D 2.3.5 b Consensus Making Environment

Anna V. Zhdanova (UIBK)

The author thanks for discussions
Stefan Decker (NUIG), Carlos Enguix (NUIG), Robert Stevens (UoM),
Max Völkel (UKARL) - alphabetically

Abstract.
Two main aspects presented in this deliverable are as follows: (i) development of a prototype for a framework allowing and motivating consensual collaborative ontology construction and (ii) application of the developed infrastructure to scenarios on Semantic Web community portals with involvement of real users.
Keyword list: ontology construction, Web communities, social networking
EU-IST Network of Excellence (NoE) IST-2004-507482 KWEB
Deliverable D2.3.5 b v2.01 (WP2.3)

Document Identifier:	KWEB/2006/D2.3.5 b
Class Deliverable:	KWEB EU-IST-2004-507482
Version:	v2.01
Date:	February 08, 2006
State:	Final
Distribution:	Public

Knowledge Web 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-2004-507482.

University of Innsbruck (UIBK) – Coordinator Institute of Computer Science, Technikerstrasse 21a A-6020 Innsbruck Austria Contact person: Dieter Fensel E-mail address: dieter.fensel@uibk.ac.at	**École Polythechnique Fédérale de Lausanne (EPFL)** Computer Science Department Swiss Federal Institute of Technology IN (Ecublens), CH-1015 Lausanne. Switzerland Contact person: Boi Faltings E-mail address: boi.faltings@epfl.ch
France Telecom (FT) 4 Rue du Clos Courtel 35512 Cesson Sévigné France. PO Box 91226 Contact person : Alain Leger E-mail address: alain.leger@rd.francetelecom.com	**Freie Universität Berlin (FU Berlin)** Takustrasse, 9 14195 Berlin Germany Contact person: Robert Tolksdorf E-mail address: tolk@inf.fu-berlin.de
Free University of Bozen-Bolzano (FUB) Piazza Domenicani 3 39100 Bolzano Italy Contact person: Enrico Franconi E-mail address: franconi@inf.unibz.it	**Institut National de Recherche en Informatique et en Automatique (INRIA)** ZIRST – 655 avenue de l'Europe – Montbonnot Saint Martin 38334 Saint-Ismier France Contact person : Jérôme Euzenat E-mail address: Jerome.Euzenat@inrialpes.fr
Centre for Research and Technology Hellas / Informatics and Telematics Institute (ITI-CERTH) 1st km Thermi – Panorama road 57001 Thermi-Thessaloniki Greece. Po Box 361 Contact person : Michael G. Strintzis E-mail address: strintzi@iti.gr	**Learning Lab Lower Saxony (L3S)** Expo Plaza 1 30539 Hannover Germany Contact person: Wolfgang Nejdl E-mail address: nejdl@learninglab.de
National University of Ireland Galway (NUIG) National University of Ireland Science and Technology Building University Road Galway Ireland Contact person: Christoph Bussler E-mail address: chris.bussler@deri.ie	**The Open University (OU)** Knowledge Media Institute The Open University Milton Keynes, MK7 6AA United Kingdom. Contact person: Enrico Motta E-mail address: e.motta@open.ac.uk
Universidad Politécnica de Madrid (UPM) Campus de Montegancedo sn 28660 Boadilla del Monte Spain Contact person: Asunción Gómez Pérez	**University of Karlsruhe (UKARL)** Institut für Angewandte Informatik und Formale Beschreibungsverfahren – AIFB Universität Karlsruhe D-76128 Karlsruhe

E-mail address: asun@fi.upm.es	Germany Contact person: Rudi Studer E-mail address: studer@aifb.uni-karlsruhe.de
University of Liverpool (UniLiv) Chadwick Building, Peach Street L697ZF Liverpool United Kingdom Contact person: Michael Wooldridge E-mail address: M.J.Wooldridge@csc.liv.ac.uk	**University of Manchester (UoM)** Room 2.32. Kilburn Building, Department of Computer Science, University of Manchester, Oxford Road Manchester, M13 9PL United Kingdom Contact person: Carole Goble E-mail address: carole@cs.man.ac.uk
University of Sheffield (USFD) Regent Court, 211 Portobello street S14DP Sheffield United Kingdom Contact person: Hamish Cunningham E-mail address: hamish@dcs.shef.ac.uk	**University of Trento (UniTn)** Via Sommarive 14 38050 Trento Italy Contact person: Fausto Giunchiglia E-mail address: fausto@dit.unitn.it
Vrije Universiteit Amsterdam (VUA) De Boelelaan 1081a 1081HV. Amsterdam The Netherlands Contact person: Frank van Harmelen E-mail address: Frank.van.Harmelen@cs.vu.nl	**Vrije Universiteit Brussel (VUB)** Pleinlaan 2, Building G10 1050 Brussels Belgium Contact person: Robert Meersman E-mail address: robert.meersman@vub.ac.be

Work package participants

The following partners have taken an active part in the work leading to the elaboration of this document, even if they might not have directly contributed writing parts of this document:

- National University of Ireland Galway (NUIG)
- University of Innsbruck (UIBK)
- University of Karlsruhe (UKARL)
- University of Liverpool (UniLiv)
- Vrije Universiteit Amsterdam (VUA)
- Institut National de Recherche en Informatique et en Automatique (INRIA)
- Free University of Bozen-Bolzano (FUB)
- Centre for Research and Technology Hellas / Informatics and Telematics Institute (ITI-CERTH)
- University of Sheffield (USFD)

Changes

Version	Date	Author	Changes
1	09-10-2005	Anna V. Zhdanova	Public internal version
1.1	11-10-2005	Anna V. Zhdanova	Update after QC comments
2	13-15-12-2005	Anna V. Zhdanova	Second version
2.01	08-12-2005	Anna V. Zhdanova	QA comments integrated

Executive Summary

The goal of the work reported here is to identify current limitations of community portals, introduce consensual community-driven ontology management as a new approach to ontology construction and demonstrate the added value to community portals of being community-driven.

Three main parts of this work are (i) development of a framework allowing and motivating collaborative ontology construction and reuse for the final user (a person and a community), (ii) building a prototype on the basis of this specification, namely the People's portal, and (iii) application of the developed infrastructure to scenarios on Semantic Web community portals with involvement of real users.

The scope of work on the framework for community-driven ontology management is in enrichment with community-supporting features the established practices for ontology management in the areas of ontology development and population, storage, alignment and versioning. The objective of community-driven ontology management is to provide means and motivations for a large number of users to "weave" and adopt the Semantic Web.

The consensus making environment (aka People's portal infrastructure) allows end users to define the content structure (i.e., develop ontologies), populate ontologies and define the ways the content is managed on Semantic Web community portals where the People's portal infrastructure is applied. Content management features on the People's portal include personalization support, dynamic reaching of a consensus on the basis of heterogeneous ontologies.

The People's portal was deployed as a part of an intranet at DERI – Digital Enterprise Research Institute and as an extension to the portal of a Semantic Web community (KnowledgeWeb network of excellence). In all the empirical studies, the community's response and behavior were observed.

In conclusion, comparison to the functionalities of the existing (Semantic) Web community environments approaches and the empirical results prove feasibility and the advantages of community-driven ontology management. Empirically, communities were capable to introduce on the community portals such ontology items as Classes, Subclasses, Properties, Instances, ontology mappings, and reuse these items afterwards.

Contents

Executive Summary **6**

1 Introduction **8**

1.1 Motivation 8
1.2 Content and Scope of Work 11
1.3 Relation to the State of the Art 15
1.4 Structure of the Document 18

2 Community-Driven Ontology Management **20**

2.1 Bottom-Up Approaches – Why and How They Work 20
2.2 Community-Driven Ontology Development and Population 26
2.3 Consensus Modeling 28
2.4 Social Networking Model of a Community 34

3 Implementation and Case Studies **41**

3.1 People's Portal - Framework Implementation 41
3.2 Case Study 1 – Digital Enterprise Research Institute 46
3.3 Case Study 2 – KnowledgeWeb 52

4 Results and Related Work **60**

4.1 Results for Case Study 1 (DERI) 60
4.2 Results for Case Study 2 (KnowledgeWeb) 63
4.3 Related Work 70

5 Conclusions and Future Work **73**

5.1 Further Applications for Community-Driven Ontology Management: Gene Ontology Community 73
5.2 Conclusions 77

Acknowledgements **79**

References **79**

1 Introduction

"The reasonable man adapts himself to the world;
the unreasonable one persists in trying to adapt the world to himself.
Therefore all progress depends on the unreasonable man."
George Bernard Shaw

In the first, introductory chapter of this deliverable, I start with motivation of my work by expressing an outlook on the state of the art in the area of community Web portals and identification of current limitations there. Further, I define content and scope of the work, relate my work to the state of the art and give an outline of the deliverable.

1.1 Motivation

Nowadays, a multitude of related to business or leisure community portals has been created (e.g., “Yahoo Groups”, “Orkut”, “Ecademy”, “LinkedIn”, 43things.com, flickr.com among highly popular ones), and many community portals have proved to be highly popular and successful by acquiring millions of members [O'Murchu et al., 2004]. However, the existing community Web portals are rather inflexible when it comes to specification of user profiles, the content of the portals, the ways this content is organized, and search options. Currently, the existing community Web portals simply specify what and how the users can contribute there and search about. The specification comes from the Web portal creators and their view of the domain, which is usually comprehensive, but is definitely limited, and thus, makes the portal out of interest for the users after they want to go beyond this view. However, a far larger degree of the portal’s flexibility and adaptation to the portal’s member’s real demands can be achieved by bringing the Semantic Web technologies [Berners-Lee, 2001] to the existing community Web portals. Intelligent application of Semantic Web technologies would allow the portal’s members to specify what their (informational) demands are and how to fulfill these demands.

In this work, I address the challenge of giving the control over the domain ontologies in the core of community portals to the communities themselves. The typical resulting functionality enabling a community to construct its own ontology is illustrated with the scenario below. The scenario is specific, but it should be viewed as an instance of a general pattern of an arbitrary information exchange that can take place at community portals.

The scenario is as follows: John and Mary are registered at a community Web portal in a social networking domain, where they have their profile information (such as name, surname, e-mail addresses, hobbies, etc). Assume that the Web portal does not have an ontology attribute (slot in the profile form) to allow the users specify their phone numbers, but John and Mary want to exchange the phone numbers between each other. A screenshot of a member profile at a community portal without a community accessing ontology management functionalities is shown in Figure 1.

On the current portals the scenario usually runs as: John and Mary will have to use e-mail or an instant messenger or some other available mediums, but no longer the portal and its functionality. Further, if all the communities of John and Mary exchange all non-semantically annotated data by e-mail or any other chosen medium, the community members will obviously suffer from the numerous e-mails they get and/or overload with interfering requests for information and huge workflows of irrelevant data.

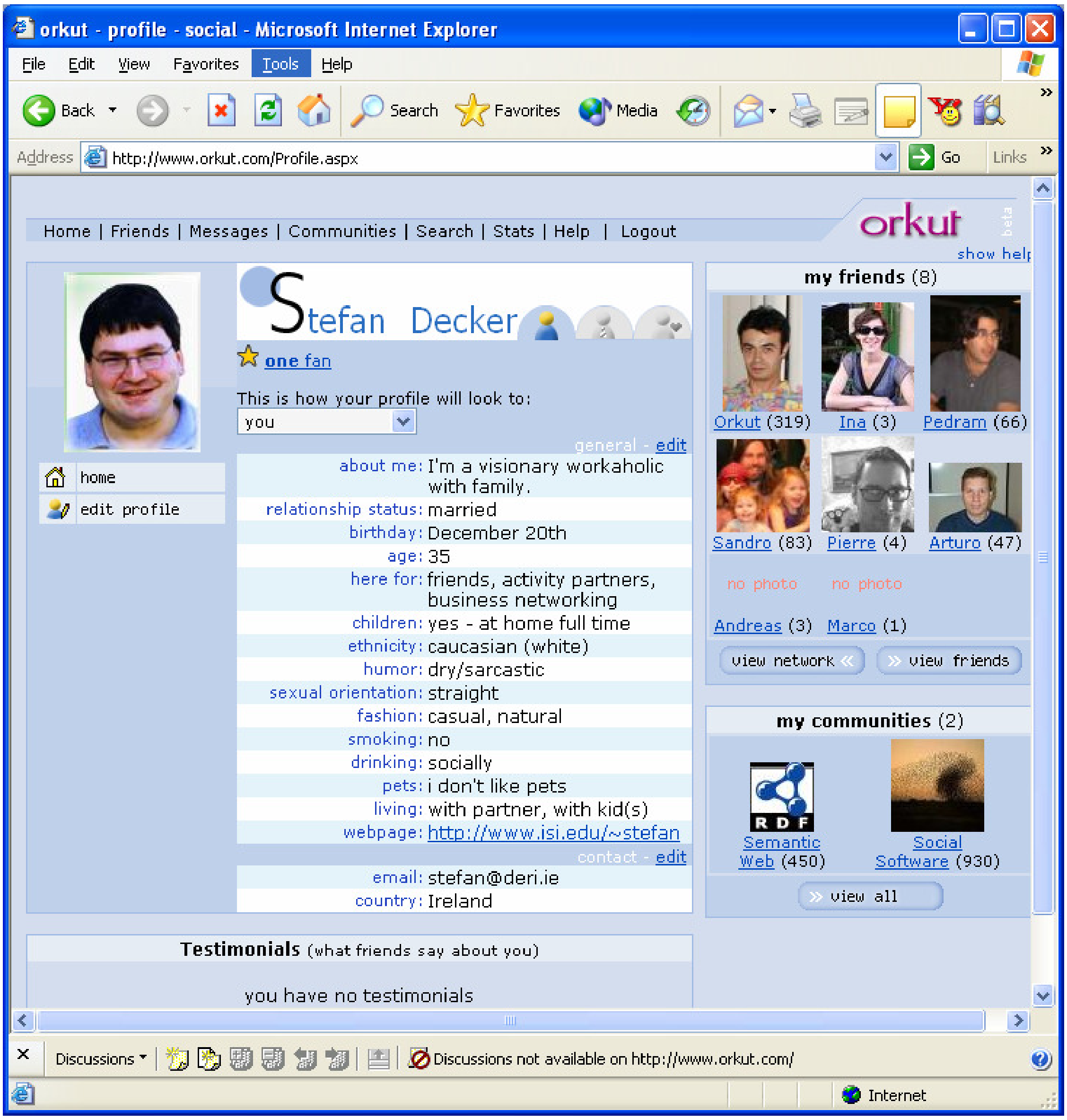

Figure 1: Member Profile on a Community Portal

On the portals supporting community-driven ontology management introduced in this deliverable the scenario includes the following steps: Using the ontology management functionalities integrated into the Semantic Web portal environment, John goes to the community portal and creates a new concept "phone number" in a community ontology. Then he fills his phone numbers in his profile (he can do this now, since the concept is introduced by him). Mary reuses the concept introduced by John and fills in her phone numbers in her user profile. Thus, John and Mary can exchange information (e.g., easily find each other's telephone numbers on request or have them delivered by default) using the Semantic Web portal functionalities, without the need to involve any external mediums.

The obviously foreseen outcome of the proposed approach to the users' broad access to the content ontology editing options is that the ontologies of the Semantic Web community portal run into risk of becoming badly structured, unreliable, far too large and redundant to support the activities of the community and an individual efficiently. Handling this problem by making the ontological structure adapted to the human communication formalisms by organizing an efficient way to operate and render ontological data is a challenge for the practical use of our approach. Specifically, this and similar challenges are addressed in Chapter 2 of this work (Community-Driven Ontology Management).

Optimization of the community-driven ontology management activities are based on development of personalization and community support on Semantic Web. The objective of the traditional personalization is fairly straightforward. It is to deliver information that is relevant to an individual or a group of individuals in the format and layout specified and in time intervals specified. While personalization was applied extensively on the ordinary Web for the individual users (especially in eCommerce area) [Aggarwal et al., 2002; Instone, 2004; Schiaffino and Amandi, 2004], the studies for community and consensus aspects of personalization in the Semantic Web context are still lacking. Meanwhile, semantic representation of the data on the Web and social networking information provide an immense potential of for formalizing personalization processes and employment of developed reasoning techniques to improve implicit personalization. The state of the art in personalization and community fields are mainly in establishing the theoretical basics for the further work on the application level, e.g., developing languages, such as a view language that picks up the unique situation of data in the Semantic Web and allows easy selection, customization and integration of Semantic Web content [Baumgartner et al., 2005]. Certain attempts to specify ontology-based delivery of Semantic Web content were undertaken in the OntoWebber [Jin et al., 2002]. However, the resulting methodology and implementation have been proven to be too complex for understanding by and spreading among the communities, and thus the practice for community-driven ontology management will pursue simpler solutions with wider applicability.

Introduction of community-driven ontology management is important as it adds the following values to conventional management on community portals:

1) Ontology management is an expensive process. In community ontology management, the expenses are shifted from the portal maintainers to the communities employing ontologies. This shift results in ***adequate investment distribution*** among the ontology items (e.g., classes and properties). Specifically, the ontology items of higher importance to the community gain more support in terms of more associated resources, e.g., instance data, granularity in description, cross-ontology mappings.
2) The ontologies which are constructed, aligned and further operated by the communities ***represent the domain and connection with other domains more comprehensibly*** than the ontologies designed and maintained by an external knowledge engineer. External knowledge engineers are typically the bottleneck to the ontology comprehensiveness, as they are not capable to capture all the variety that might take place in a community and associated communities.
3) The community-driven ontology management approach provides a ***higher dynamicity and up-to-dateness*** to the outside-world changes in time, comparing to the conventional ontology management approach. When ontologies are constructed by external knowledge engineers, all the changes need to be captured and introduced by these engineers. With external knowledge experts, the delay in realizing and introducing the changes might take days, weeks or even months. This delay is unacceptable for many dynamic domains, where new terminology regularly and rapidly changes (e.g., business or sport).

1.2 Content and Scope of Work

In order to overcome the limitations identified in the previous subsection, the work within this deliverable has two main objectives:

1) development of a specification for a framework allowing and motivating community-driven ontology management, i.e., ontological data construction and reuse for the final user (a person and a community). The specification must provide solutions for overcoming the limitations of the current Web portals.
2) building a prototype on the basis of this specification, namely the People's portal;
3) applying the prototype to two case studies.

The work advances fields of ontology management by introduction of *community-driven ontology management* (1.2.1), contributes to the practices of *distributed ontology engineering* (1.2.2) and the *implemented framework* (1.2.3) is applied in *case studies* within community portals (1.2.4).

1.2.1 Community-Driven Ontology Management

Ontology management software components essential for supporting basic ontology-based development activities are used Semantic Web applications. Every Semantic Web application needs partial or complete ontology management support. Specification and development of ontology management components were previously funded and carried out in EU and USA projects (in particular, EC IST projects such as DIP[1], SEKT[2], KnowledgeWeb[3], Esperonto[4], SWWS[5]). Progress in development of community Semantic Web environments brings in new positive influence, usage scope and wider acceptability to the basic ontology management components by setting new requirements such as enabling communities manage their own ontologies, making the ontology management knowledge services more flexible, reusable and proven in real-life scenarios thus attractive enough to make the Semantic Web accepted by the communities.

The scope of the work is in reuse of the existing ontology management practices and tools and enriching them with features for supporting community-driven ontology management with applying such management on Semantic Web community portals. One may envision the resulting community-driven ontology management toolkit containing the following components adapted within the scope of community-driven ontology management:

Community-driven Ontology Editing Service: It is an editor for editing ontologies (creating and updating ontology and instances). The front end is the user-friendly interface, which helps or guides users to easily create and update (add, delete, and modify) ontology and its instances. The backend is the data storage management systems, which can be databases, file systems, plain text files. A specific requirement for an ontology editor to be community-driven is an opportunity to integrate it tightly with Semantic publishing and delivery component, and enable consensual editing for multiple users, i.e. communities. This requirement is grounded on flexibility degree that is needed to provide in a community environment enabling community members to change and influence community processes and structures.

Community-driven Ontology Storage and Query Management Service: The goal of this component is to efficiently store and query small and large amounts of ontology data and metadata by providing fast indexing, searching and querying to ontologies and its instances. Most current ontology storing and querying components from the functional perspective are similar to database and database management system components. In addition, the first Semantic Web search engines start to appear (such as Intellidimension Semantic Web search engine[6]). However, there is a long road to go to making Semantic Web database-like components and Semantic Web search engines mature and attractive to use. Taking into account that the communities publish their information on the Semantic Web in a distributed manner in simple ways (such as putting online FOAF files), in project work, the focus in storage and querying will be on maintaining repositories of Semantic Web content and composition/decomposition of distributed

[1] DIP: http://dip.semanticweb.org
[2] SEKT: http://sekt.semanticweb.org
[3] KnowledgeWeb: http://knowledgeweb.semanticweb.org
[4] Esperonto: http://esperonto.semanticweb.org
[5] SWWS: http://swws.semanticweb.org
[6] Intellidimension Semantic Web Search: http://semanticwebsearch.com

content, easy to maintain from the storage and creation point of view, thus involving critical community masses.

