the information stored in the portal. Querying of the information can be done using either ontology-based search forms or keywords or via a “classical” search interface. In the former case, the system exploits the concepts and relationships from the ontology to allow the user to interact with this information on a conceptual level providing the user with a “mental map” of the available information.

The specific aim of the current deliverable is keep the content of the OntoWeb portal in pace with the general progress of the OntoWeb network, i.e. offer access to all the deliverables, information on network members, about events important to the network or other interesting resources.
## Content Table

OntoWeb Portal: Content Upkeep ..................................................................................... 1
OntoWeb Consortium ........................................................................................................ 1
OntoWeb Consortium ......................................................................................................... 2
Revision Information ........................................................................................................ 3
Abstract ............................................................................................................................... 4
Content Table ...................................................................................................................... 5
Content .................................................................................................................................. 6
1. Introduction ....................................................................................................................... 6
2. Data upkeep (AIFB) .......................................................................................................... 8
3. Meta-data synchronisation (STAR Lab) .......................................................................... 11
4. Conclusion ....................................................................................................................... 14
References ............................................................................................................................ 14
Content

1. Introduction

The OntoWeb portal as such, its GUI, its functionalities, its extended semantic search and browse facilities as well as its latest improvements have already been extensively described elsewhere [1,2,3,4,5]. We refer the reader to the references for more details on the mentioned issues.

The graphical representation of the OntoWeb portal architecture (see Figure 1) provides an overview for the reader to situate the descriptions given in the sections below. It can be clearly seen that the portal has two main components:

1. the Zope Object Server (see [1] for the latest technical description).
2. the Dogma Server (see [5] for a more technical description, to be combined with [3] for the latest information on the GUI)
The former is responsible for storing the data (via syndication or structured forms) and exporting an annotated version of it, while the latter stores the instance (= annotated export) and provides for the semantic (since ontology-based) query facility. Because Zope (and Plone) is a general framework that combines a database with a web environment, one can also query the content of the database directly (see Figure 2). Keeping the content of the portal up to date thus necessarily involves these two main components.

**Figure 2: OntoWeb home page and functionalities offered**

The remainder of this deliverable is structured as follows. In a first section, efforts by partner AIFB to keep the content of the Zope Object Enterprise Server up to date are described (section 2). Subsequently, VUB STAR Lab details how an annotated export of the content is integrated in the DOGMA Instance base server (section 3). A short conclusion ends this deliverable (section 4).
2. Data upkeep (AIFB)

The OntoWeb portal provides several content exchange mechanisms (described in OntoWeb Del. 6.4). These mechanisms are now instantiated and used within several other community portals. Currently the following portals are part of the complete infrastructure which is depicted in Figure 1:

- AIFB portal
- OntoWeb portals
  - OntoWeb.org
  - OntoWeb Edu
  - OntoWeb Roadmap
- planned extension: KM-Vision.org

The AIFB portal, OntoWeb.org and KM-Vision.org are relying on different ontologies A, B and C) since they cover different domains and are targeted to different audiences. The three portals related to the OntoWeb thematic network rely on the same ontology (named B). In fact, the initial ontology for the OntoWeb.org portal has been meanwhile extended to cover also the contents of the other two OntoWeb portals.

To keep the content up-to-date we therefore rely on an automatic support to exchange and synchronize content. From a technical point of view, the three portals rely more or less on the same technology as a baseline, viz. the Open Source ZOPE (http://www.zope.org/) web application server framework and some of its extensions. This baseline technology provides us with basic facilities to automatically synchronize content on the fly between these portals. Content-filters allow to specify on a rather detailed level what will be exchanged. If needed, a publication workflow which involves human reviewers in the loop ensures the quality of the provided content.

Content is provided by different parties for the OntoWeb.org portal. As such, OntoWeb.org serves as a main entry point to numerous heterogeneous data sources which can be classified as described in the following. Firstly, people directly maintain e.g.
personal information, publications etc. by using forms provided as part of the portal. Secondly, they can provide meta data on their Web sites and publish the URL of these Web sites within the OntoWeb.org portal. An RDF/S crawler regularly pushes this content into the portal. Thirdly content of the Edu and Roadmap portals is imported into the main OntoWeb.org portal.