Community-driven Ontology Matching Service: The ontology aligner supports ontology mapping processes that now mostly are performed manually with good interface support. A basic ontology inference provides consistency checking, related class or relation name identification, instance updates etc. The front end is the user interface for semi-automatic ontology mapping (such as recommendation lists and help for defining the mapping rules). The back end is the inference support (ontology inference engine). The upgrade of a regular ontology aligner to a community ontology aligner is adding a widely available repository of ontology mapping solutions that result from the usage of the ontology aligner. Special ontologies are used to specify relevance, reusability and reliability of certain ontology mappings from repositories (employing social networking and statistical information). The ultimate goal of the community alignment service activity is to enable knowledge services of external applications to reuse (i.e., gain benefit from) these annotated mapping repositories and alignment services.

Community-driven Ontology Versioning Service: The versioning service represents different versions of the ontologies, including backward consistency support and related instance versioning. The front end provides a report on version information, changes and their effects, for example, the difference of two versions of the ontologies. The back end supports backward consistency in the different versions of the ontologies and their instances update. The Ontology Versioning Service is to be interoperable with Ontology Editor, Ontology Storage and Query Manager and pluggable inference engines for performing additional optional tasks such as checking consistency. On top of the ordinary functionality of an ontology versioning service, a community versioning service needs to have a set of simple understandable interfaces, be available and easily accessible on the Semantic Web, and track the changes taking place in distributed ontologies and instance data sources, reporting relevant inconsistencies and its resolutions to community versioning service users.

Flexibility and *reusability* of the resulting software components are the key requirements in the long run for the work that has a goal to unite digital communities. In particular, these requirements imply that ontology management software components

- should be stand-alone modules, though easy to integrate with each other and additional external widely-spread ontology management components
- should be open source, and disseminated via broad open source development dissemination channels such as sourceforge.net, in order to gain actual usage and influence
- except for the Community-driven Ontology Editing service, should be separated from any presentation functionalities and editing user interfaces to be easily applicable in any Semantic Web community application. The only target user group for all mentioned above ontology management components are Semantic Web community application developers.

1.2.2 Distributed Ontology Engineering

Though the Semantic Web has ontologies and machine-processable instance data in its basis, in practice, the quantity of available ontologies for reuse and sharing is very limited. For example, the resource SchemaWeb[7] (maintained by the IST project SWAD-Europe – the aim of which is to support W3C's Semantic Web initiative in Europe) nowadays is a major resource for publishing ontologies and it links to ca. 250 ontologies only. The linked ontologies are mostly ontologies describing a specific domain (e.g., Person, Publication, Project). The mentioned quantity of available ontologies refers to ontologies specified in multiple existing different ontology languages (e.g., RDFS, OWL). Many of these ontologies are not supported by a large amount of instance data. Some domains are supported by several ontologies (e.g., Person and Publication), while many domains are not supported by ontologies at all. Finally, the number of domain independent ontologies that can be widely applied is negligent, and ontologies for certain aspects like Semantic Web publishing, data delivery and community and personalization support ontologies are not available. All these factors diminish ontology usage and thus success of the Semantic Web.

To overcome the current ontology engineering bottlenecks, two main ontology engineering directions are considered:

- development of domain-independent, easy to use and deploy ontologies in order to support community-driven ontology and instance data management and personalization
- development of domain ontologies, specifically, collaboratively and bottom-up by domain experts irrespectively from their ontology engineering experience

A particular attention is paid to interoperability of the produced ontologies, provision of large amounts of instance data complying to these ontologies and an access to the outside systems to appropriate part of instance data – which makes the ontologies valuable for the outside world. Another side of the dissemination activity is contributing to establishment of best practices for discovery and publishing of ontological content: currently the Semantic Web mainly consists of interlinked RDF documents, and there are no standards of even wide spread conventions neither for embedding Semantic Web into ordinary Web, nor for efficient discovery of the Semantic Web content.

1.2.3 Implementation

The People's portal [Zhdanova, 2004] is an implementation of community-driven ontology management and is focused on letting the Web developers and users to create the Semantic Web content through constructing, populating and using the People's portal, and thus becoming the Semantic Web developers and users. Among key factors of the People's portal prototype potential success in actual production and use of the Semantic Web content is its integration with adding-value community Web portals, i.e., individuals

[7] SchemaWeb: http://www.schemaweb.info

bring their knowledge on the Web in a natural way by actually using the portals. Among the key ideas behind the People's portal is that the Semantic Web is more likely to come true if large user communities are provided with means and motivation to weave the Semantic Web (i.e., bring their knowledge on the Web in a structured way), in a similar way as the means and motivation to weave the Web were provided before [Berners-Lee, 1999]. The mission of the People's portal is to provide the means and motivation to weave the Semantic Web for a large number of Web users.

The People's portal specification and implementation are domain neutral, thus as a solution, the People's portal can be applied to ontologizing numerous domains, from dating to car manufacturing. In this work's prototypes, community Semantic Web portals employing the People's portal are in the domains dealing with people. Specifically, the portals collect and operate with information about people, such as members of a research institute. The domains focused on people were chosen due to an already existing large amount of practices to represent people on the Web (hence popularity, understandability and advanced level to compare with) and superiority of such domains to other domains in the number of Web portal members they can draw.

In order to enable Web users and developers to effectively create Semantic Web content, the People's portal, a solution for community-driven ontology management, is developed here. The solution includes the aspects of distributed, dynamic, weakly coordinated ontology construction, ontology and instance data versioning support and ontology alignment. In addition, specification and implementation of personalization and community support and consensus reaching between parties bearing different ontologies contributes to overcoming heterogeneity and dynamicity that hinder effective ontology reuse.

1.2.4 Case Studies

In order to prove feasibility of community-driven ontology management and identify its limitations, the implementation was applied to two case studies. The first, Digital Enterprise Research Institute case study is a community-driven model of an intranet environment of a research institute. The second, KnowledgeWeb case study, is a community-driven communication medium for a community of researchers working in related fields. Both case studies are implemented and applied as Semantic Web community portals.

1.3 Relation to the State of the Art

The work is aimed at specifying Semantic Web best practices and ontology-based software components for community-driven ontology management that are applicable to wide reuse in community Semantic-based environments. Adapting ontology management software solutions to be applicable to community environments is the work that should be done to make the existing ontology software mature. Enabling ontology management software components work for uniting communities (i.e., making Semantic Web accepted) is the next step after making Semantic Web real, since ontology-based unification and interoperability problem are aimed to be resolved for multiple heterogeneous sources/communities. Thus, the target of the work is to strengthen, support and evolve digital communities and community Semantic Web environments via making Semantic Web best practices widely adopted and Semantic Web accepted with community-driven ontology management.

The scope and the goal of the work imply production of widely understandable and easy-to-use modeling solutions and software for community-driven ontology management, and provision of services making community annotations and ontology management accessible for humans, tools, and applications on the Web. Therefore, the work carried out influences and enhances the state-of-the-art in the following areas:

- **Community-driven ontology management**: Ontology management methodologies such as requirements, guidelines and algorithms for ontology storage, querying, alignment and versioning as well as supporting tools were specified and developed in previous projects (such as IST projects DIP, SEKT, KnowledgeWeb, SWWS, Esperonto, WonderWeb). However, many of these methodologies and tools are at a general-purpose abstract level and were not elaborated for the task they should be used by the definition of Semantic Web - appliance in community environments. In practice, most ontology methodologies and tools are difficult to use for uniting communities due to the lack of features supporting community activities, such as supporting communities in managing and evolving their own ontologies.
- **Making ontology management accepted by the masses**: Once Semantic Web methodologies have been produced, practice and promotion of ontology use are of paramount importance. In particular, W3C launched Semantic Web Best Practice and Deployment Working Group[8] to provide hands-on support to developers of Semantic Web applications. However, the focus of this group is mainly on assisting with the correct usage of recent W3C specifications, but not on establishing Semantic Web community-driven infrastructures that propagate best practices. In this work, flexible ontology management software components are following the real-life use-cases and theoretical methodologies. The resulting components provide novel functionalities that are currently missing in ontology management tools and alleviate the work for developers in applying Semantic Web technologies in practice.
- **Personalization, community and individual support, consensus making**: The state-of-the art in personalization and community support on the Web comprises such techniques as ontology views, collaborative filtering and personalization on the basis of user'/customers' profiles and online traceable behavior [Baumgartner et al., 2005; Instone, 2004; Schiaffino and Amandi, 2004; Won, 2002]. This work brings in new, domain-independent community and personalization ontologies,

[8] SWBPD WG: http://www.w3.org/2001/sw/BestPractices/

frameworks and Semantic Web practices in applying them to heterogeneous, multicultural and multilingual communities and also to individual community members.

- **Semantic Web publishing**: For Web languages such as HTML, CSS and XML that are already widely used, a set of publishing methods, techniques and tools is worked out and widely renown. However, for new knowledge representation formalisms (such as RDF/S and OWL Semantic Web languages and new emerging formalisms), flexible and easy-to-use publishing technologies and tools do not exist. This work contributes to development of publishing technologies, domain-independent presentation and publishing ontologies and tools as well as raise awareness and competence in developed new methods within communities of web developers.
- **Distributed environments**: In environments with distributed character (arising geographical distribution of web-resources, P2P communication such as email and instance messaging), numerous theories and supporting applications were developed (e.g., file sharing networks such as Gnutella, FastTrack, Napster and semantically-enabled Bibster [Haase et al., 2004]). Information exchange and file sharing can take place only upon a condition that a community involved in distributed networking exists and supported.

In Table 1, we list typical representatives of renowned community-related Semantic Web environments with organizations producing them and ontology management components reused. Currently, we observe (1) diversity of employed ontology management components, (2) absence of widely adopted practices for setting up and maintenance of Semantic community environments, (3) lack of community modeling solutions and lack of community-driven ontology management functionalities endowing Semantic Web with growth and added-value, (4) low reusability – an ad-hoc installation is required to get each new instance of an organization's community environment, and in most cases this installation can be performed by the initial developers only, (5) low flexibility – as an instance of environment is installed, its functionalities are hard to combine with functionalities of other environments, (6) absence of cross-community and cross-environment interoperation. Observing the state of the art makes it clear that contributions to the area of community-driven ontology management are absolutely essential for making Semantic Web technologies widely taken up by the developers and ubiquitously used by communities.

URI of a Typical Representative of a Semantic Web Portal	Producer - Organization	Ontology Management Support
http://www.swed.org.uk/swed	Hewlett-Packard, Bristol, UK	Jena
http://knowledgeweb.semanticweb.org	UPM, Madrid, Spain	WebOde
http://www.aifb.uni-karlsruhe.de	UKARL/AIFB, Karlsruhe, Germany	KAON, Bibster, AIFB SEmantic portAL
http://museosuomi.cs.helsinki.fi	University of	*Ontodella* logic

	Helsinki, Helsinki, Finland	server
http://flink.semanticweb.org	Free University of Amsterdam, Amsterdam, the Netherlands	Sesame
http://news.kmi.open.ac.uk/kmiplanet/	The Open University, Milton Keynes, UK	WebOnto

Table 1: State of the Art in Semantic Web Portals

The work has an outcome in terms of easy-to-use and easy-to-adopt community-driven ontology management solutions and provision of a wide outreach of these solutions to the communities. The specific results of the work are:

- **Software components for community-driven ontology management** appropriate **for multi-domain deployment on Semantic community environments**, including innovative practices for security, trust and privacy on the Semantic Web
- Domain-independent ontology-based infrastructures for **personalization and community support** with features of collaborative ontology construction and reuse, and social networking
- Human user interfaces and services for tools and applications for **community aware interoperation, distributed accomplishment of tasks**
- Domain-independent ontology-based methodologies for remote **aggregation, publishing and delivery** of Semantic Web (meta)data with a specific consideration of personalization and community support traits for content aggregation from multiple distributed sources
- **Wide spread** of the best practice solutions among web communities within and beyond the scope of the case studies of the project and human-computer interaction **research reporting successes and challenges in adaptation of community-driven ontology management by human user.**

1.4 Structure of the Document

The deliverable is structured as follows. In Chapter 2, specification of community-driven ontology management - a conceptual framework allowing and motivating consensual collaborative ontology construction and reuse for the final user - is provided. The People's portal, which is an implementation of the framework, is described in Chapter 3. In the same chapter, two use cases of the People's portal are presented and analyzed. In the first use case (Digital Enterprise Research Institute), the People's portal was applied as a part of intranet environment of a research institute. In the second use case (KnowledgeWeb on the People's portal), the focus is on acquisition of ontological information about people involved in the area of Semantic Web for facilitation of joint research and social activities. In Chapter 4, a brief survey of related work and evaluation of the proposed approach are provided. Specifically, effectiveness of the performed community-driven ontology management, users' feedback to the People's portal and my personal view on the limitations of the community-driven ontology management are presented. Conclusions and future work - further perspectives for community-driven ontology management in general - are discussed in Chapter 5.

2 Community-Driven Ontology Management

This chapter of the deliverable is organized as follows. At first, I give a motivation for the proposed approach, i.e., community-driven ontology management. Specifically, (1) bottom-up approaches (i.e., grounding for community-driven ontology management), (2) community-driven ontology development and population, (4) community-driven ontology matching, (3) consensus modeling. Finally, (4) an ontological social networking model, which is derived from the theory on community-driven ontology management and is supported in the case studies, is described in the last section.

All subsections of this chapter overlap in the three main topics of this deliverable: **people** (communities, social networks), **ontologies** (ontology construction, ontology management, ontology tools), and **portals** (community portals, end user aspects). To simplify reading of the chapter, the main focus/topic of each chapter's section is marked in Table 2.

topic/section number	1	2	3	4
people			✓	✓
ontologies	✓			
portals		✓		

Table 2: Main Topics in the Sections of Chapter 2

When reading the deliverable, the reader may choose to focus on the sections most interesting and relevant to him/her depending on his/her background. Overall, chapter 2 contains the motivation, problem statement, model and principles of community-driven ontology management within community portals. The statements of chapter 2 are followed in the implementation and use cases of this work (described in chapter 3).

2.1 Bottom-Up Approaches – Why and How They Work

"*If you want to build a ship, don't drum up people together to collect wood*
and don't assign them tasks and work,
but rather teach them to long for the endless immensity of the sea"
Antoine de Saint-Exupery

In this section, success stories in the bottom-up construction of simple schemata and their spread (i.e., acquisition of an ontology status) are introduced, and the bottom-up way towards large-scale ontologies of more potential usage is outlined. We show success of bottom-up ontology development and its limitations on the other hand: the current bottom-up ontology development is not sufficient for establishment of full ontological support in many domains. Development of infrastructures, such as the People's portal, enabling vast amount of users to participate in community-driven ontology management is an important next step in making the Semantic Web applicable for numerous scenarios encountered in real life.

2.1.1 Existing Practical Distributed Ontologies

There are several examples of ontologies that became widely accepted and reused for the purpose of distributed data exchange and integration (see Table 3 for the most populated ontologies on the Web). Very often these ontologies were organically grown and quickly found a large number of creative users, even though for a long time they were not endorsed by any of the popular standards committees. The most common domains of human activities drew many alternative proposals for the specification of a conceptualization of these domains. Two examples of the most often described domains are represented by ontologies describing a *person* and ontologies describing a *document*. Many alternative versions of ontologies describing people and documents are found in online ontology libraries such as Protégé Ontologies Library[9] and SchemaWeb[10]. The reason of high frequency for describing people and documents is likely to be rather simple: the Web is to a large extent driven by people and consists of documents published by them. Below, we provide typical examples of the person and document ontologies that gained a high degree of popularity.

EIGHT BEST POPULATED ONTOLOGIES (GENERATED IN JUNE,2004)

Onto. Name	Namespace URI	# of Docs. Populated
RDF	http://www.w3.org/1999/02/22-rdf-syntax-ns#	$> 1,129,749$
FOAF	http://www.foaf-project.org/	$> 1,126,002$
DC	http://purl.org/dc/elements/1.1/	$> 1,117,433$
RDFS	http://www.w3.org/2000/01/rdf-schema#	$> 1,129,749$
MCVB	http://webns.net/mvcb/	$> 8,838$
RSS	http://purl.org/rss/1.0/	$> 7,560$
vCard	http://www.w3.org/2001/vcard-rdf/3.0#	$> 6,229$
Bio	http://purl.org/vocab/bio/0.1/	$> 6,183$

Table 3: Eight Best Populated Ontologies[11]

[9] Protégé Ontologies Library: http://protege.stanford.edu/ontologies/ontologies.html

[10] SchemaWeb: http://www.schemaweb.info

[11] The table is taken from "How the Semantic Web is Being Used: An Analysis of FOAF Documents" by L. Ding, L. Zhou, T. Finin, A. Joshi, Proceedings of the 38th International Conference on System Sciences, January 2005.

Person ontologies:

1) VCard[12] is a schema to specify electronic business card profile. Factually, vCard is a simple ontology to describe a person with 14 attributes such as Family Name, Given Name, Street Address, Country, etc. The ontology is proposed with the precise way to describe the instance data using RDF, so that the data conforming to this description can be accessed and reused by other applications.

2) FOAF[13] (Friend Of A Friend) is a schema which is similar to VCard in a way that FOAF also is a small ontology to describe a person. FOAF schema provides 12 attribute types, that are similar to the attribute vCard provides: First Name, Last Name, Email address, etc., and the precise way to describe the instance data using RDF is also proposed by the FOAF-project. However, FOAF is more expressive comparing to VCard in a way that it enables to create links between people. I.e., one can express with FOAF that s/he knows (is a friend of) some specific person. Thus, FOAF allows to track connections between people, thus providing more opportunities for practical reuse of ontology instance data. In addition to conventional search and retrieval of the ontology instance data, FOAF provides the means to use personal URIs as data to link people's semantic annotations in a common network. Thus what is of importance is that FOAF is also one of the ways to support cross-metadata referencing on the Semantic Web.

Document/web publication ontologies:

1) Dublin Core stands for a vocabulary aimed to be used to semantically annotate web resources and documents. The vocabulary consists of 15 attributes to describe a document or a web resource and contains parameters that express the primary characteristics of the documents, e.g., Title, Creator, Subject, Description, Language, etc. The vocabulary (ontology) is propagated by Dublin Core Metadata Initiative[14], an organization dedicated to promoting the widespread adoption of interoperable metadata standards and developing specialized metadata vocabularies for describing resources. The goal of promoting a widespread adoption of the standard is claimed to be enabling of more intelligent information discovery systems.

2) RSS is variably used as a name by itself and as an acronym for "RDF Site Summary", "Rich Site Summary", or "Really Simple Syndication". The RSS ontology specifies the model, syntax, and syndication feed format and consists of 4 concepts: "channel", "image", "item", "textinput", each of them having 3-6 attributes like "title", "name", "description". RSS was developed in early 1999 to populate Netscape's My Netscape portal with external newsfeeds ("channels") and thus pioneered syndication; that is, provision of a channel of information by

[12] VCard: http://www.w3.org/TR/vcard-rdf
[13] FOAF: http://www.foaf-project.org
[14] Dublin Core: http://dublincore.org

representing multiple resources in a single document. Since then RSS has taken on a life of its own and now thousands of Web sites use RSS as a "what's new" mechanism to drive traffic their way.

2.1.2 Ontology Promotion and Distribution in Practice

Clearly, after a to-be-standard, i.e., yet another way to describe a certain domain is proposed, this way has to gain recognition from a considerable community to become a really used standard. It is typical, that for a popular domain several parties propose and push forward their schemas describing the domain to other parties to adapt. Thus, coming to an agreement whose schema is to be used as a standard, and who has to adapt is an important issue that requires a solution. Here, we propose a list of criteria for ontologies that contribute to the promotion and distributed character of ontologies and illustrate these criteria with the successfully expanded formats introduced above.

1) Being integrated in successful tools for Semantic Web engineering
RSS, VCARD formats are included in Jena-2. Jena-2 is a mature API for OWL, RDF, DAML+OIL data and ontologies, and is recognized as one of the best existing Semantic Web tools at the moment according to the Semantic Web tool assessment by SemWebCentral[15]. The fact that Jena-2 is an open source environment contributes to the affordability and thus widespread of Jena-2. Clearly, integration with Jena-2 for RSS and VCARD formats leads to a broader dissemination and usage of these formats.

2) Being extended by other ontologies
FOAF ontology is extended by the Relationship[16] ontology that allows to specify the links between people more precisely than the FOAF's "knows". The Relationship ontology specifies a vocabulary for describing relationships between people, containing around 20 terms such as "friendOf", "childOf", "employedBy", "worksWith", "hasMet", etc. Obviously, being extended by a third party is an acqnowledgement of usefullness and appopriateness of the ontology and provides more chances of further reuse and extension.

3) Being integrated in applications and web resources
Serialization of contents using RSS use resources of BBC, CNET News.Com, iTunes, Telegraph (UK), New York Times, Yahoo! News, wired.com are news for general audience (technology, culture, business, politics) and slashdot.org (technology) news, etc. On the other hand, involvement of a predefined ontology in an application or a web resource is also likely to lead to inclusion of this ontology in software toolkits developed to support this resource. And as far as software is also reused, the ontologies encoded in the software have an opportunity to be widely promoted complying with the first criteria for ontology promotion and distribution identified in this list.

4) Being simple

[15] SemWebCentral's Semantic Web tool assessment: http://www.semwebcentral.org/assessment
[16] Relationship ontology: http://purl.org/vocab/relationship

All the described above ontologies (VCard, FOAF, Dublin Core, RSS) are indeed simple, each of them consists of ca. 15 commonly known items in a flat structure. Creators of some of these simple ontologies explicitly stated that lightweight of their ontology in their design goals (e.g., RSS ontology developers put lightweight as the first design goal). Simplicity of an ontology makes understanding of this ontology easy to a human and simplifies its implementation support and reuse. Being simple also contributes to being multipurpose, thus the same ontology can be reused in different contexts.

5) Being based on widely accepted formats
All the described above ontologies (VCard, FOAF, Dublin Core, RSS) have an XML/RDF encoding specification. Since HTML/XML/RDF standards are the main processable formats supported and used on the Web, an ontology proposal has to support these standards for the sake of simplicity of its reusability. Being based on widely accepted standards is also beneficial for interoperability, versioning, mediation support – all third-party developed tools can be reused for the promoted ontologies. Another important issue is extendibility of the XML/RDF standards that caters to the extendibility of the ontologies designed on the basis of these standards and complies with the trends and objectives of the Web.

2.1.3 Simple Ontologies are not Enough

The Need for Extendible Large Scale Ontologies with Distributed Character

The RSS working group states that as RSS continues to be re-purposed, aggregated, and categorized, the need for an enhanced metadata framework grows. Channel- and item-level title and description elements are being overloaded with metadata and HTML. Some producers are even resorting to inserting unofficial ad hoc elements (e.g., <category>, <date>, <author>) in an attempt to augment the sparse metadata facilities of RSS.

The other communities who appreciate usefulness and value of RSS also report that it has reached its limits. There is a demand for more advanced portal syndication which RSS can not satisfy. One initiative in developing technologies to overcome the limitations of simple ontologies for Web publishing comes from Apache Software Foundation and proposes portal syndication with Web services and Cocoon [Ivanov, 2004]. Another initiative is Atom[17] that is aimed to define a feed format for representing and a protocol for editing Web resources such as Weblogs, online journals, Wikis, and similar content. The feed format is to enable syndication, and the editing protocol is to enable agents to interact with resources by nominating a way of using existing Web standards in a pattern. To overcome the limits of externally distributed small-scare ontologies, organization of user-driven ontology extension, support and metadata communication within Web portals is considered in the approach of the People's portal [Zhdanova, 2004].

[17] http://www.atomenabled.org

The reasons why staying within the scope of simple ontologies (e.g., exchanging FOAF profiles and posting cross linked news stories from RSS) is not enough and far too limited for the existing Web are as follows:

- embedding and personalizing rich content and behaviour from remote Web applications are becoming necessity for catering to specific user needs
- extension of simple ontologies, discovery and communication of these extensions are becoming necessity for bringing semantics to a larger amount of Web content
- mapping between simple ontologies and their alignment with other extendible ontologies are becoming necessity for large–scale data integration.

The introduced solutions by the RSS working group to handle the RSS limitations are as follows. One proposed solution is the addition of more simple elements to the RSS core. This direction, while possibly being the simplest in the short run, sacrifices scalability and requires iterative modifications to the core format, adding requested and removing unused functionality. A second solution, and the one adopted in the RSS specification, is the compartmentalization of specific functionality into the pluggable RSS modules. This is one of the approaches used in this specification: modularization is achieved by using XML Namespaces for partitioning vocabularies. Adding and removing RSS functionality is then just a matter of the inclusion of a particular set of modules best suited to the task at hand. No reworking of the RSS core is necessary.

Obviously, the problems and solutions for RSS ontology above are also valid for other simple widely spread ontologies. Having simple and easy to understand ontologies and ontology pluggable extensions on the user side, the complex processes of combination and reuse of these ontology components in ever-changing specification and conceptualization processes of the outside world are left encapsulated on the middleware and application side. Clearly, the development and especially reuse of the pluggable extension modules involve complex problems that are not resolved at the moment. These problems arise from the support requirements for practical large-scale extendible ontology management, such as:

- easy and quick extension opportunity to cater to dynamically arising and changing needs of ontology users
- discovery of existing pluggable extension modules
- composition of existing pluggable extension modules
- decomposition of existing pluggable extension modules
- matching of existing pluggable extension modules and core ontologies with other external ontologies and modules
- tools to support ontology extensions proposed from the user's side, discovery, composition, decomposition, matching and reuse of created earlier ontologies and extensions.

Thus, preserving the successful approach of simple usable ontologies and resolution of the issues above are clearly to be considered as major challenges in the practical state-of-the art distributed ontology management, and are addressed with creating supporting infrastructure for community-driven ontology management.

2.2 Community-Driven Ontology Development and Population

"Either you think, or else others have to think for you and take power from you, pervert and discipline your natural tastes, civilize and sterilize you."
F. Scott Fitzgerald

Here, we describe extension policies for ontologies and editing policies for instance data in community environments providing a ground for consensus making processes in community environments. We identify operations with ontologies at three levels (at the level of an individual user, a community as a whole, and at the portal/community environment level) and distinguish two ontology types: ontologies specifying content or profile data, and ontologies specifying personalization data. These levels and types (shown in Figure 2) serve as a basis for the community-driven ontology management deployment at Web portals and allow introduction of similar editing and storage policies for the ontologies and data that are assigned to the same level and type. The ontology extension and data editing policies that are enacted at the levels of individual users, communities and portals in a consensus framework are as follows.

1. *User profile ontologies*: All portal users extend profile ontologies in a by-the-way, routine manner with no interaction of central controllers and external experts. Bringing in external ontologies and bringing out ontologies constructed within the portal environment are possible.
User profile data are provided and edited by community members, individually.

2. *User personalization ontologies* are extended by any community member who has expertise and capability to support new ontology items with personalization rules or services. Here and below, the user/community <u>has expertise and capability to support ontology items</u> if the user/community can provide functionalities to maintain new ontology extensions and employ them in adding-value scenarios (e.g., search with attributes from new ontology extensions). Such user/community with expertise and capability can be a portal creator(s) or an external service provider. Bringing in external ontologies and bringing out ontologies constructed within the portal environment are possible.
User personalization data are provided and edited by community members, individually.

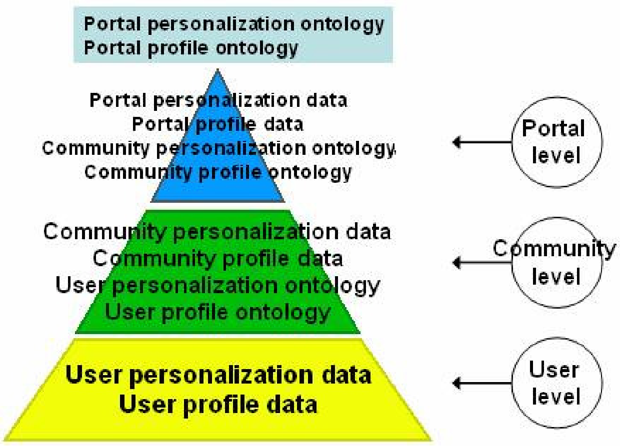

Figure 2: Layering Ontologies and Instance Data

3. *Community profile ontologies* are extended by any community member who has expertise and capability to support ontology items with rules or services. Extension of these ontologies is done on the basis of user profile ontologies.
Community profile data are generated automatically by analyzing user profile data (e.g., a per cent of community members that have their own cars can be obtained as a community profile data item). Also direct introduction of the community data is possible in the cases when this data can not be received as a result of analysis of other ontologies and ontology data (e.g., the name of the community).

4. *Community personalization ontologies* are mainly generated automatically by adapting user personalization ontologies with focus on the requested content and delivery times. Direct introduction of the community personalization ontology items is also possible by anybody who has expertise and capability to support ontology items with personalization rules.
Community personalization data are generated automatically by analyzing user personalization data. Direct introduction of the community data is also possible for the cases when this data can not be received as a result of analysis of other ontologies and ontology data (e.g., the name of the community).
Comparing to the ontologies of the community level, the ontologies of the *portal level* are associated with software used by a community (e.g., different communities can be registered on the same portal, and vice versa, the same community can be distributed among multiple portals), and not explicitly with communities themselves.

5. *Portal profile ontologies* are extended by anybody who has expertise and capability to support ontology items with application integration rules or services.
Portal profile data are provided by a community of users. For example, a member of the community can specify/confirm mappings for certain items of ontologies from the community level. After specification, these mappings are stored as portal profile data and can be reused by other communities. Direct introduction of the portal data is also possible in the cases when this data can not be received as a result of analysis of other ontologies and ontology data (e.g., the name of the portal).

6. *Portal personalization ontologies* are extended by anybody who has expertise and capability to support ontology items with personalization rules or services for application integration.
Portal personalization data are specified by the communities of users, in practice, most primarily portal owners.

2.3 Consensus Modeling

"It's only words... unless they're true."
David Mamet

In this section, we present our model of consensus process in the setting of community-driven ontology evolution on the Semantic Web, clarify how the proposed model complies to the Web content publishing principles, and, finally, illustrate its usage with personalization and community support scenarios.

2.3.1 Definitions

In this subsection, the concepts of ontology, personalization and community support are discussed. Relevance of these concepts to the consensus making problem is shown. Finally, our notion and requirements of consensus on the Web are provided.

Ontologies
Ontology is a *specification of a shared conceptualization* [Gruber, 1993]. "Shared" requires consensus in community employing ontologies as the means of information exchange. Consensus as common understanding and agreement can only be the result of a social process involving individuals and communities. Thus, ontologies have a dual status in information exchange:

- Ontologies as pre-requisite for consensus: Agents can only exchange information when they have already agreed on a common specification reflecting a consensual point of view on the world.
- Ontologies as a result of consensus: Ontologies as consensual models of meaning can only arise in situations where agents agree on a certain model of the world and its interpretation.

Personalization and Community Support
Personalization is traditionally defined as the ability to customize each individual user's experience of electronic content [McCarthy, 2001]. The objective of personalization for the purpose of delivery of personalized information is fairly straightforward. It is to deliver information that is relevant to an individual or a group of individuals in the format and layout specified and in time intervals specified [Won, 2002]. While personalization was applied extensively for individual users (especially in eCommerce area) [Aggarwal

and Yu, 2002; Instone, 2004; Kamei et al., 2003; Schiaffino and Amandi, 2004], the problem of supporting communities with personalization-based information exchange on the Semantic Web context is still open.

By a community support, we understand delivery of certain objects by a community member that are reused or shared by the other community members and thus unite the community. The objects delivered by an individual community member are the basis for information exchange in the community and the information exchanged itself. Thus, these objects may range from portal content to ontology mapping schemas.

Reaching Consensus
In the Webster's dictionary "consensus" is defined as "agreement; accord; consent". In the Semantic Web context, the consensus can be reached at the data level (e.g., "how much should I pay for this service?") and at the metadata level (e.g., "how should we refer to this concept?"). An ability to reach a consensus at both levels is a must for successful cross-application interaction.

The necessity to cope with the following issues makes support of understanding and agreement between two or more parties a difficult task: (i) *dynamicity*, i.e., rapid change of the outside world, its conceptualization and specification of conceptualization, (ii) *heterogeneity*, i.e., presence of various description formats and ontological histories. In addition, the process of reaching a consensus is often combined with the requirement of (iii) *maintaining the integrity of the parties' original ontology bases*. This requirement meets the common need to have an opportunity of access the data via once used schemata and protocols while extending capabilities to adapt to new concepts, facts, rules and processes.

2.3.2 Consensus Process Stepwise

In the light of different ontologies (describing users, communities, cross-platform interoperation), we specify the consensus process basing on actions of individual users and interactions across communities and platforms.
As for ontologies and policies to edit them, we subdivide actions constituting the process of reaching a consensus into the following categories:

- *Individual actions* – actions taken by individual users and having an effect on individual users only;
- *Community actions* - actions taken by individual users and having an effect on more than one individual users;
- *Cross-community actions* - actions taken by individual users and having an effect on more than one individual users belonging to different communities;
- *Cross-platform actions* - actions taken by individual users and having an effect on more than one individual users of different environments (such as portals, platforms, communication media).

Individual actions:

- *Create* – create a new ontology or information item from scratch;
- *Create with reuse* – create a new ontology or information item employing existing ontology or information items. Discovery of and access to a reused ontology item are the necessary conditions of *create with reuse* action. *Create with reuse* also includes a simple reuse, not necessarily accompanied by creation of a considerable added value.

A user joins a community if he/she creates with reuse an ontology or data item basing on an item reused by other (more than one) individual user(s). The strength of connection with a community may be represented in a range from 0 (not reusing any items assigned to the community) to 1 (reusing all items assigned to the community).
Therefore, all individual actions are directed towards weakening or intensifying connections and relations with communities.

Community actions:

- *Join/leave community* – joining or leaving community takes place on the basis of reuse of items created by the community.

Cross-community actions:

- establishing links between communities for gaining benefit for one community from another community and enabling interoperation of these communities.

Cross-platform actions:

- establishing links between portals/platforms to benefit from interoperation of environments and enabling interoperation of these environments.

We model ***consensus*** as a result of a reiterating *process* consisting of the following three steps:

1) *Creation* or *creation with reuse* of an ontology or data item(s) that are estimated as highly relevant by an individual.
2) Discovery of relevance of created or created with reuse items to other individuals The discovery process consists of the following steps:

   a. Ranging communities and individuals as more and less relevant to an individual, e.g., depending on presentation of external ontology items in the individual and community profiles, dynamics and tendency in the evolution of individual and community profiles.

   b. Reception of information on individual and community actions, e.g., as a summary starting from more relevant communities and individuals to less relevant communities and individuals. Reception of information on similar actions (e.g., efforts that can bring benefit via making alignment) and complementing actions (which can influence or be influenced by actions of an individual) is of special importance for estimating relevance.
3) Returning to step (1) with estimation of relevance renewed by a discovery process.

Therefore, consensus is a result of a sequence of individual actions grounded on individual estimates of relevance based on information flows received from the community.

2.3.3 Consensus Making Principles

In the setting of community-driven ontology evolution and actions constituting a social consensus forming process involving individuals and communities, the basic principles of content evolution on the Web should be satisfied. Below we identify core principles of the consensus making on the Semantic Web, which make the proposed framework compliant with the Web.

- New ontology and data items for both content and personalization appear *only because of the efforts of individual community members* who initiate the new items.

- Theoretically (not taking into account restrictions which can be caused by offline societies, e.g., via laws), ontology items introduced by anybody on the Semantic Web *can not be deleted or modified*, they *can be supported or not* by communities. Similarly, any content can be published on the Web, but certain content is read, accessed, used, referred (i.e., supported) by the Web users, and certain content is not. Only introduction of new ontology items is supported in the proposed consensus framework, but not deletion and modification of existing ontology items. After a community member introduces a new item, the item will exist in the system, and the other community members have no possibility to delete and modify the item. Further, the members can support the initiated item by *putting an effort to comply with the initiative* (e.g., by reusing the item and including the item into the personal ontology view) or decline the new item by *ignoring* the item (i.e., not reuse and not put any efforts into the initiative). Thus, we adhere to the principle of backwards consistency in ontology development. The rationale to support the common software development principle of backward consistency in ontology development is to allow the application developers refer to the ontology items which are most appropriate for their tasks (disregarding the fact that these items might not be supported by the majority of the community).

- If two similar ontology items are found in a community environment, *a community member can map the similar items* (for example, for making use of instance data from both items). He/she can indicate the similarity of these items via a mapping pointing that certain ontology items are considered to be related by a certain user. An ontology mapping can be introduced by any user employing a standard functionality that assists to introduce the mapping in compliancy with the community or portal profile ontology. In case a mapping is introduced in a community or portal profile ontologies, the whole community or all portal users can benefit from reusing the mapping.

- *Each ontology item has a measure of importance*, e.g., popularity in the community and relatedness to the community. The value of such measure can be an indicator of how many times an item was instantiated in the community (in association with community personalization ontology). According to the value of the item's importance measure of the community, a decision on how to generally treat the item is executed by personalization rules. For example, a decision on the item's placement on the screen for a default community member can be made via community personalization ontologies and rules as proposed. Apart from the community, the measure of importance of an ontology item can be adapted and applied to an individual user also (e.g., if an user has initiated the item, the item is marked as being important to him/her in association with the user personalization ontology).

2.3.4 Proof of Concept Scenarios

The proposed framework is feasible for facilitating implementation of the consensus-related scenarios on the Semantic community Web portals. We illustrate this feasibility describing certain scenarios and showing the benefits of the involvement of the framework.

Scenario 1: An owner of a new online shop needs to create personalized treatment of customers (e.g., similar to the one created on Amazon.com). The possible ways to satisfy the need are to implement such support completely, or adapt an existing solution. Both ways are costly without a community and user oriented ontological support.

Benefits granted by the consensus framework: Personalization schemata and rules comprise separate ontology-based components and can be applied easily and interchangeably to multiple environments. Therefore, an online shop owner would be able to easily apply personalization solutions specified, agreed upon and used by owners of other online shops.

Scenario 2: A Web user with no ontology engineering experience finds that a community Web portal used by him/her lacks an ontology concept specifying a phone number. The user wants to propose an ontology concept of phone number he/she uses elsewhere to the community associated with the Web portal.

Benefits granted by the consensus framework: Ontology management is assessable to the broadest possible spectrum of community members, and visual ontology representations (web-forms, graphics and natural language descriptions) are the ones viewed in the portal's user interfaces and commonly shared in human-portal interaction. For the regular Web users (non-professional ontology engineers), ontology extension and population are downsized to provision of natural language descriptions, filling out forms and triggering implicit personalization and ontology instantiation (e.g., resulting from observing actual use of the environment such as calculation of item popularity measure). Meanwhile, the ontology structures and mappings introduced at the natural language and user-form level have potential to be reused at the level of machine-machine interoperation.

Scenario 3: A community member wants to be informed about the trends happening in his/her communities and potentially interesting trends happening in other communities. For example, a biologist wants to be notified about published papers, conferences and other activities associated with the concept "protein" in the communities of researchers in chemistry and biology, and he/she wants to know which papers and activities are considered to be important for one or another community.

Benefits granted by the consensus framework: Modeling community decisions takes place at the community level. Specifically, the community members can be timely notified about community trends, e.g., on appearing new concepts or growing or decreasing popularity (i.e., importance) and support of existing concepts. According to these notifications, the members can make decisions on whether to adhere to community trends.

Scenario 4: Information of sensitive character about a community needs to be accessed by an authorized person without accessing profiles of individual users.

Benefits granted by the consensus framework: Creation and evolvement of a depersonalized community member profile encapsulating personal data takes place at the community level. A depersonalized profile of a community member is necessary for acquiring data in community profile ontologies and complies with privacy support. (Privacy guarantee is necessary for obtaining accurate statistical data on sensitive issues, as applying the data on people's preferences and interests diminishes the concerns in providing the data [McCarthy, 2001]. In addition, as indicated by Won [Won, 2002], "…there is in general no cause for concern if information about an individual, even sensitive information, is used merely as a part of broad statistical information (e.g., the number of people in Dallas who purchased a BMW 528i in 2000…)".).

Scenario 5: The owner of a Web application wants to use implicit personalization features to release users of his/her application from an extensive manual data input.

Benefits granted by the consensus framework: Enhancement of implicit personalization is done at a community and portal level. Implicit personalization is an opposite of explicit personalization. Traditionally, implicit personalization is based on user behavior analysis (e.g., products purchased, pages browsed). Normally, users are turned away by explicit personalization such as need to fill in forms, subscribe to mailing lists, etc. [Instone, 2001]. With the proposed framework, implicit personalization can be done on the basis of analysis of ontology data at the community level, being also a base for efficient solutions with respect to the users with underspecified profiles.