**Examples for Content Upkeep**

We enable knowledge exchange among different portals by taking advantage of sharing the same kind of content, i.e. in our case sharing the same concepts in the domain ontologies. Common concepts in the ontologies of the AIFB portal, OntoWeb.org (and KM-Vision.org) are *events* and *publications*. In a traditional setting one would maintain content related to these concepts separately in each portal. But, since there are overlapping interests, several chances for re-using already provided knowledge occur to reduce the maintenance burden. Particularly we found the following dependencies between the portals:

- Events maintained in OntoWeb.org are also closely related to our working group at the AIFB, but they are typically not related to the other working groups at our institute.
- All publications of our working group at the AIFB should also be available at the OntoWeb.org portal.
- OntoWeb.org and KM-Vision.org are very closely related, actually the same kind of events are published in both.

Exemplary, we demonstrate a real-world scenario in which publication lists of the researchers at our institute AIFB are imported (synchronized) with the OntoWeb.org portal and the KM-Vision portal.

Since each researcher is required by the AIFB administrative to enter each publication in the portal, it provides a large and nearly complete amount of data. These information are then used for the AIFB portal to generate publication lists by month, year and researcher.

For instance, a researcher working at the AIFB wants to publish his latest publications in
the AIFB portal. Therefore, he uses the AIFB publication interface and enters manually the publications respectively uploads a BibTex file. As a result, the AIFB portal generates publication lists of each author and corresponding co-authorship whereby the portal distinguish between internal and external authors. Internal authors are all researchers working at the AIFB while external authors are well-known colleagues from other institutes or organisations. In case of an internal author or co-author the portal adds this publication to the publication list of the co-author, too.

The scope of the OntoWeb.org portal covers the work of the AIFB. Therefore, the OntoWeb.org portal imports the publication metadata from the AIFB. Whereby the OntoWeb.org portal provides a similar import interface for publications.
3. Meta-data synchronisation (STAR Lab)

Up till now there was no software connection between the data store (AIFB) and the meta-data store (VUB) that constitute the two building blocks of the OntoWeb portal (see Figure 1). As the content in the ZOPE server didn’t change frequently, this has not been a major issue. Also, the user could always use the ZOPE-based interface to look for information. However, as the content in the portal grew and updates became more frequent, a mechanism had to be foreseen to synchronise the instance server of the VUB part with the content of the AIFB ZOPE server (see Figure 3). On the latter side, a specific export mechanism has been arranged: the content of the ZOPE DB is dumped into six RDF files and made available on the web. On the former side, a combination of a directed crawler [7] and an import facility [6] based on a SOAP-connection has been implemented. The crawler fetches the meta-data files from the site (http://ontoweb.aifb.uni-karlsruhe.de/CMI) and transmits them via SOAP to the import module of the instance base. If needed, the crawler could also visit other site to collect RDF files. As no checks are done regarding the ontology used for the annotation tags, this is not recommended. The importing module consists of some Enterprise Java Beans that take care of actually storing the RDF triplets in the instance base. The new meta-data simply replaces the previous content. It is easy (using the Unix/Linux “cron” command) to automatically activate the crawler.

The result of activating the crawl & import facility is illustrated by Figure 4. It shows one of the five pages with results that are classified as “events”. Remark that the “Second
European Summer on Ontological Engineering and the Semantic Web” appears twice (once in capitals with acronym and once non capitalised without acronym).

Figure 4: OntoWeb semantic search/navigation facility showing synchronised content

When comparing the content of this page with the original content of the ZOPE DB, we also see two entries with the same typographic characteristics (one with and one without acronym), which constitutes a clear proof that the content of the two servers is identical –
at least the meta-data produced on basis of the common ontology. E.g., concerning educational resources another (more granular) ontology has been used to annotate. The resulting annotations (originating from the OntoWeb Edu portal by partner UO) and metadata are not visualized by the semantic browser – even if they have been entered in the instance base – as the labels of these meta-data tags are not included in the ontology agreed upon by AIFB and VUB.

Figure 5: OntoWeb “classic” search/navigation facilities
4. Conclusion

On basis of its architecture and functionalities, the OntoWeb portal constitutes a very good example of a semantic portal in general. This deliverable explains to what extent and in which way the content (data and meta-data) of the portal has been kept up to date during the last phase of the OntoWeb project.

References