In conclusion, I show that the proposed consensus framework process is compliant with the major requirements of the consensus process identified above:

Adaptation to dynamicity – the portal environment allows the users to change ontologies and ontology data as soon as the need for the change appears. The changes take place as soon as they are introduced, and information about the new opportunities is delivered to the associated (and possibly potential) community members.

Adaptation to heterogeneity – interoperation between communities and individuals with different ontological histories is achieved with minimal efforts by reusing once acquired adaptation solutions that were introduced at the community and portal level (e.g., ontology mappings).

Maintenance of the integrity of the parties' original ontology bases – integrity of original ontology bases is supported by allowing to extend ontologies only (without allowing to modify and delete the items once introduced in the environment) and by community and personalization features (e.g., an individual user can delete certain items from his/her view). Generally, an individual does not need to create his/her ontological schemas once he/she has adapted sub-ontologies of a one or more communities or brought in a new ontology schema and established the mappings between his/her schema and an existing community ontology(ies) for interoperation and community support. Personalization profile connects the user's original ontology base to a network of ontologies that are interlinked with shared multiple mapping patterns and thus supported by the community preserving the integrity of the original ontology bases.

2.4 Social Networking Model of a Community

In this section, I describe how the communities and social networks/folksonomies were represented conceptually and numerically, as well as community dynamics notification algorithm employed in the People's portal implementation. On the basis of these representations, more abstract scenarios described in previous sections (such as consensus modeling and information delivery) are practically executed in the People's portal use cases.

2.4.1 Conceptual Modeling

A social networking/folksonomy model employed in the use cases is in agreement and is built on top of Peter Mika's model for semantic social network representation [Mika, 2005]. In this subsection I repeat the major points of Mika's work which are relevant for the model proposed and introduce additional terms where necessary.

As stated by Mika [Mika, 2005], in order to model networks of folksonomies at an abstract level, such model is represented as a tripartite graph with hyperedges. The set of vertices is partitioned into the three (possibly empty) disjoint sets $A = \{a_1, \ldots, a_k\}$, $C = \{c_1, \ldots, c_l\}$, $I = \{i_1, \ldots, i_m\}$ corresponding the set of actors (users), the set of concepts (tags, keywords) and the set of objects annotated (bookmarks, photos etc.) In effect, the common bipartite model of ontologies (concepts and instances) is extended by incorporating actors in the model. Specifically, in the model employed here, I consider *persons* (also more generally called as *subjects*) belonging to the set A, and ontology instances and literal values (also more generally called as *objects*) belonging to the set C. *Subjects* and *objects* are also referred with a common name as *nodes*.

In a social tagging system, users tag objects with concepts, creating ternary associations between the user, the concept and the object. Thus the folksonomy is defined by a set of annotations $T \subseteq A \times C \times I$ [Mika, 2005]. Such a network is most naturally represented as hypergraph with ternary edges, where each edge represents the fact that a given actor associated a certain instance with a certain concept. In particular, we define the representing hypergraph of a folksonomy T as a (simple) tripartite hypergraph $H(T) = \langle V,E \rangle$ where $V = A \cup C \cup I$, $E = \{\{a, c, i\} \mid (a, c, i) \in T\}$. I also refer to edges connecting actors/subjects and concepts/objects as *links*. Factually in the People's portal ontology construction environment, links are most often represented as arbitrary properties that connect subjects with objects.

Tripartite graphs with hyper-edges can be reduced to three bipartite graphs (also called two-mode graphs) with regular edges. These three graphs model the associations between actors and concepts (graph AC), concepts and objects (graph CO) and actors and instances (graph AI). For example, the AC valued bipartite graph is defined as follows:

$AC = \langle A \times C, E_{ac} \rangle$, $E_{\mathrm{ac}} = \{(a, c) \mid \exists i \in I : (a, c, i) \in E\}$, $w : E \to \mathrm{N}$, $\forall e = (a, c) \in E_{ac}$, $w(e) := |\{i : (a, c, i) \in E)\}|$

Therefore, the bipartite graph AC links the persons to the concepts that they have used for tagging at least one object. Each link is weighted by the number of times the person has used that concept as a tag. This kind of graph is known in the social network analysis literature as an affiliation network [Wasserman et al., 1994], linking people to affiliations with weights corresponding to the strength of the affiliation. An affiliation network can be used to generate two simple, weighted graphs (one-mode networks) showing the similarities between actors and events, respectively. Ontology construction on the People's portal-driven community environments is mainly supported at the level of the AC graph presented above, namely involving subjects (actors), links (edges) and objects (concepts).

The process of folding a bipartite graph (the extraction of a one-mode network) can be most easily understood by looking at the matrix form of the graph. Let's denote this matrix as $\mathbf{B} = \{b_{ij}\}$. As discussed before, $b_{ij} = 1$ if actor a_i is affiliated with concept c_j . We define a new matrix $\mathbf{S} = \{s_{ij}\}$, where $s_{ij} = \sum_{x=1}^{k} b_{ix} b_{xj}$. In matrix notation $\mathbf{S} = \mathbf{BB}'$. This matrix, known as the coaffiliation matrix, defines a social network that connects people based on shared affiliations. In our case the links are between people who have used the same concepts with weights showing the number of concepts they have used in common. The dual matrix, $\mathbf{O} = \mathbf{B}'\mathbf{B}$ is a similar graph showing the association of concepts, weighted by the number of people who have used both concepts as tags. Note that in both graphs the diagonal of the corresponding matrices contains the counts of how many concepts or persons a given person or concept was affiliated with in the bipartite graph. In the People's portal, these values are used to normalize the association weights and then retrieving communities based on the relative weights. In case of the **S** social network, for example, this means that the relative importance of links between persons is taken into account.

2.4.2 Numerical Modeling

In this section, I propose a numerical model to specify communities and relations within these communities on the basis of a more general conceptual model described in the previous section.

Connection Strength

Rewording the formalization of the previous section, a community is modeled as follows. Subjects (i.e., persons or actors) can be connected to each other only via links with the same objects (i.e., concepts). This modeling also complies with a definition of a community as a group having common interests. In the model, these interests are represented by objects.

Strictly speaking, direct links between two subjects do not exist. A subject can only be connected to another subject in the following way via an object and two or more links: "Subject1 – Link1 – Object – Link2 – Subject2". A link between a subject and an object are bi-directional. Each direction of a link has a value assigned to it. The value assignment represents the fact that a connection of one subject to an object may be stronger than a connection of another subject to the same object.

Formally, the value of the link is calculated as follows. *link_value*(*link_a*) is defined for any model where *link_a* exists between an object and a subject. The value of the function is in the range $(0,1]$.

Practically, one can determine strength/value of each link by examining subjects and objects associated with this link. Basing on the theoretical principles on language, communication and communities discussed in Chapter 1, I put forwards the following two factors as crucial in influencing the connection strength/value of a link between subjects:

- <u>Popularity of objects</u>: Growing popularity of objects (or how many subjects are linked to these objects) weakens the connection strength between subjects linked via these objects. For example, being connected with someone having an object "Community portals" as a common research topic is stronger than being linked with someone having a common concept "Female" as a "Gender" attribute.
- <u>Capacity of subjects</u>: The more objects are linked to/embraced by a subject, the weaker connections of this subject are to other subjects via these objects. In other words, the more activities a subject is involved in the less attention/time/effort is distributed to the object from the subject's side. For example, if a researcher claims to work in 10 projects, this most often means that the time invested in each of these projects is less than it would be in case when a researcher works in just one project. Here, being involved in many projects with different partners results in weakening the connection strength between partners.

Strictly speaking, modeling connection strength between two subjects can be made more complex with taking in account additional factors and when trying to establish a very precise balance between the two main factors mentioned above. For example, in a system where a person is allowed to marry only one person, being connected to someone via an object "Marriage" is stronger than having the same connection in a community where a person may marry several persons. However, popularity of objects and capacity of subjects are in any case seen as inverse proportional to the connection strength or value of the link.

Remaining generally correct and adding value from the practical point of view, it can be stated the strength of the link between a subject and an object is inversely proportional to the subject's capacity and the object's popularity. Remember that the connection strength function *link_value* is not symmetrical, i.e., subjects can be attached to one another with different strength: one subject may be linked closer to another subject, than the later to the first subject.

The value of the function *link_value* between *subject_1* and *subject_2* from the point of view of *subject_2* is calculated as follows.

$$link\_value(link\_1) = \{ \frac{1}{popularity(object\_1)} \frac{1}{capacity(subject\_1)},$$
$$\forall link\_1 = (object\_1,\ subject\_1) \in E, \exists\, link = (object\_1,\ subject\_2) \in E \ \}$$

Here, *subject capacity* and *object popularity* are metrics signifying on the number of links connected to the node. These metrics are formally specified below. Specifically, the measures *subject_capacity* and *object_popularity* are specified via the measure *links_connected*(*node*), which returns the number of links connected with a node.

Subject Capacity

Informally, *subject capacity* reflects the number of things a person/agent is involved with, the number of activities a person/agent participates in, etc. Subject capacity is identified by the number of objects the subject is connected to.

Formally, *capacity*(*subject_1*) is defined for any model where *subject_1* exists. The value of the function is in the range $[0, \infty)$ and is calculated as follows.

$$capacity(subject\_1) = links\_connected(subject\_1),$$
$$where\ links\_connected(subject\_1) = \{e \mid \forall e = (subject\_1, object) \in E,\ object \in C\}$$

Object Popularity

Informally, *object popularity* reflects the number of persons/agents which are associated via any kind of link with the object. As it was already defined above, objects factually are represented by instances, both string values and resources connected with a subject via a property.

Formally, *popularity*(*object_1*) is defined for any model where *object_1* exists. The value of the function is in the range $[0, \infty)$ and is calculated as follows.

$$popularity(object\_1) = links\_connected(object\_1),$$
$$where\ links\_connected(object\_1) = \{e \mid \forall e = (subject, object\_1) \in E,\ subject \in A\}$$

Closeness Measure

As mentioned above, two subjects can only be connected via an object or objects and other subjects and links, but not directly to each other via a link. Therefore, in order to calculate connection strength between subjects, we look via which objects these subjects are connected and how popular or important these objects are.

As the communities are dynamics and are permanently subject to changes, practically a change in closeness between two subjects is caused by a person profile change on a community portal. When a person assigns new ontology objects to him/her, the closeness measure values towards other people connected with the person change, links to new people may appear and already existing links may disappear.

Formally, *closeness*(*node_1*, *node_2*) is defined for any model where *node_1*, *node_2* and paths between *node_1* and *node_2* exists. The value of the function is in the range $(0, \infty)$ and is calculated as follows.

$$closeness(node\_1, node\_2) = \sum_{paths(node\_1, node\_2)} \prod_{link\_1 \in paths(node\_1, node\_2)} link\_value(link\_1)$$

Here a path between *node_1* and *node_2* is defined as a one or more links when following them one by one, reaching *node_1* from *node_2* is possible. And *link_1* is said to belong to a path when it forms a part of the path between *node_1* and *node_2*. Function *paths*(*node_1*, *node_2*) returns all the paths leading from *node_1* to *node_2* in the given model, or $E_{node\_1node\_2}$ in the graph notation.

As the reader may already notice, the closeness function can be used to calculate closeness between two objects, similar to the way the closeness is calculated between two subjects. Pragmatically, the function reflects how close one subject's view on the world (Weltanschauung) to the view of another subject, i.e., how many common objects they share and how strongly they are committed to these objects. When a subject can be reached via a path consisting of several links, a product of the respective link values is taken. Such modeling correlates with the fact that one direct link is stronger than several transitive links, e.g., being a friend is a stronger relation than being a friend of a friend. When a person/agent can be reached from another person/agent via several paths, the products for every path are summed up in order to receive the value reflecting all the relations connecting two persons/agents. The function is asymmetric in the same way as the function *link_value* is asymmetric, i.e., one node can be connected stronger to another than the later to the first one.

2.4.3 Community Dynamics Notifications

In consistency with the consensus model described in section 2.3 of this deliverable, keeping a community member up-to-date regarding the community dynamics (i.e., changes which take place in the community) is crucial for keeping the community representation correlated with reality and evolving.

Important events for the community members to be informed/notified of include the following.

Notification of a member and a community upon user profile change

A community member is to be notified upon changes in the profiles of community members who are connected to him/her via shared objects.

A notification process for community members on the profile change is as follows.

1) a community member changes his/her profile
2) community members who are notified of the change are identified
3) closeness degree between community members is re-calculated
4) the member who changed the profile and his/her communities are notified about the change, previous and new closeness degrees and changes in closeness degrees (including members indirect links to whom appeared/disappeared because the initiative community member changed relations to certain objects in his/her profile)

Notification of a community/community member upon appearance of a new object

A community member is to be notified upon appearance of new objects in the community space (i.e., ontology), as these objects may appear relevant to a person and a community member(s) may consider assigning them to his/her profile. Selection process of a (sub-)community, which is notified upon appearance of a new object in an ontology, may employ analysis of already existing links between subjects and objects and use of *closeness* value between different nodes. For instance, a community member may be notified about a new object if the closeness value between him/her and the person who introduced this object in the ontology is not smaller than a certain threshold value.

Notification of a community/community member upon popularity change of objects

A community member is to be notified upon popularity change for the objects in the community space (i.e., ontology), as these objects may appear relevant to a person and a community member(s) may consider assigning or removing links to them in his/her profile. Selection process of a (sub-)community, which is notified upon an object popularity change in an ontology, may employ analysis of already existing links between subjects and objects and use of *closeness* value between different nodes. For instance, a community member may be notified about a change in object popularity if the closeness value between him/her and this object in the ontology is not smaller than a certain threshold value.

Depending on a specific use case, different methods to implement these notifications can be pursued. Such methods can include visualization of an information on the web site automatic notification via an email, RSS feed or a mobile device. Specific community-sensible information delivery ways are intended to be consistent with the information delivery principles [Węcel and Zhdanova, 2005].

3 Implementation and Case Studies

In this chapter, the architecture and the semantic formalisms laying in the core of the People's portal and the functionality are explained.

3.1 People's Portal - Framework Implementation

Overall, the prototype of the People's portal is built employing the Web technologies and tools, specifically, Java and JSP as programming languages, existing Semantic Web toolkits for the state-of-the-art ontology management and languages of XML/RDF/OWL family for knowledge representation.

3.1.1 Architecture

In this section I introduce the architecture of the system. In Figure 3, the overall system arrangement and the modules of the People's portal are shown.

The *People's portal implementation* includes modules directly extending conventional ontology management modules with community-oriented functionalities. The modules for community-driven ontology management allow users to develop and instantiate ontologies, access to the user profiles, modify data shared by the communities, reach consensus by reuse and perform further operations of community-driven ontology management.

I provide the system description following the implementation layers as they are depicted in Figure 3.

Platforms and core software toolkits: The core ontology management modules and Semantic Web applications of community-driven ontology management run on the Tomcat server [Tomcat]. The framework is implemented on the basis of Java technology[18] and is applicable to all major existing platforms, such as Windows and Linux.

Data processing and ontology toolkits: In order to skip implementation of core ontology and data management functions such as parsing, querying, storing, a few external toolkits were directly or indirectly reused. Specifically, Xerces[19] was used for operating with XML data. Jena 2 [Carroll et al., 2004] was used as the main toolkit for processing and managing ontological content. In addition, OWL API [Bechhofer et al., 2003] and KAON API [Volz et al., 2003] were reused by employing the Ontology Alignment API implementation [Euzenat, 2004].

[18] Java technology: http://java.sun.com
[19] Xerces Java Parser: http://xerces.apache.org/xerces-j

Ontology and user management: The core ontology management toolkits (such as Jena) do not provide ontology versioning and ontology storage functionalities specific to distributed and often conflicting structures of community portals, and external compatible modules are not available for these purposes. Therefore additional light-weight ontology management functionalities such as semantic change logging, storage update policies, usage statistics calculation were implemented. In addition, introduction of user management logics was necessary, i.e., functions which define user representation and data storage information for individual users. These modules are connected directly with the core ontology management APIs. Partially user management practices were reused from the functionalities inbuilt in Tomcat [Tomcat]. For instance, user registration and restricted access management were supported with the Tomcat functionalities in the DERI use case.

Community and ontology libraries: Community modules capture representation of community models. Specifically, the libraries include functions to calculate communities of an individual member, subject capacities and object popularities, proximity measures between individuals. Also the libraries contain high level functions for community-driven ontology management.

Base libraries: Base libraries module was established at the People's portal in order to perform routine utility functions such as conversion of URIs and community member representations, technical interaction support such as automatic email sending upon community member profile change.

Integration layer: Integration layer includes adapters and wrappers to access case study specific user interfaces, external reused systems and integrate external data. Specifically, the layer includes such specific modules as conversion of the XML data received from the LixTo engine [Baumgartner et al., 2003] to the People's portal RDF-based formats and modules responsible for linking the core system functionality with the user interfaces.

UIs and external APIs: The functionality of the modules is delivered with Web user interfaces and can be accessed by human users through an ordinary Web browser such as Internet Explorer. The interfaces include the ones performing acquisition of ontological content and the ones performing information delivery and visualization for the end users. In addition to being available to the end users, the Semantic Web data generated by the People's portal applications are available for other Web applications. One of the reuse cases of the People's portal data by external applications is YARS [Harth and Decker, 2005][20].

[20] YARS – Yet Another RDF Store: http://sw.deri.org/search

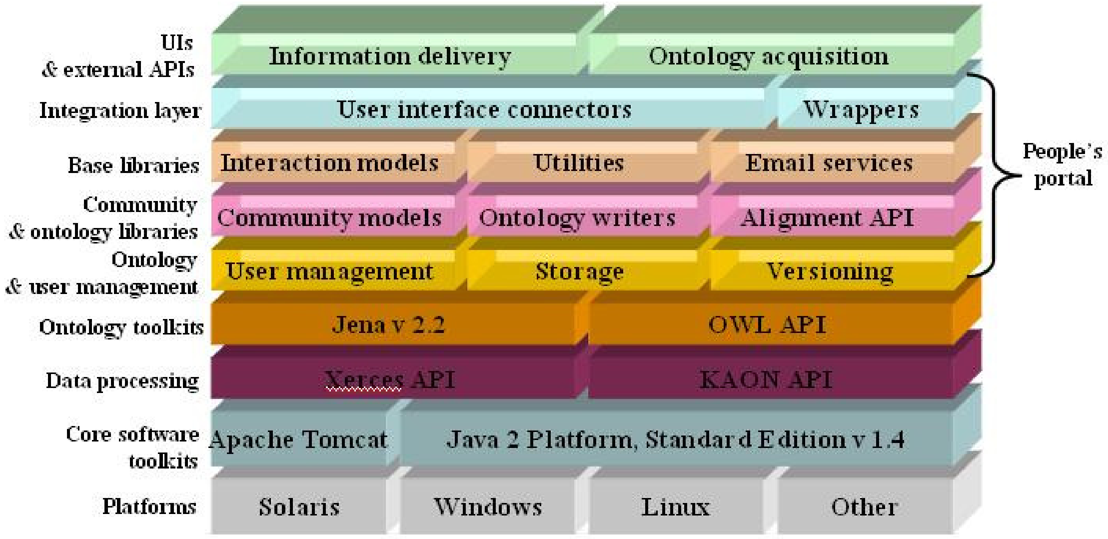

Figure 3: Overall Architecture of the People's Portal

In the People's portal implementation and case studies of the work, components developed at other universities and companies are reused:

- Jena 2 [Carroll et al., 2004] - a framework developed by HP Labs for manipulating with metadata in Java applications. The framework includes APIs and support for ontologies specified in RDFS, DAML, OWL, RDQL, reasoning and persistent storage support; used as the underlying technology for ontology storage, access, manipulation;
- API for Ontology Alignment [Euzenat, 2004] – an proposal for a consensual format capturing ontology alignment; used as a format for ontology alignment, extended and applied to the community-driven ontology alignment;
- The LixTo Wrapper [Baumgartner et al., 2003] – technology for locating and extracting desired data from the Web; used in the Digital Enterprise Research Institute case study implementation to perform initial ontology population from the existing data on the Web.

The code of the project is open source. At the time of writing, everything needed to install, run and modify the system is available at:

http://homepage.deri.org:8080/community/pportal.zip (v0.1)

http://homepage.deri.org:8080/community/pportal_v0.2.zip (v0.2)

The reader is encouraged to experiment with and extend the system.

3.1.2 Functionality

The infrastructure supports acquisition and exploitation of ontological structures by a community. In particular, the following community-driven ontology management functions are supported by our prototype:

- *Editing* – the community members are enabled to extend the domain ontology via graphical Web interfaces, adding classes, subclasses, properties, instances and relating instances. The interfaces are generated dynamically depending on the structure and content of the RDFS or OWL domain and community-supporting ontologies. An example of an employed knowledge acquisition interface is shown in Figure 5.

- *Storage* – community related data are stored by means of a common centralized repository. Private instance data are stored and accessed in a distributed manner.

- *Alignment* – the implementation of the community alignment service[21] allows semi-automatic mapping between ontologies and saving the approved mappings in a publicly available storage, e.g., as an OWL file accessible over the Web. The community ontology alignment service provides a basis for interoperation across communities by linking ontology items used by various parties.

- *Versioning* – the instances are distinguished as community-related and individual-related. Community-related instances are generally relevant to more than one individual at a time and therefore are displayed to many individuals (e.g., information about research projects). Individual-related instances can be restricted to a particular individual (e.g., private phone numbers). Different versioning policies are applied to community-related and individual-related instances. For community instances after instance modification, a new instance is introduced and the visual name of the previous instance is modified. Private individual-related instances are rewritten when changed and a semantic change log is maintained at the community level.

- *Aggregation* – distributed content can be processed by the aggregation module to produce an input to other components, such as the publishing and information delivery component in Figure 3.

In addition, the following principles were taken into account in design and implementation of the community environment:

- *Ontology layer pyramid support* – Ontologies at different layers, such as user, community, portal layers, and different dimensions, such as profile and personalization dimensions, of the community environment are stored and evolved applying different storage and evolution policies [Zhdanova, 2004].

[21] Ontology alignment service: http://align.deri.org

- *Distributivity support* – A core principle of the environment design is to import/integrate ontologies and data from various locations on the Web. Therefore, ontology linking and metadata aggregation are supported.

- *Automatic ontology population* – In order to reduce data input overheads for the members of the community Semantic Web environment, ontologies can be instantiated automatically. In particular, our experience with ontology instance acquisition from HTML Web pages employing the LixTo toolkit [Baumgartner et al., 2003] indicates that acquisition of initial datasets is highly important for getting the users involved in community portal activities. At the same time, the means and sources of automatic information acquisition need to be chosen carefully, e.g., in our case the efforts spent on automatic acquisition of ontology instances for ca. 100 user profiles from existing HTML pages with LixTo were comparable with the efforts required for the manual insertion of the data.

- *Publishing and delivery of Semantic Web content* – The community members are enabled to introduce and see/get notified of the changes that were caused by their actions in the Semantic Web community environment also on the ordinary Web. The role of added-value publishing and delivery of information to the human user is crucial for user involvement in community-driven Semantic Web environments. Earlier practical experiments show that users do not get involved on a permanent basis with an environment that provides solely functionalities for collaborative ontology construction [Farquhar et al., 1997].

- *Support of a social networking numerical model* – As described in the section 2.4 of this deliverable, the conceptual and numerical community and social networking models are supported in the implementation, together with interaction/notification processes. In particular, link values, closeness of subjects, involved communities and other elements of the model are calculated in the People's portal implementation and communicated to the community members.

- *Restricted access and user profiling* – Access policies for the community ontologies are normally required to protect the communities from unauthorized ontology management. At the simplest level, these policies can be supported by user profiling and password protection, as done in our implementation. Further, access policies can be implemented taking into consideration community and social networking information provided by the users [Golbeck et al., 2003].

- *Community-based consensus reaching* – By informing community members about the ontology evolution, the portal infrastructure facilitates the convergence of ontologies, i.e., it helps members to find a consensus in defining the shared ontologies.

3.2 Case Study 1 – Digital Enterprise Research Institute

URL: http://homepage.deri.org

In this section, mission, objectives, community, effort, used software of the Digital Enterprise Research Institute (DERI) case study are described. Major findings/results of this case study are presented in section 4.1 of this deliverable.

Mission

The mission of the *Digital Enterprise Research Institute* case study is acquisition of ontological information associated with DERI from DERI employees for creation of a dynamic intranet environment and facilitation in production of external DERI web pages.

Objectives

The DERI case study has the following research objectives:

- Demonstrate introduction of community-driven ontology management on community portals practically adds theoretically expected values, such as dynamicity and adequate knowledge representation;
- Demonstrate that acquisition of ontological data from regular community members is possible and appropriate, as well as to find out the extent to which the proposed approach is valid and the usage issues it raises.

Community

The case study covered co-workers of DERI[22] and contributed towards transferring DERI into a semantic organization, namely, uniting via a portal a community of people who are of the same "class", e.g., who use the portal for their common (work) activities. Therefore, work in direction of conversion of the DERI website into a Semantic Web portal powered by the People's portal environment took place.

At the time of August, 2004, DERI as a community united by having the same working place had 92 members: 35 members were listed on the DERI Innsbruck website and 57 members were listed at the DERI Galway website. 16 members of DERI Innsbruck and 13 members of DERI Galway had their own homepages and links to them displayed on the DERI "members" page. Thus, *more than two thirds* of DERI members do not have their own homepage on the DERI website. It is appropriate to note that the DERI

[22] DERI: http://www.deri.org

community is growing fast, and includes people with different qualification, such as researchers, managers, technicians. Hence, the firstly adopted methods of the personal homepage and web site support start to fail due to scalability issues (e.g., many people have many activities) and skill issues (e.g., it is not efficient to request a manager to construct his/her own homepage by means of editing an html template as the DERI researchers constructed their homepages in the earlier times of DERI).

Providing a personal Semantic Web DERI homepage for each DERI member and methods to edit it are visible results of the first Semantic DERI prototype. An example of a personal DERI member homepage generated by the portal environment is presented in Figure 4. These homepages are available to be linked from the DERI web-site already now. They also contain a link to an alternative (personally set up) member homepage.

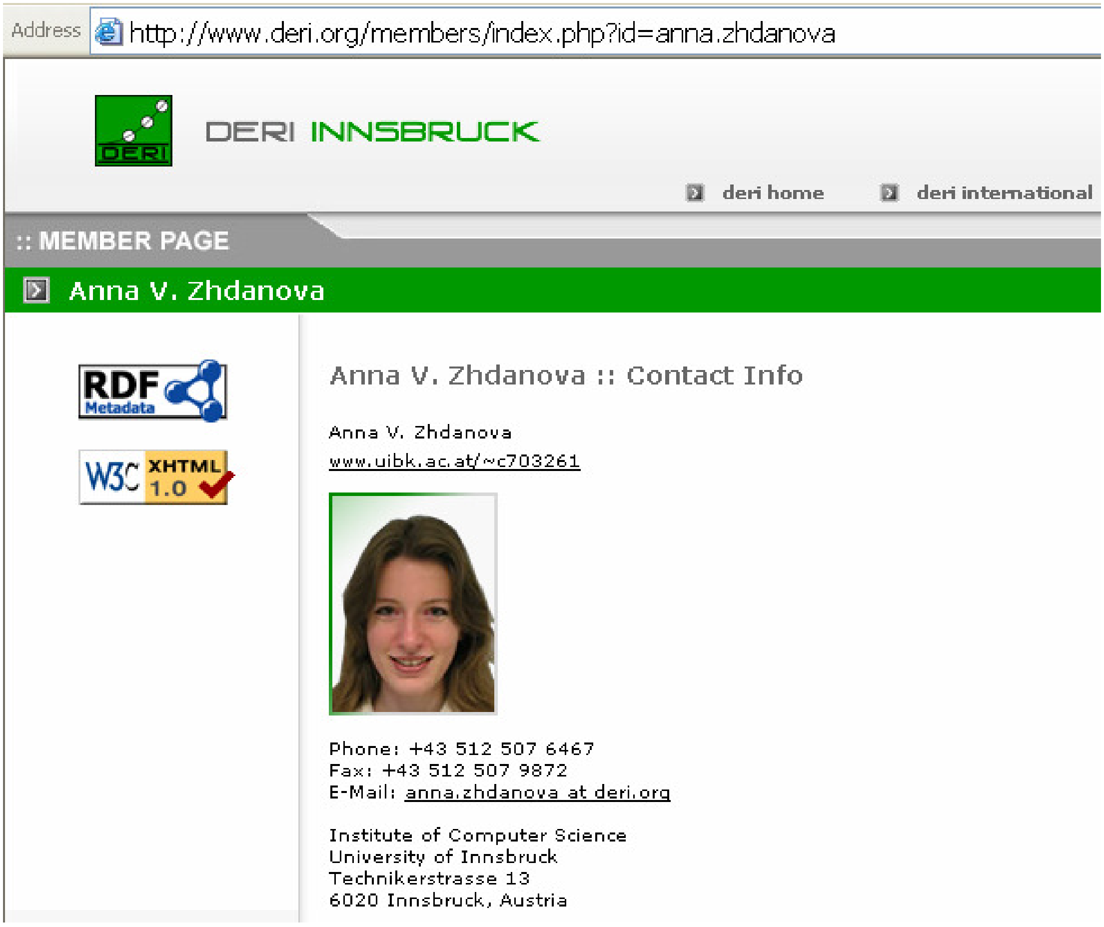

Figure 4: An Example of a Personal DERI Member Homepage

As a DERI member, a user does not need to do anything to have his/her own personal homepage: the semantically enabled DERI Web portal does a substantial amount of publishing work for the DERI member.

The overall benefits for the DERI members and DERI web-site maintainers and visitors are:

- the DERI members gain time, because they do not need to edit the homepage manually (e.g., HTML code);
- the DERI members gain time, because they do not need to learn a special language (e.g., HTML) to edit the homepage;
- the DERI members gain time, because they do not need to adapt the homepage according to the new versions of the homepage templates;
- the DERI webmasters gain time, because they do not need to edit on ontology instance on each page where this instance is present. This is achieved by automatic extraction of this instance via ontology each time the instance appears at the Web site;
- the DERI Web site is more dynamic and the DERI webmasters gain time, because the part of the changes at the ontology instance level are introduced by the portal users 24x7 and immediately published by the portal environment without being presented to and introduced by a webmaster in his/her working hours;
- the DERI website is more community-driven, since the Semantic Web portal environment enables the community members propose and support new ontology items, which can be included at the portal presentation level.

Effort and Used Software

People's portal implementation with all the surrounding components as described in the section 3.1 of this deliverable was employed for enabling the case study.

The case study required analysis of reusable domain-specific ontologies to be taken as a starting point for ontology employed and evolved in the DERI community environment. Also, in addition to the external Semantic Web based pages with visualization of available information (as shown in Figure 4), ontology acquisition interfaces for community members in order to contribute knowledge were designed.

Domain Ontologies

The ontologies involved in the prototype were approached to the ontology specified in the Semantic Web portal working group ontology deliverable [Möller et al., 2004]. Attention to the issue of compatibility with existing wide-spread ontologies (such as FOAF) has been paid. However, direct reuse of the portal ontology specified in the Semantic Web portal ontology deliverable [Möller et al., 2004], has been found difficult because of the following factors:

- not all the immediately required concepts/properties were found in the Semantic Web portal ontology (e.g., property for a work fax number of a person);
- certain modeling solutions in the ontology proved unreasonable at the moment of possible deployment (e.g., the concept "Project" that is a subclass of the concept "Agent" - and by definition acquires its properties. In particular, the concept has acquired a property "GivesTutorials" that is senseless to instantiate in the context of the concept "Project".);
- not all the ontology items were modelled having a widely accepted specifications in their core (e.g., "Publication" concept).

In my opinion, main reasons why these factors came into play are fairly straightforward:

- the ontology at the time of its construction was not meant/considered to be used for publishing/visualizing data from the portal in general, and for publishing the DERI web-site in particular. Thus, the ontology construction work resulted in having many more complex Semantic constructions which were not worth to support with the visualization implementation;
- the ontology development approach was top-down, meanwhile the bottom-up approach (where the specific arising needs are tackled as soon as they arise) would work better.

In the DERI community environment, a support for multiple, smaller sized ontologies (specifically, parts of the FOAF ontology) with a commonly shared knowledgebase was implemented. Therefore, different (also external) agents can access the ontology and the data in the way which is mediation-free for them. The ontology alignment component can contribute to resolution of an adequate access problem in more complex cases.

Therefore, the approach of community-driven ontology construction from the very beginning looked superior to the approach of expert-imposed ontology deployment. Last but not least, the later approach (and its limitations) were already very well known and experienced at multiple research groups [Staab et al., 2000; Corcho et al., 2003; Stollberg et al., 2004a].

Ontology Acquisition Interface

The ontology population (instantiation) part of the prototype is delivered together with a simple web-based ontology editor that allows every portal member to extend the existing ontology. The importance of ontology extension functionalities on the Semantic Web community portals is in allowing the community to specify what kind of content they draw to their portal and in bottom-up growth of the quantity Semantic Web pages without which the Semantic Web is impossible [Zhdanova, 2004].

The idea of having certain real-life actions (e.g., publishing new instances at the portal) taken place immediately should be applied with care to a Semantic Web portal of an organization such as DERI, because the ontology acquisition approach is novel and might invoke undesired consequences if the community makes mistakes in using the system (e.g., due to an obvious lack of knowledge on how to deal with these kind of systems). However, in the Semantic DERI use case, the approach can be used and be helpful irrespectively of the depth of its deployment: by letting the users to extend the existing ontology, we learn more about user's interests and receive additional instance and ontology data that can be included (probably, in some cases, after some transformations) in the next and other stable/publishable ontology and data versions in various applications.

A view on how ontology extension/editing (ontology acquisition) functionality were incorporated in regular user Web forms is presented in Figure 5 and Figure 6. These views are generated directly from OWL and RDFS ontologies and their instance data.

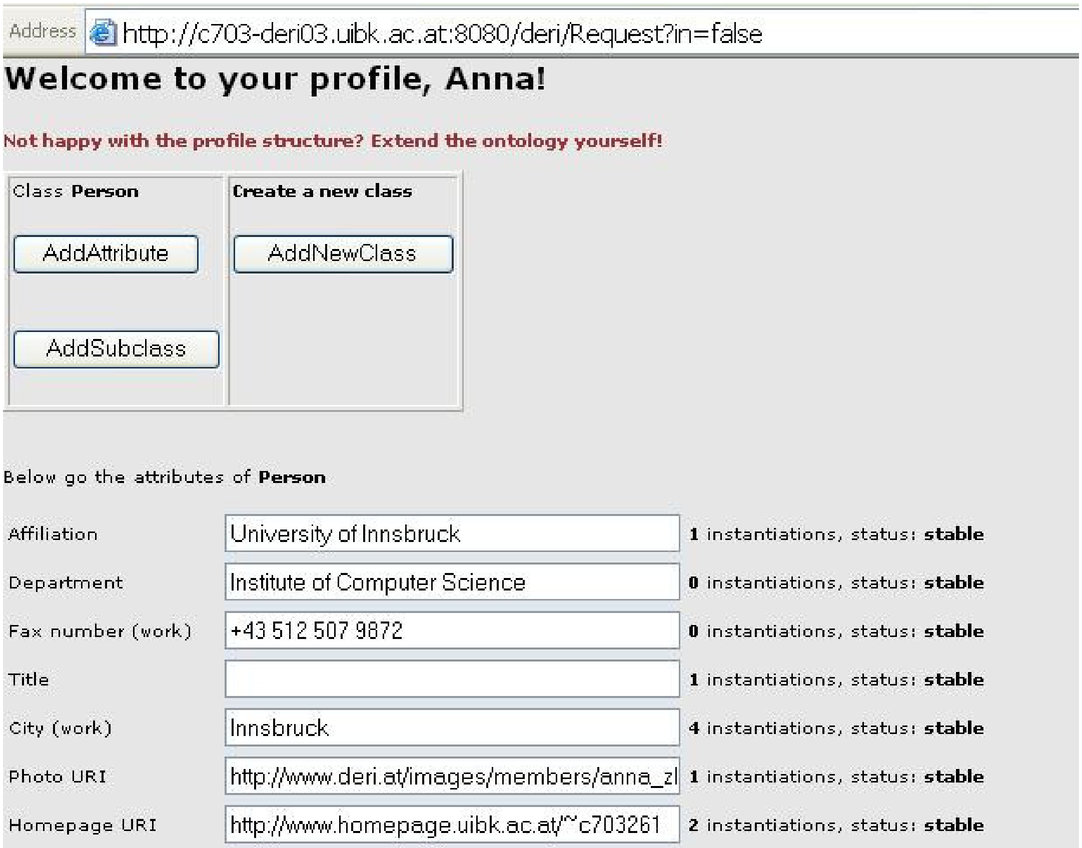

Figure 5: Simple Ontology Editing on Community Portals

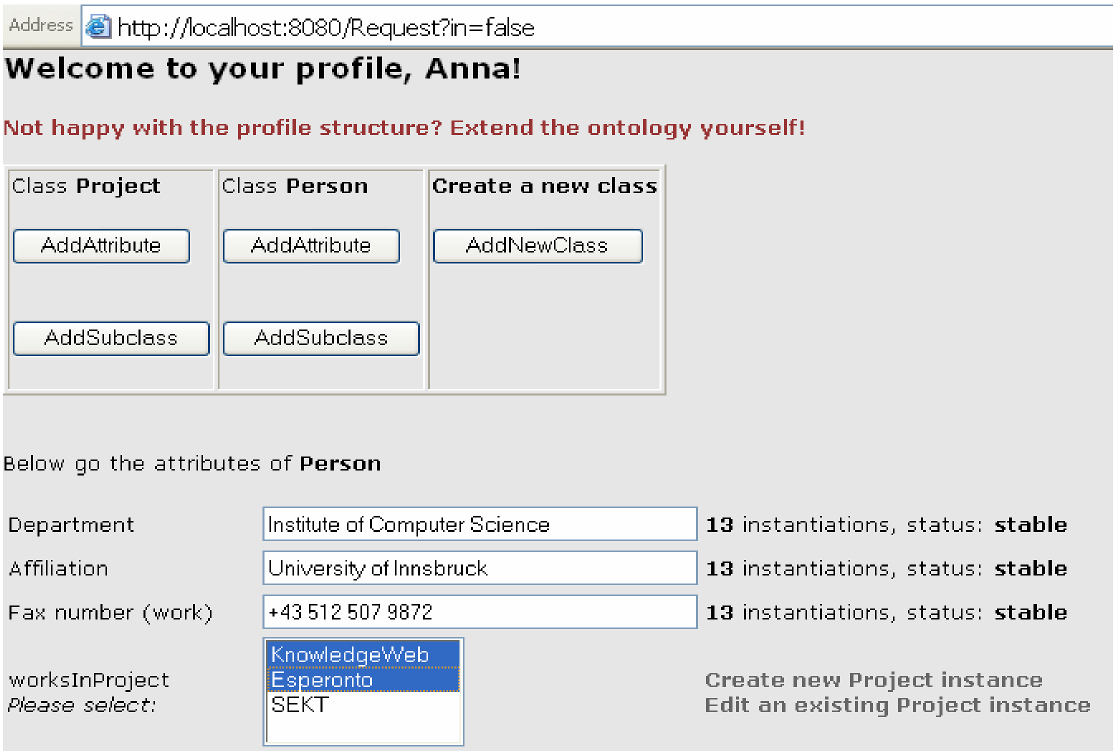

Figure 6: Outlet to Complex Instantiation in a Community Environment

Comparing to Figure 5, in Figure 6, the possibility to relate "class to class" is shown (whereas in Figure 5 only relating "class to literal" is possible) and an outlet to introduction of community instance data (for example, new research projects that can be referred by anyone in the community) is shown.

Here, reusing the community instance data makes it possible to collect and easily reuse more "complex" information. An example of such information at the instance level is data of who works in which project, in which working group, on which topic, etc.

Community-related instances introduced with this environment are versioned and can be reused in consensual ways, depending on the date of creation, date of validity, instance popularity in the community.

When a community member updates his/her profile, a respective RDF annotation of this member is produced and stored at the environment. The community member is also free to copy the annotation elsewhere and reuse it in other scenarios and applications (see Figure 7 for an example of a personal annotation produced by the People's portal).

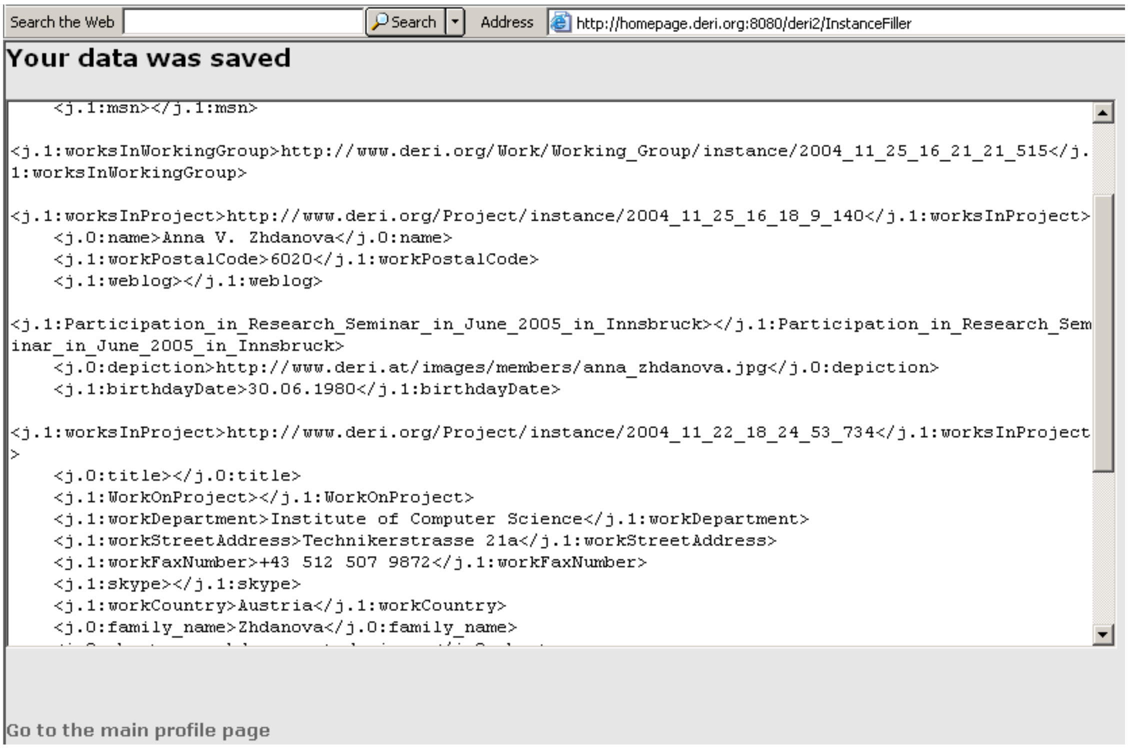

Figure 7: Personal Annotation from the People's Portal

A special gateway was created for DERI web-masters to set up profiles of new DERI members and (if necessary) change existing profile. The gateway has been in action since December 2004 and since then is used by the DERI web-masters every time new DERI members appear (i.e., nearly every month) to introduce basic information about these members.

3.3 Case Study 2 – KnowledgeWeb

URL: http://people.semanticweb.org

In this section, mission, objectives, community, effort, used software of the KnowledgeWeb case study are described. Major findings/results of this case study are presented in section 4.2 of this deliverable.

Mission

The mission of the case study *KnowledgeWeb on the People's portal* is acquisition of ontological information about people involved in the area of Semantic Web for facilitation of joint research and social activities.

Specifically, the application was foreseen to be used by a Semantic Web community for the following purposes:

- business: informing colleagues about current interests and activities, discovering and track communities, finding partners for joint deliverables, papers, event organization, project proposals;
- social and curiosity: getting to know people involved in the Semantic Web area personally and getting introduced;
- exploring the Semantic Web community: finding out how many male vs. female folks are in Semantic Web area, which counties they come from originally, who share interests with whom, etc.

Objectives

The case study *KnowledgeWeb on the People's portal* had the same research objectives as the DERI case study:

- Demonstrate introduction of community-driven ontology management on community portals practically adds theoretically expected values, such as dynamicity and adequate knowledge representation;
- Demonstrate that acquisition of ontological data from regular community members is possible and appropriate, as well as to find out the extent to which the proposed approach is valid and the usage issues it raises.

Community

Comparing to the DERI case study, the case study *KnowledgeWeb on the People's portal* addresses a larger community, factually the whole community that is associated with the Semantic Web research topic.

The core community of the case study was community of researchers involved in the European KnowledgeWeb network of excellence.

The summary of the KnowledgeWeb network of excellence as present in the Annex I of the project is as follows.

“The current World Wide Web (WWW) is, by its function, the syntactic web where structure of the content has been presented while the content itself is inaccessible to computers. The next generation of the Web (the Semantic Web) aims to alleviate such problem and provide specific solutions targeted the concrete problems. The Web resources will be much easier and more readily accessible by both human and computers with the added semantic information in a machine-understandable and machine-processable fashion. It will have much higher impact on eWork and eCommerce as the current version of the web already had. There is, however, still a long way to go transfer the semantic web from an academic adventure into a technology provided by software industry. *Supporting this transition process of Ontology technology from Academia to Industry is the **main and major goal** of **Knowledge Web***. This main goal naturally translates into three main objectives given the nature of such a transformation. (1) Industry requires immediate support in taking up this complex and new technology. Languages and interfaces need to be standardized to reduce the effort and provide scalability to solutions. Methods and use cases need to be provided to convince and to provide guidelines for how to work with this technology. (2) Important support to industry is provided by developing high-class education in the area of semantic web, web services, and Ontologies. (3) Research on Ontologies and the semantic web has not yet reached its goals. New areas such as the combination of semantic web with web services realizing intelligent web services require serious new research efforts. Spoken in a nutshell, it is the mission of Knowledge Web to strengthen the European software industry in one of the most important areas of current computer technology: Semantic web enabled eWork and eCommerce. Naturally, this includes education and research efforts to ensure the *durability* of impact and support of industry."

As Semantic Web is a very dynamic area of research, and KnowledgeWeb is targeted at a large number of participants from different areas, deploying a community portal with functionalities for community-driven ontology management appeared to be a very relevant action in this community setting.

Later, as a matter of fact, "knowledgeweb on the people's portal" became an application for the whole Semantic Web community. As it became visible in the process of application exploitation, the application proved to be more demanded by and useful for Semantic Web community as a whole rather than solely within the KnowledgeWeb network. Presumably, the demand on the community-driven portal was higher in the Semantic Web community in general, because the designed usages of the portal (finding people with similar research interests, getting to know the community, etc.) are more interesting for the people who are relatively new to the community and want to make decisions about getting involved there. In this sense, judging from involvement of community members in an application usage, KnowledgeWeb, running on its second year, looked more as a tight group of old acquaintances (where everyone knows well everybody else) rather than a loose network (where many members are unknown to each other, and there are still many open questions).

Effort and Used Software

People's portal implementation with all the surrounding components as described in the section 3.1 of this deliverable was employed for enabling the case study.

In order to have basic starting data in the application from the very beginning (i.e., prevent multiple data entry and ambiguity), core data on people involved in the Semantic Web area were aggregated and reused from the official KnowledgeWeb NoE portal (http://knowledgeweb.semanticweb.org). The data were in RDF format and therefore, easily processible within the People's portal.

The case study required a number of new human-readable pages and interfaces, as the mission and expected usage of the "knowledgeweb on the people's portal" were different from the expected usage of the DERI community environment. "knowledgeweb on the people's portal" was even more human-centered application than the DERI community environment: the whole portal is built around personal data of the community members and connections such data derive. In Figure 8, the main entry page to the "knowledgeweb on the people's portal" community portal is demonstrated. Similarly to the DERI case study, an ontology acquisition interface was set up (see Figure 9) and specific visualization and information delivery web-interfaces were implemented in order to browse profiles of community members (see Figure 10), and, specifically, browse communities of people connected to a particular person via shared community instances (see Figure 11).

Figure 8: Entrance Page of knowledgeweb on the people's portal

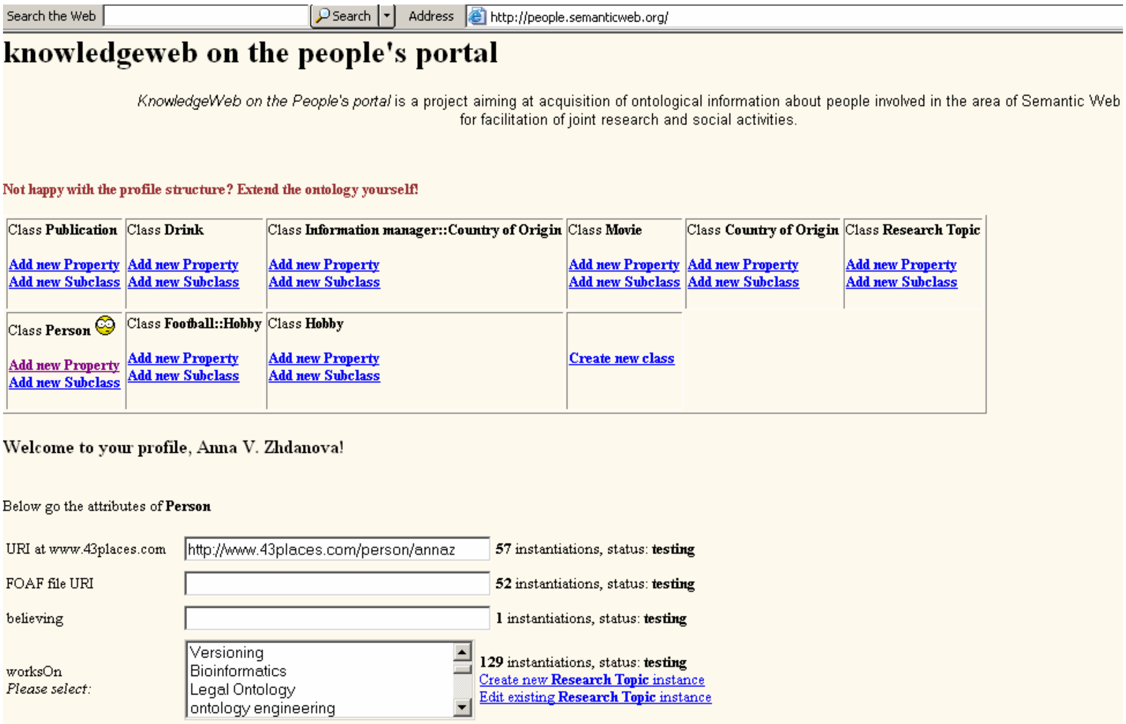

Figure 9: Ontology Acquisition: knowledgeweb on the people's portal

Comparing to the DERI case study, the KnowledgeWeb case study required implementation of more functionalities allowing community members to communicate with each other, as the Semantic Web community has weaker ties than a research institute and additional functionality items are needed for uniting the community.

After KnowledgeWeb case study was launched, it was extensively tried (ca. 450 hits on the first days) as well as generated email feedback, but the ratio of involved and returning users appeared to be relatively low. To tackle the problem of not having a permanently used case study application, "knowledgeweb on the people's portal" was extended with community dynamics analysis features with a goal to get a larger number of community members permanently using it as an entertainment and knowledge exchange platform for social and research activities.

The interactivity extension went into directions of:
(1) enabling people to explore new spaces, e.g., discovering and delivering facts about community which are of potential interest to them,
(2) competitiveness, e.g., setting up a publicly available and regularly updated overview of who is assumed to be getting out the most usefulness of the system.

These directions are supported by means that include displaying related information on the human readable Web pages as well as reaching people beyond Web browsers, such as via email messaging.

Many systems actually practice similar means to involve users (Orkut, LinkedIn, Learner [Chklovsky, 2003]), but (i) none do it in connection with community-driven ontology management or ontology acquisition and (ii) none reveal data on effectiveness of different means to involve users in ontology construction. The last two aspects ensure that the interactivity extension of the "knowledgeweb on the people's portal" use case generates new and valuable research results.

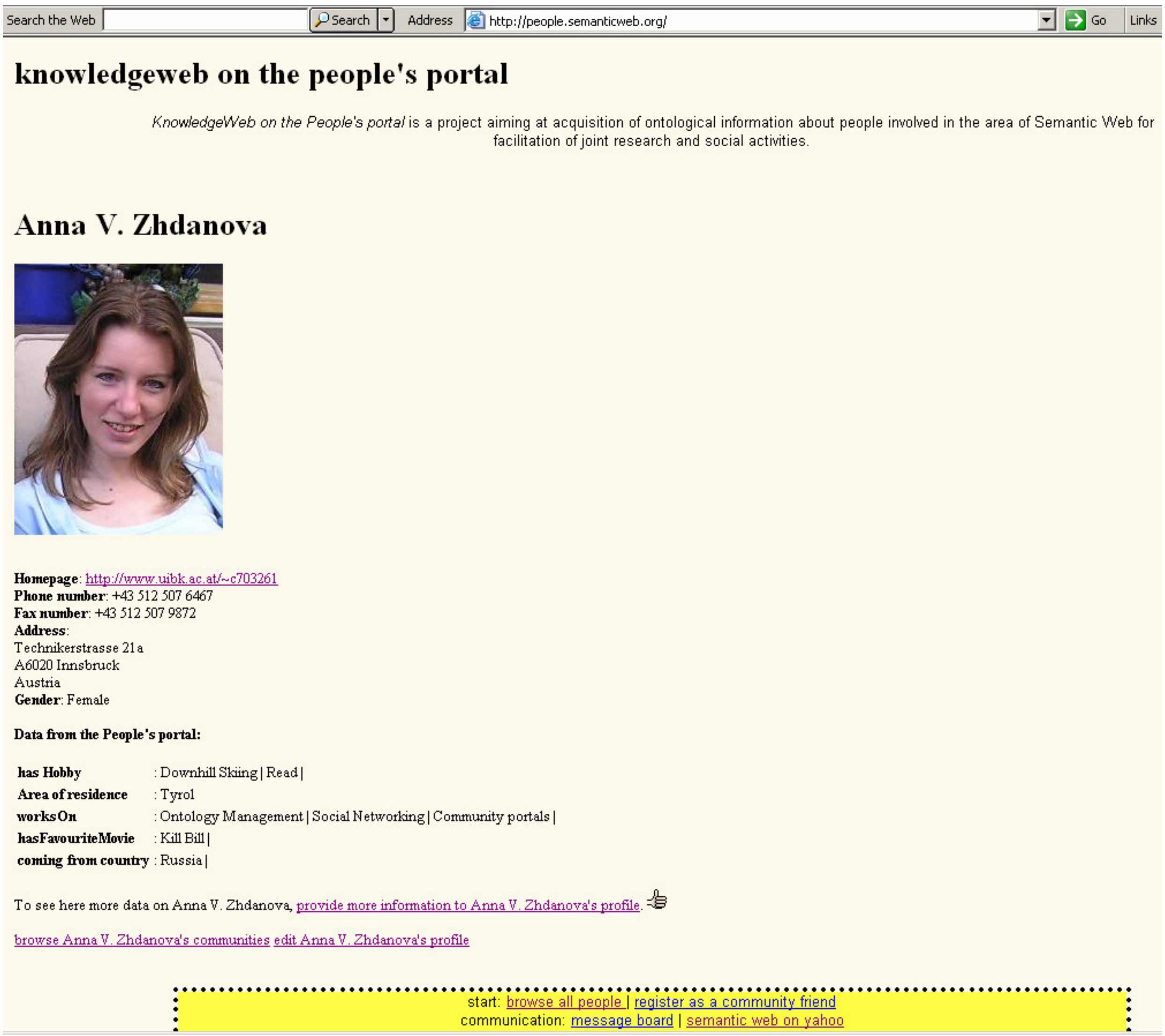

Figure 10: Profile Visualization: knowledgeweb on the people's portal

Figure 11: Community Visualization: knowledgeweb on the people's portal

The community and social networking model employed at the "knowledgeweb on the people's portal" case study was the one described in Chapter 2 of this deliverable. The metrics used to calculate the community information and dynamics are as well, though slightly modified towards simplification: only communities located within one object distance from the subject are taken into account in calculation of the closeness measure. The later measure was introduced to speed up the processing speed of the application: much of the People's portal data are generated on the fly in real time. An example of an email sent to a community member upon her profile change with a community summary is in Figure 12. Similar emails were sent to the closest community members of a person who modified his/her profile in order to keep them up-to-date with the community dynamics.

Notifications on the community dynamics were sent only to community members active in the environment, and within one object distance from the community member who modified his/her profile. Email addressed employed were picked up from the KnowledgeWeb portal, and when possible overruled by the email addresses acquired by the “knowledgeweb on the people’s portal application” itself.

From: PeoplesPortal@deri.org
To: anna.zhdanova@deri.org
Cc:
Subject: profile changes: knowledgeweb on the people's portal

Dear Anna V. Zhdanova!

You or somebody else has changed your profile at "knowledgeweb on the people's portal".

Links from the following people to you may be changed:
Mustafa Jarrar, new connection strength 0.044, previous connection strength 0.018, change degree 0.025
Diana Maynard, new connection strength 0.097, previous connection strength 0.097, change degree 0.0
Fabien Gandon, new connection strength 0.092, previous connection strength 0.072, change degree 0.02
J%F6rg Diederich, new connection strength 0.117, previous connection strength 0.117, change degree 0.0
Dumitru Roman, new connection strength 0.5, previous connection strength 0.5, change degree 0.0
Wolf Siberski, new connection strength 0.08, previous connection strength 0.08, change degree 0.0

Links from the following people to you appeared:
Paolo Bouquet, closeness degree 0.1
Jan Henke, closeness degree 0.04
Robert Tolksdorf, closeness degree 0.054
Aditya Roy, closeness degree 0.073

You are welcome to browse your profile, browse your communities and edit your profile.

Best regards,
People's portal

http://people.semanticweb.org

This email was generated automatically.
You can change your email settings by following the link http://homepage.deri.org:8080/kweb/email-option.jsp?name=Anna V. Zhdanova.

Figure 12: Email Notification from People's Portal upon a Profile Change

4 Results and Related Work

"One must be a god
to be able to tell successes from failures without making a mistake."
Anton Chekhov

In this section, I present (i) evidence of the ontology acquisition power of the People's portal environment, and (ii) present the results of running the People's portal on public web-sites:

- http://homepage.deri.org
- http://people.semanticweb.org

Specifically, the following data are presented:

- The amount and kinds of ontological knowledge acquired (percentage of new classes, new properties, new instances, etc.) and reused, and community-driven ontology management processes performed,
- The user feedback that was gathered as the environment was running,
- My own impressions (with discussion) of further ways to enhance the environment.

4.1 Results for Case Study 1 (DERI)

One of the goals of this work was to show feasibility of application of community-driven ontology management on Semantic Web portals. In order for this effort to succeed, in addition to technical competence in the community-driven ontology management processes, it is desirable to ensure that visitors of the portals have a positive experience. To find out what the contributor perceptions are, they were invited to use and comment on the system deployed as a part of DERI intranet and website.

The DERI community environment is publicly available since October 20, 2004 (first release), and December 12, 2004, (the second version, improved on the basis of users' feedback). During a 70 days trial period (from October 20, 2004 to December, 31 2004) more than 60 feedback messages where expressed from users who tried and played with the environment. The content and frequency of the requests are listed in Table 4.

Summarized it can be stated that most users expressed concerns regarding the Web interface of the environment (which was not the primary goal of the case study), and not the Semantic Web part.

Users were requesting more information on the visual knowledge acquisition interfaces, asking to change visual characteristics of personal homepages, etc. Security and password protection was another issue addressed, as the first release of the environment was open to any Web user. With the second version, the problem was eliminated by securing all the data with a login and password system. Topics related to the Semantic or community-driven ontology management part of the environment (i.e., topics 5, 6, 7 in Table 4) were addressed less than the topics around human-readable Web. The few remarks directly addressing the community-driven ontology management were of two types: some users were confused and others troubled by the opportunity to take more influence than usually in a community environment. Confusions (i.e., lack of knowledge about effects that can be achieved in the environment and how) can be significantly diminished by introducing better user interfaces naturally involving inexperienced users in the ontology management process. The users who are afraid of the potential effects of world wide community-driven ontology management (e.g., semantic spamming or intentional, malicious misuse) should be provided with means to protect themselves against the undesired effects, e.g., by an opportunity to being subjected only to a limited number of communities and services.

Item	Topics of requests	# of requests, version 1	# of requests, version 2	Total # of requests
1	Ontology editor: functionality/visualization	7/1	0/3	11
2	Security, passwords	6	4	10
3	Human-readable Web-pages (looks, future pages)	7	2	9
4	Visibility of Semantic Web content on ordinary Web pages / content negotiation	2/6	0	8
5	URIs, community ontology	4	3	7
6	Considerations regarding usage of the environment	4	3	7
7	“Who is responsible for editing what?”	1	3	4
8	Bug reports	2	1	3
9	E-mail address protection from spam	0	2	2

Table 4: Classification of the Requests to the DERI Community Environment

In Table 5, we summarize the actions of community members during the trial period, such as introduction of new ontology items and instances. Generally, we observed that basic entities like classes, subclasses, properties and instance data were successfully acquired and reused by the community. Examples of correctly acquired ontology items are listed in the second column of Table 5. However, certain propositions made by the community were not obvious to understand, and were considered as modeling mistakes. An example of "incorrectly" modeled ontology part is assigning names of specific projects (such as "SEKT" and "DIP") as property names for a class **Project**, which led also to "incorrect" instance assignment solutions as indicated in the third column of Table 5. However, "incorrect" modeling solutions were normally not supported (i.e., reused) by the community, which demonstrates the community's capability to replace an expert in selection of appropriate modeling solutions, advancing the community-driven ontology management approach.

Types of ontology items	**"Correctly" acquired**	**"Incorrectly" acquired**
Classes	Yes, e.g., "Hobby"	No
Subclasses	Yes, e.g., "Lecture" as a subclass of "Teaching"	Yes, e.g., "Manager" as a subclass of "Project"
Properties	Yes, e.g., "weblog" for a "Person"	Yes, e.g., property "SEKT" for "Project"
Instances	Yes (from initial ontology) – e.g., new instances for "Project", Yes (from community ontology) – e.g., "weblog" had more than 5 correct instantiations	No (from initial ontology), Yes (from community ontology), e.g., value "DIP" for property "SEKT" of a "Project"

Table 5: Community-Driven Ontology Acquisition in the DERI Case Study

In the DERI case study, we have observed all the advantages of community-driven ontology management as discussed in Chapter 2:

1) *Adequate effort investment in ontologies* – The owners of the environment were free from constructing ontologies. All the ontology construction efforts were delegated to the community members. In comparison, conventional construction of an ontology for the same domain took more than half a year in time, involvement of several experts and multiple discussions [Möller et al., 2004], which brought unduly expenses to the working group without a guarantee of an adequate representation of the modeled community as a result.
2) *Comprehensive domain representation* – As ontology construction was delegated to the community members, only ontology items important for active community members were introduced and instantiated in the environment. The ontology which

resulted from a community-driven ontology editing process was substantially different to the ontology constructed by the experts in the area [Möller et al., 2004]. The differences are obvious already at the upper ontology level: the expert-constructed ontology has **Agent**, **Event**, **Location**, **Publication**, **Tool** and **Topic** as the core classes, whereas the core classes of the community-constructed ontology are **Person**, **Project**, **Working Profile**, **Work**, **Teaching**, **Topic** and **Hobby**. These results indicate that experts are not capable to specify the community knowledge comprehensively, as a community would do it itself. For example, here, teaching activities were considered insufficiently important by the experts, and a wish of the research community members share information about their hobbies on the Semantic Web was unexpected and overlooked.

3) *Dynamicity and up-to-dateness* – In the case study, the ontology items and instances were introduced as soon as a community member missed an item or an instance and took an action to introduce it. Whereas with a typical, expert-controlled approach, setting of new items would take significant delay in appearance of the item as well as its extensions and instances. For example, in the case study, a property stating that a **Person** can have a weblog was introduced by a community member soon after the environment's launch, and shortly after this introduction, more than five weblog values were acquired from other community members.

4.2 Results for Case Study 2 (KnowledgeWeb)

Similarly to the DERI case study, the ontology acquisition process and the community behavior at the case study community environment were observed. The statistics and data presented in this section have been acquired over the period from August 18, 2005 till December 5, 2005, starting from the time when "knowledgeweb on the people's portal" went online.

Over the first three months of application exploitation, "knowledgeweb on people's portal" has been used sufficiently intensively to summarize observations and make conclusion on the extent the targeted research objectives were fulfilled: more than 950 unique visitors were recorded to be interacting with the application. A complete monthly Web-site usage statistics can be found in Table 6.

Month	Pageviews	Unique Visitors	First Time Visitors	Returning Visitors	Page Reloads
2005-08	354	27	16	11	327
2005-09	570	83	68	15	487
2005-10	65	22	15	7	43
2005-11	2415	627	594	33	1788
2005-12	475	197	180	17	278

Table 6: Monthly Visitor Statistics for the KnowledgeWeb Case Study

Pursuing observation of the community-driven ontology construction feasibility, I provide a summary of usage analysis for the KnowledgeWeb case study in the same tabular form and classification as for the DERI case study, described in the previous section. The summaries include numbers for different request types sent as a feedback to the application (see Table 7), and acquisition observations for different ontology items classified by their types (see Table 8).

Item	Topics of requests	Total # of requests
1	Ontology editor: functionality/visualization	8
2	Security, passwords	3
3	Human-readable Web-pages (looks, future pages)	6
4	Visibility of Semantic Web content on ordinary Web pages / content negotiation	0
5	URIs, community ontology	2
6	Considerations regarding usage of the environment	6
7	"Who is responsible for editing what?"	1
8	Bug reports	5
9	E-mail address protection from spam	1

Table 7: Classification of the Requests to knowledgeweb on the people's portal

Some of the issues already mentioned in Table 7, became more acute in the "knowledgeweb on the people's portal" comparing to the DERI case study. Specifically, issues on trust and security (present under items 2, 6, 7 in Table 7) were raised substantially more often than in the first case study, see extracts from the emails received from the application users below.

"…the e-mail address is put directly into the page. This is an invitation for crawlers to collect it and use it for SPAM. Please do either omit the e-mail address or use one of the common obfuscation techniques…"

"…in many countries it is not allowed to store data about persons without their explicit permission or when they have not themselves provided those data to you for a stated purpose (e.g. in Belgium, Austria, ...). Your current implementation seems to ignore that."

"…you seem to have collected personal data without asking people for permission (at least, I have never been consented to the publication of my data on the portal)…"

"Very interesting. Photos, telephone numbers, and addresses of women? In America this is unheard of. Are the women outside of the "land of the free" this courageous?"

Such feedback to the case study was not surprising, as the application dealt with private data reuse on a public space, and this issue was spotted to be a sensitive issue on the (Semantic) Web. In particular in 2004, a well-known and popular web-site aggregating and republishing publicly available FOAF files (used to run at www.plink.org) was closed down by its owner, because the later was not willing to stand against the private data reuse pressure. Summarizing, the received feedback confirm that policies (or accepted and agreed upon norms for the reuse of Semantic Web data) gain more and more attention ensured by the growing amounts of publicly available Semantic content that becomes more and more easier to harvest and reuse.

Types of ontology items	**"Correctly" acquired**	**"Incorrectly" acquired**
Classes	Yes, e.g., "Publication"	No
Subclasses	Yes, e.g., "Football" as a subclass of "Hobby"	Yes, e.g., "Information Manager" as a subclass of "Country of Origin"
Properties	Yes, e.g., "Qualification" for a "Person"	No
Instances	Yes (from initial ontology) – e.g., new instances for "Hobby", Yes (from community ontology) – e.g., new instances for "Qualification"	No (from initial ontology), None noticed (from community ontology)

Table 8: Community-Driven Ontology Acquisition in the KnowledgeWeb Case Study

The case study gave an opportunity to gain certain insights on the global Semantic Web community and here I summarize several important ones among them.

As theoretically expected to be, some ontology items were more popular than the others. Typically for an application targeted at a research community, adding instances to a concept "Research topic" concept and assigning them to personal profiles were among the most popular actions in the environment. Particularly, objects related to a “Research Topic” concept were roughly two times more popular than objects related to a concept “Hobby” and roughly three times more popular than objects related to a concept “Movie”.

The following research topics around the Semantic Web area were acquired from and shared within the community:

- Social Networking
- Object Role Modeling
- Versioning
- Bioinformatics
- Legal Ontology
- Ontology Engineering
- Philosophy
- Knowledge Acquisition
- Conceptual Modeling
- Semantic Coordination
- Ontology Alignment
- Regulatory Ontologies
- Multimedia Generation
- Context Aware Computing
- Business Rules
- Mediation
- Peer-to-peer
- Ontology tools
- Database
- Semantic Grid
- Semantic Web Services
- Multimedia Semantics and Reasoning
- Industrial Data Integration
- Community Portals
- Lexical Semantics

In addition to a vast spectrum of research areas the community generated vast spectrums of instance of their hobbies, favorite movies, favorite drinks, counties the communities come from, etc. All these data proves to be diverse. As an example, for such data, I list the countries which were introduced as community instances at the environment in the context of where the community members come originally from:

- Norway
- Northern Ireland
- China
- Palestine
- Italy
- Ireland
- Poland
- Argentina

- Netherlands
- UK
- USA
- Brazil
- Germany
- Cuba
- Belgium
- India
- Spain
- France
- Greece
- Russia

To see the contrast between the counties where people come from originally and the countries wherefrom people are located and access the community application, I list the data showing the domains from which the application was accessed (see Table 9).

Hits	Country
774	Austria
464	Network
359	Italy
346	Germany
221	United Kingdom
218	US Commercial
169	France
114	Netherlands
108	Spain
75	Ireland
57	Poland
54	Finland
53	Brazil
47	Greece
47	Canada
42	US Educational
42	Belgium
40	Non-Profit Organization
37	Switzerland
35	Australia
26	Romania
15	Russian Federation
14	Namibia
11	New Zealand
9	Korea (South)
8	Norway
7	Mexico
7	International
7	Sweden
6	India
6	Hungary
5	Turkey
5	Japan
4	Bulgaria
2	Denmark
2	Cyprus
1	Israel
442	Unknown

Table 9: Visitor Distribution per Country on knowledgeweb on the people's portal

Another specific observation in the KnowledgeWeb case study was on the interactivity mechanisms reporting community dynamics to community members. Factually, the "knowledgeweb on the people's portal" environment draws more interest in the community than it was expected, and communities generate more activity than estimated (especially during the first weeks when the application was "new"). As a result, the automatic notification mechanisms, as they are described in section 3.2 of this deliverable, acted frequently, and in few cases resulted in "overfeeding" members who signed in early in the environment, generating the following email feedback:

"…can you make those messages from the People's Portal stop PLEASE. They are so frequent…"

"I received this week four emails from the people portal. … Do you think that it can be possible to get an update of your profile one time per week or one time every two weeks?.."

Conclusion on this observation is that (i) community behavior is not always obvious to predict; (ii) when designing automatic notification algorithms, one has to take user's views and attitudes into consideration and possibly employ more flexible communication schemes (e.g., using less obtrusive communication technologies such as RSS feeds or more flexible algorithms).

As a general observation regarding potential involvement of regular users in community-driven ontology management, one can estimate that among community members who sign in into the community (or edit their profiles there):

- ca. 100 per cent are able to overcome the barrier of introduction of new community instances other community members can refer to, following to an already existing ontology;
- ca. 20 per cent are able to overcome the barrier of introduction of new ontology schemata (e.g., relatively complex structures such as classes and properties), which other community members can follow.

The ontology acquisition ratio figures roughly correlate with the DERI case study. Generally, the KnowledgeWeb case study demonstrates feasibility of the community-driven ontology management and shows that the applications become richer in knowledge representation, communication opportunities and more dynamical when community-driven ontology management is incorporated in the usage scenarios instead of pre-defined ontologies. The application has also practically confirmed an acute need for community-driven ontology matching in community-driven ontology construction: already as soon as community-constructed ontology acquired ca. 15 properties for a concept "Person", two of these properties were semantically the same ("Country of Origin" and "coming from country").

As a general note I would like to note that running a community environment such as "knolwedgeweb on the people's portal" looks rewarding, and appears to be an interesting research field to be involved in at this point of time. The application clearly drew community interest, which could be seen not only by high visiting rates, but also by the positive email feedback received.

"…very nice portal, and good phd research :-)…"

"I really like to see Semantic Web applications in use…"

"Looks really promising."

"…it's fun to play around."

"…it looks good."

"…looks really cool."

"nice idea, already signed up"

"…great idea, finally a portal where all members of the semantic web community can register and see what the others are doing (additionally to the Semantic Bank of the ISWC05 e.g.) … Thanks for your time and work!"

4.3 Related Work

Much of existing work was referenced in Chapter 2 while discussing the specifics of the proposed methodology. In this section, I list work/approaches which can be compared to the proposed approach as a whole. Further, with examples of typical knowledge-based portals and consensus making tools, I show that the approach proposed in this deliverable renders a higher flexibility to individuals and communities in defining themselves.

4.4.1 Similar Approaches

Ontology development and editing policies are quite simple on most of the current Semantic Web portals [Stollberg et al., 2004]: ordinary portal users do not participate in construction of ontologies, though they often can introduce their ontology instances (e.g., as in KnowledgeWeb[23] and Esperonto[24] Semantic Web portals based on ODESeW [Corcho et al., 2003]). Exceptionally, the users can propose changes to ontology structure, but these changes need to be approved by the main ontology editor [Pinto et al., 2004]. Obviously, this approach to ontology development and editing is not dynamic, does not consider heterogeneity, personalization and community aspects, is not scalable, and thus can not serve as a basis for organization of an effective communication process. Though the People's portal environment supports functions that are typical for Semantic Web portals in general, it is different, because of allowing the portal members to specify knowledge representation issues of their Semantic Web portal, and thus, develop their own portal themselves.

In analogy with FOAF project[25], the People's portal environment provides means (similar to foaf-a-matic) to create semantic annotations on people's personal details or other portal content the portal members might want to bring in. The specifics of the People's portal

[23] KnowledgeWeb portal: http://knowledgeweb.semanticweb.org
[24] Esperonto portal: http://esperonto.net
[25] FOAF project: http://www.foaf-project.org

environment is that its users actually produce machine readable pages to make use of the portal, whereas FOAF project approach focuses on the promotion and improvement of a specific ontology, but not on the FOAF ontology application, usage and dynamic user-driven evolution. Meanwhile, recent research has shown effectiveness of knowledge acquisition from web users, and the same research also brought understanding that in order to be a success knowledge acquisition applications need to move out from the game and toy area and be tightly integrated with applications that are of actual use to the community [Chklovski, 2003].

In comparison to Wiki and Open Directory Project[26] approaches, where "netizens" are encouraged to bring structured knowledge on the Web, the People's portal environment aims at reaching more semantic granularity in specifying the portal content. The People's portal environment provides the means for collaborative development of ontologies. However, it is different from environments for explicit web-based collaborative ontology development [Domingue, 1998] [Farquhar et al., 1997], which resulted to be of limited practical usage. The People's portal environment makes the users involved in creation, extension and reuse of ontologies implicitly in order to increase the value of the portal.

There exist approaches to community information aggregation, visualization and delivery to an end-user on the Semantic Web. For example, Decker and Frank [Decker and Frank, 2004] address this problem by combining the current Semantic Web developments in a Social Semantic Desktop, which will let individuals collaborate at a much finer-grained level as is possible and save time on filtering out marginal information and discovering vital information. Delivery of community-driven Web content will also interoperate at a Semantic level with mobile devices, first projects start to appear, e.g., Semapedia[27]: an application of Web-based Wikipedia to mobile environments. Community-driven ontology management and consensus-based ontology construction will surely benefit by employing special communication techniques developed in these active areas.

Another recent trend comprises very popular portals allowing communities to create their own vocabularies and tag the items/information they want to exchange with arbitrary tags from their vocabularies. The following applications fall in category of such portals:

- http://del.icio.us – This community portal allows communities tag and share their bookmarks, search the bookmarks on the basis of the
- www.43things.com and www.43places.com – These community Web portals allow describing by community-created tags and sharing information about the things people do (www.43things.com) and about the places where people travel or want to travel (www.43places.com).
- www.flickr.com – This community portal allows community members to tag with arbitrary tags, search and share for photos.
- http://base.google.com – This community application was recently launched (in November 2005) and reminds functionality of the People's portal most of all among the portals mentioned here and known to me. The application allows

[26] Open Directory project: http://dmoz.org
[27] The Physical Wikipedia: www.semapedia.org

regular Web user to contribute their arbitrary items (pictures, text, ads, web-sites) for searching and sharing and annotate these items using pairs of an arbitrary attribute and an arbitrary value. Most popular/shared attributes and attribute values come up in the upper level of Google search interfaces and are proposed to be used for searching and browsing the available items.

Though none of the portals above is based on Semantic Web technologies, they clearly show the massive trend of the Web in becoming more structured and annotated in a community-driven manner, via social processes and contributions of regular Web users. In this respect, the People's portal environment appeared to be planned from the very beginning to make a contribution in a trend that now proves to be dominating in acquisition of the structure on the Web. Ontology acquisition from regular users has not yet become a common practice on the Web, but current trends are convincing that sooner or later this will be common practice. Therefore, the People's portal implementation and experiences provide a pioneering insight in one of the most possible futures of one of the most dominating trends on the Web.

5 Conclusions and Future Work

"Everything that has a beginning has an end."
The Oracle, "Matrix Revolutions", 2003

In this chapter, I outline further potential applications of the proposed framework for community-driven ontology management and People's portal and provide conclusions outlining main contributions of the deliverable to the field.

5.1 Further Applications for Community-Driven Ontology Management: Gene Ontology Community

One of the very promising directions for future work is applying the principles for community-driven ontology management to research communities. In particular, life sciences are seen as an important domain of Semantic Web application: for example, recently a charter for "Semantic Web for Health Care and Life Sciences Interest Group (HCLSIG)" was published at W3C[28]. Therefore, here, one more potential use case for community-driven ontology management and the People's portal which I investigate in this section is provision of a consensual ontology construction support to a community associated with the gene ontology (GO) [Con01].

Main goals of the GO community are as follows:
- collect, structure and distribute/disseminate information in the field of genetics;
- create a common vocabulary for talking about major attributes of gene products in order to achieve a "de facto" integration.

The broader goal of OBO [Stevens et al., 2003] is to cover the range of biology which is currently largely in English, and thus facilitate querying and analysis.

The gene ontology is an important example in community-driven ontology construction, because the GO community is far ahead of other communities in *consensus-grounded* and *collaborative* construction of ontologies [Bada et al., 2004]: the ontology size, the dynamics rate and number of people involved in the project make the GO community one of the largest case studies of its kind available.

Bringing in community-driven ontology management to the GO community would be targeted at the **following audiences**:

[28] HCLSIG Charter: http://www.w3.org/2001/sw/hcls/charter.html

- Developers of various community environments (for them to illustrate by example the influence of a community on ontology construction process and the corresponding tool support to make the environments benefit from its communities at the highest degree),
- Developers of tools supporting ontology versioning (to give an idea on which ontology change operations are especially useful and can be successfully captured and processed by the community),
- Computer scientists community (for us to spot gaps in the market with the GO kind of case study).

It is obvious that a community is created to reach certain goals, as for instance the GO community was created. At the same time, the reality demonstrates that once the community Web environment starts to run, the environment is most likely to be used to satisfy goals other than the ones set by the community environment hosts [Shirky, 2003]. Here come "usages" of ontologies and community environments. In fact, "usages" can redefine the "goals". For example, software producers might see an added-value that communities get from their software whereas the product was not designed to provide the discovered added-value functionality.

In Figure 13, we show the **main feedback attitudes** we distinguish in community environments, which are as follows.

- When we consider *individual user level*: "usage", how people use the community environment,
- When we consider *community level*: explicit and implicit feedback, what people say explicitly and which implicit message they bring by interacting with the environment,
- When we consider *community maintainers/software level*: "goals": which goals and purposes the community creators pursue when setting up community software.

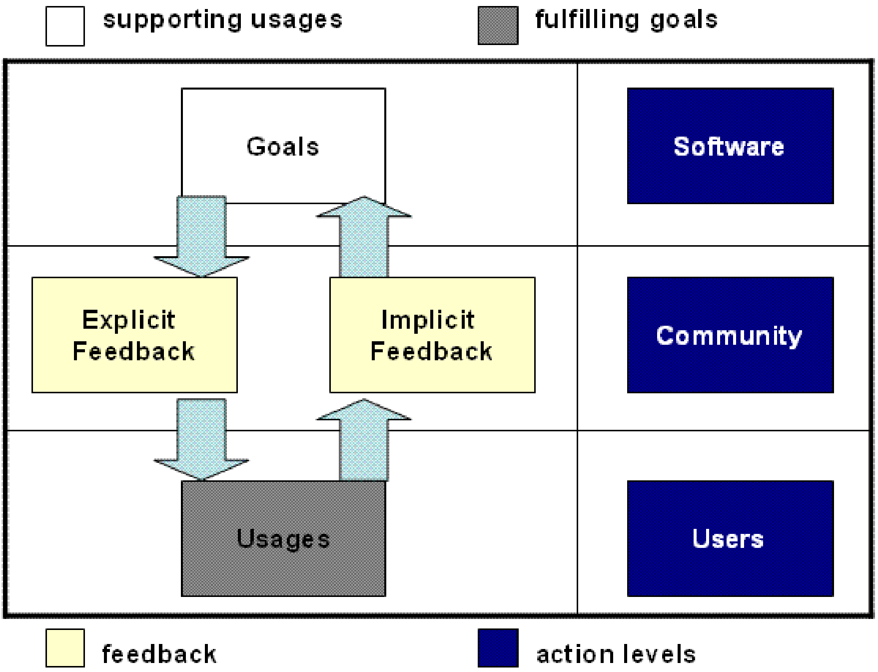

Figure 13: Goals and Usages in Community Environments

Here we analyze the gene ontology community applying the paradigm depicted in Figure 13. Specifically, usages and community software infrastructure of the gene ontology community are described below.

The gene ontology community **reaches its goals** and **performs its usages** employing the following **technical infrastructure**:

To *collect information*: mailing lists, F2F meetings, sourceforge account,
To *structure information*: CVS, sourceforge account, editors such as DAG-Edit, formalisms such as OBOL,
To *distribute/disseminate information*: websites geneontology.org, sourceforge.net, CVS, converters to different ontology languages such as to OWL.

Changes in the gene ontology are listed in monthly reports[29]. The monthly reports contain a concise summary of what has happened in the GO ontologies over the past month: new terms, term name changes, new definitions, term merges and obsoletions, significant term movements, and stats for the ontologies. Information on items from the SourceForge tracker that have been closed over the past month is also available.

Addressing the issues of explicit and implicit feedback in the GO community, one has primarily notice that *active curation* of the gene ontology construction is one of the GO success factors [Bada et al., 2004]. GO construction is moderated by around 40 gene

[29] GO community monthly reports: http://www.geneontology.org/MonthlyReports/

ontology team members. Though involvement of a broad community of ontology users is limited to their provision of suggestions on ontology modification. Such approach to ontology construction can be seen as restrictive in the light of current consensus modeling solutions which provide community members more opportunities to be involved in ontology construction [Zhdanova and Martín-Recuerda, 2005].

Explicit feedback (i.e., what community members request to change) is mainly performed via SourceForge. Specifically, any community member can submit a suggestion on gene ontology modification, e.g., as a “curator request” for issues on the ontology terms. Four categories are offered to choose from when a request is submitted: “new term request”, “other term-related request”, “term obsoletion”, and “none”. Explicit feedback features from sourceforge.org have been available from February 2002 and on March 2002, such souceforge requests start to get resolved by the GO curators.

In Figure 14, we indicate how many explicit “curator requests” to change the gene ontology were proposed by the community (lower line). As for the ontology construction itself, research shows a steady increase in the gene ontology: both in terms added and in relations between these terms. In fact, quantity of relations between terms is shown to grow considerably rapidly than the quantity of terms [Mungall, 2004]. On the graph in Figure 14, we summarize all kind of changes (both in terms and in relations between them) performed at the gene ontology over time (upper line of the graph represent actual changes in the gene ontology over time).

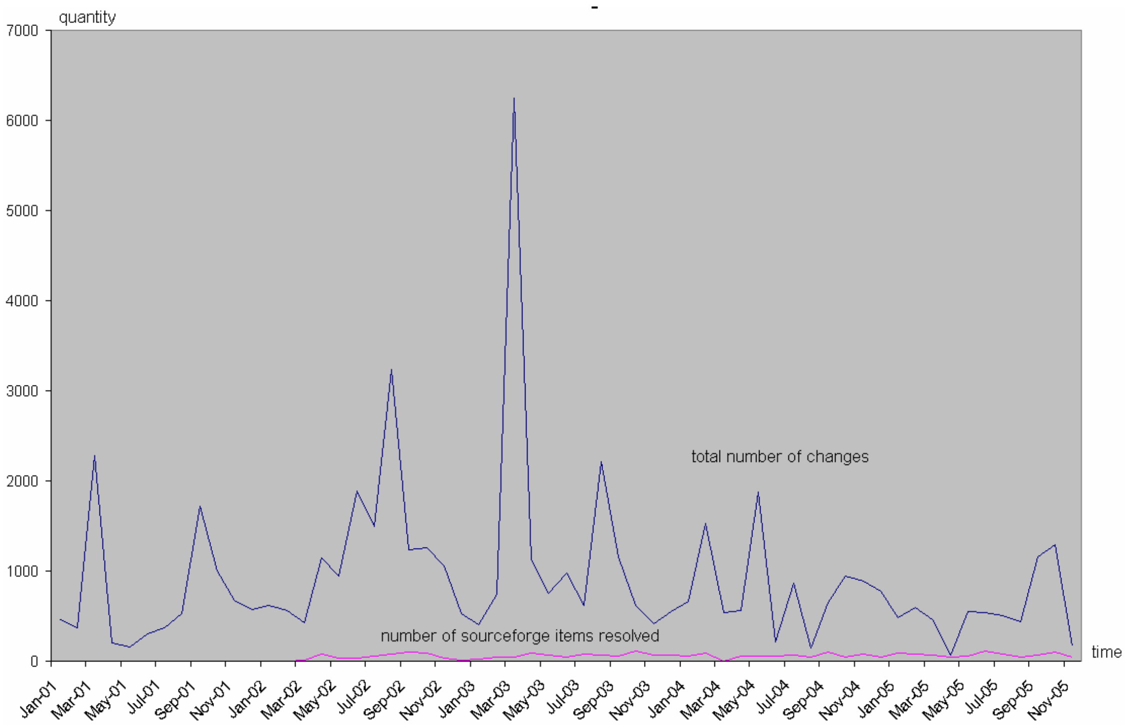

Figure 14: Total Number of Changes and Fixed Requests in the Gene Ontology

Conclusions for data analysis. Analyzing the data on gene ontology dynamics, certain inconsistencies and problems can be seen in the light of community-driven ontology construction. The following issues in GO construction were identified with respect to dynamics and community involvement:

- dynamics of ontology development does not correlate to the development of the actual domain, biology: specifically, at certain points of time substantially more changes are made merely because the curators are more active;
- pre-established categories of ontology change are not equivalently important (e.g., "new terms" are introduced significantly more often than "term merges" take place). Therefore, initial (not user community driven) categorization of the GO construction operations proves to be a rather ad-hoc separation. This example demonstrates that a pre-categorization as well as a predefined by experts ontology cannot be comprehensive;
- implicit feedback (how the gene ontology is actually used) is currently not considered in ontology construction;
- certain relatively old (e.g., dated from 2002) curator requests are still marked as "open", which shows that the communication process in the community can be improved by employment of an infrastructure allowing support of alternative versions and enabling communities to agree on some parts of these ontologies;
- sourceforge requests from the community are far from directing the majority of changes: as one can see from Figure 14, most of the changes done in the GO are still curator/expert-driven.

The proposed in this deliverable consensus making environment prototype is targeted as being able to resolve the above identified bottlenecks in the current community environments.

5.2 Conclusions

The novel contributions and advantages of the principles and features of the proposed community driven ontology management framework include (see also [Zhdanova, 2004; Zhdanova et al., 2004; Zhdanova et al., 2004a; Zhdanova and Keller, 2005; Węcel and Zhdanova, 2005; Zhdanova, 2005; Zhdanova and Fensel, 2005; Zhdanova et al., 2005; Zhdanova, 2005a; Zhdanova and Martín-Recuerda, 2005]):

- integrated approach to community-driven ontology construction covering communities and personalization on the Semantic Web environments.
- introducing and supporting the principle of ontology and data layering, which is beneficial for distribution of data storage, traffic reduction and interoperation of separate components not necessarily native to the environment.
- active involvement of the portal members in construction of their own community portal. Involvement of the members in the portal construction releases the portal

developers from their work of setting up the portal's content and ontological structure. Delegation of the work of setting structure and content makes the applications less costly in support and provides dynamic catering to the members' needs without being brokered by the system administrator.

- the framework is specified in terms of ontologies, i.e., the framework applicable to the Semantic Web.
- the framework is specified and implemented to be deployed at Semantic-based community environments, thus, as an advantage to existing consensus-modeling implementations, costs for connection with specific ontology editors and maintenance of discussions, voting systems, user profiles and ratings are avoided.
- introduction of explicit community ontologies (that can be instantiated both implicitly and explicitly). The community features simplify discovery and search of already existing items (e.g., ontology mappings) and content (e.g., information on flights) for individuals. Another aspect of this organization is that an individual user of the environment is additionally encouraged to adopt ontologies from a community, in order not to build the ontologies themselves, and spend extra time on establishing mappings afterwards. Propagated reuse of ontologized community items, mappings and personalization patterns and decreasing of explicit user involvement in personalization issues are further advantages of explicit community identification leading to consensus.
- the framework does not allow deletion and modification of ontology items, but encourages creation of new items, mapping the items when necessary for interoperation, and keeping the users and communities informed on the evolution and support of the ontology structures in the portal environment.
- a community-driven ontology matching approach that constitutes community-driven ontology management was presented. A prototype supporting the approach was implemented and its usage was analyzed. The results demonstrate feasibility of acquisition and sharing of ontology mappings among the Web communities, thereby supporting, e.g., facilitated knowledge exchange within those communities. Also, by providing a repository of annotated mappings, which is a source of domain specific knowledge, the approach enables other ontology matching systems to produce potentially better results (e.g., a higher recall).

I have proposed the principles and features of a framework for community-driven ontology management that is deployed in a layered and distributed Semantic Web community portal architecture, specifically the People's portal environment [Zhdanova, 2004]. Feasibility of community-driven ontology management is shown: among the active community members ca. 100 per cent are able to overcome the barrier of introduction of new community instances and ca. 20 per cent are able to overcome the barrier of introduction of new ontology schemata. Personalization and community support with effort-driven principles of ontology extension and support, community-driven ontology matching and timely communication of the advances in consensus direction to the community members form a new adding-value approach in solution of ontology construction problems as it can be seen from comparison with the relevant previous work.

Acknowledgements

The author thanks Francisco Martín-Recuerda for useful comments on the deliverable and Graham Hench for advice on gene ontology data statistics visualization.

References

[Aggarwal et al., 2002] Aggarwal, C., Philip, S. Yu, P. S., 2002. An Automated System for Web Portal Personalization. *VLDB 2002*, Hong Kong, China, pp. 1031-1040.

[Bada et al., 2004] Bada, M., Stevens, R., Goble, C., Gil, Y., Ashburner, M., Blake, J., Cherry, M., Harris, M., Lewis, S., 2004. A short study on the success of the Gene Ontology. *J. Web Sem. 1(2)*: 235-240.

[Baumgartner et al., 2003] Baumgartner, R., Ceresna, M., Gottlob, G., Herzog, M., Zigo, V., 2003. Web Information Acquisition with Lixto Suite. In *Proceedings of the 19th International Conference on Data Engineering*, pp. 747-749.

[Baumgartner et al., 2005] Baumgartner, R., Enzi, C., Henze, N., Herrlich, M., Herzog, M., Kriesell, M., Tomaschewski, K., 2005. Semantic Web enabled Information Systems: Personalized Views on Web Data. *International Ubiquitous Web Systems and Intelligence Workshop (UWSI 2005)*, Co-located with ICCSA 2005, Suntec Singapore, 9-12 May 2005.

[Bechhofer et al., 2003] Bechhofer, S., Volz, R., Lord, P., 2003. Cooking the Semantic Web with the OWL API. In *Proceedings of ISWC*, pp. 659–675.

[Berners-Lee, 1998] Berners-Lee, T., 1998. Notation 3 – Ideas about Web Architecture. URL: http://www.w3.org/DesignIssues/Notation3.html

[Berners-Lee, 1999] Berners-Lee, T., 1999. *Weaving the Web*, HarperSanFrancisco.

[Berners-Lee et al., 2001] Berners-Lee, T., Hendler, J., Lassila, O., 2001. *The Semantic Web*. Scientific American 284(5), pp. 34-43.

[Brin and Page, 1998] Brin, S., Page, L., 1998. The anatomy of a large-scale hypertextual Web search engine. In *Proceedings of the seventh international conference on World Wide Web 7*, Elsevier Science Publishers B. V.: Brisbane, Australia. pp. 107-117.

[Carroll et al., 2004] Carroll, J.J., Dickinson, I., Dollin, C., Reynolds, D., Seaborne, A., Wilkinson, K., 2004. Jena: Implementing the Semantic Web Recommendations. In *Proc.*

of the Thirteenth International World Wide Web Conference (WWW 2004), Alternative track papers and posters, pp.74-83.

[Chklovski, 2003] Chklovski, T., 2003. LEARNER: A System for Acquiring Commonsense Knowledge by Analogy. In *Proceedings of Second International Conference on Knowledge Capture (K-CAP 2003)*. October 2003.

[Con01] The Gene Ontology Consortium. Creating the gene ontology resource: design and implementation. Genome Research, 11(8):1425–33, 2001.

[Corcho et al., 2003] Corcho, O., Gomez-Perez, A., Lopez-Cima, A., Lopez-Garcia, V., Suarez-Figueroa, M., 2003. ODESeW. Automatic Generation of Knowledge Portals for Intranets and Extranets. In: Fensel, D. et al. (Eds.), *Proceedings of the Second International Semantic Web Conference*; Springer, LNCS 2870, pp. 802-817.

[Decker and Frank, 2004] Decker, S., Frank, M.R., 2004. The Networked Semantic Desktop. In *Proceedings of the WWW Workshop on Application Design, Development and Implementation Issues in the Semantic Web 2004*.

[Domingue, 1998] Domingue, J., 1998. Tadzebao and WebOnto: Discussing, Browsing, and Editing Ontologies on the Web. *11th Knowledge Acquisition for Knowledge-Based Systems Workshop*, April 18th-23rd. Banff, Canada.

[Euzenat, 2004] Euzenat, J., 2004. An API for Ontology Alignment. In *Proceedings of the International Semantic Web Conference*, Springer, LNCS 3298, pp. 698-712.

[Farquhar et al., 1997] Farquhar, A., Fikes, R., Rice, J., 1997. The Ontolingua Server: Tool for Collaborative Ontology Construction. *Int. J. Human-Computer Studies, 46(6),* pp. 707-728.

[Golbeck et al., 2003] Golbeck, J., Parsia, B., Hendler, J. Trust Networks on the Semantic Web, *Proceedings of Cooperative Intelligent Agents 2003*, Helsinki, Finland.

[Gruber, 1993] Gruber, T. R., 1993. A translation approach to portable ontologies. *Knowledge Acquisition, 5(2)*, pp.199-220.

[Harth and Decker, 2005] Harth, A., Decker, S., 2005. Optimized Index Structures for Querying RDF from the Web. *3rd Latin American Web Congress*, Buenos Aires, Argentina, October 31 - November 2, 2005, pp. 71-80.

[Instone, 2004] Instone, K., 2004. An Information Architecture Perspective on Personalization. In: Karat, C.-M. et al. (Eds.), *Designing Personalized User Experiences in eCommerce*, the Netherlands, Kluwer.

[Ivanov, 2004] Ivanov, I., 2004. *Portal Syndication with Web Services and Cocoon.* 1.0 Technical document.

[Jin et al., 2002] Jin, Y., Xu, S., Decker, S., Wiederhold, G., 2002. Managing Web Sites with OntoWebber, In *Proceedings of the 8th International Conference on Extending Database Technology, EDBT 2002*, Springer LNCS 2287, pp. 766-768.

[Kamei et al., 2003] Kamei, K., Yoshida, S., Kuwabara, K., Akahani, J., Satoh, T., 2003. An Agent Framework for Inter-personal Information Sharing with an RDF-based Repository. In: Fensel, D. et al. (Eds.), *Proceedings of the Second International Semantic Web Conference*; Springer, LNCS 2870, pp. 438-452.

[McCarthy, 2001] McCarthy, J. F., 2001. The Virtual World Gets Physical: Perspectives on Personalization. *IEEE Internet Computing 5(6)*, pp. 48-53.

[Möller et al., 2004] Möller, K., Predoiu, L., Bachlechner, D., 2004. *D1 v0.1 Portal Ontology.* SW-Portal Working Draft 03 August 2004.

[Noy and Musen, 2003] Noy, N., Musen, M., 2003. The PROMPT Suite: Interactive tools for ontology merging and mapping. *International Journal of Human-Computer Studies*, 59(6):983–1024.

[Noy, 2004] N. Noy, N., 2004. Semantic Integration: A survey of ontology-based approaches. *SIGMOD Record*, 33(4):65–70.

[Noy and Musen, 2004] Noy, N.F., Musen, M.A., 2004. Specifying Ontology Views by Traversal. In *Proceedings of the International Semantic Web Conference*, November 2004, Hiroshima, Japan, pp. 713-725.

[Mika, 2005] Mika, P., 2005. Ontologies Are Us: A Unified Model of Social Networks and Semantics. In *Proceedings of the 4th International Semantic Web Conference*, Springer-Verlag, LNCS 3729, pp. 522-536.

[Mungall, 2004] Mungall, C., 2004. Increased complexity of GO, URL: http://www.fruitfly.org/~cjm/obol/doc/go-complexity.html.

[O'Murchu et al., 2004] O'Murchu, I., Breslin, J.G., Decker, S, 2004. Online Social and Business Networking Communities. In Proceedings of ECAI Workshop on Application of Semantic Web Technologies to Web Communities.

[Schiaffino and Amandi, 2004] Schiaffino, S., Amandi, A., 2004. User-interface agent interaction: personalization issues. *Int. J. Human-Computer Studies 60*, pp. 129-148.

[Staab et al., 2000] Staab, S., Angele, J., Decker, S., Erdmann, M., Hotho, A., Maedche, A., Schnurr, H. -P., Studer, R., Sure, Y., 2000. Semantic Community Web Portals. *Computer Networks 33 (2000)*, pp. 473-491.

[Stollberg et al, 2004] Stollberg, M., Lausen, H., Lara, R., Ding, Y., Sung-Kook, H., Fensel. D., 2004. Towards Semantic Web Portals. In *Proceedings of WWW2004 Workshop on Application Design, Development and Implementation Issues in the Semantic Web*, 18 May 2004, New York, USA.

[Shirky, 2003] Shirky, C., 2003. A Group is Its Own Worst Enemy: Social Structure in Social Software. *Keynote talk at the O'Reilly Emerging Technology Conference*, Santa Clara, US, April 24, 2003.

[Stevens et al., 2003] Stevens, R., Wroe, C., Lord, P., Goble, C., 2003. Ontologies in bioinformatics. In Stefan Staab and Rudi Studer, editors, *Handbook on Ontologies in Information Systems*, pp. 635–657.

[Stollberg et al, 2004a] Stollberg, M., Zhdanova, A.V., Fensel, D., 2004. h-TechSight - A Next Generation Knowledge Management Platform. *Journal of Information and Knowledge Management, Vol. 3, No.1*, World Scientific Publishing, pp. 45-66.

[Tomcat] Apache Tomcat: http://jakarta.apache.org/tomcat/index.html.

[Volz et al., 2003] Volz, R., Oberle, D., Staab, S., Motik, B., 2003. KAON SERVER - A Semantic Web Management System. In *Alternate Track Proceedings of the Twelfth International World Wide Web Conference, WWW2003*, Budapest, Hungary, ACM.

[Wasserman et al., 1994] Wasserman, S., Faust, K., Iacobucci, D., Granovetter, M., 1994. *Social Network Analysis: Methods and Applications*. Cambridge University Press.

[Węcel and Zhdanova, 2005] Węcel, K., Zhdanova, A.V., 2005. Information Delivery for the End User of the Semantic Web. In *Proceedings of the ESWC 2005 Workshop on End User Aspects of the Semantic Web*, 29 May 2005, Heraklion, Greece, CEUR Workshop Proceedings, Vol-137, ISSN 1613-0073, pp. 161-175.

[Won, 2002] Won, K., 2002. Personalization: Definition, Status, and Challenges Ahead. *J. of Object Technology (1) 1*, pp. 29-40.

[Zhdanova, 2004] Zhdanova, A.V., 2004. The People's Portal: Ontology Management on Community Portals. In *Proceedings of the 1st Workshop on Friend of a Friend, Social Networking and the Semantic Web (FOAF'2004)*, 1-2 September 2004, Galway, Ireland, pp. 66-74.
URL: http://www.w3.org/2001/sw/Europe/events/foaf-galway/papers/fp/peoples_portal/ .

[Zhdanova et al., 2004] Zhdanova, A.V., de Bruijn, J., Zimmermann, K., Scharffe, F., 2004. Ontology alignment solution v2.0. *EU IST Esperonto project deliverable (D1.4).*

[Zhdanova et al., 2004a] Zhdanova, A.V, Bonifacio, M., Dasiopoulou, S., Dieng-Kuntz, R., Euzenat, J., Laera, L., Manzano-Macho, D., Martín-Recuerda, F., Maynard, D., Ponte, D., Tamma, V., 2004. Specification of Knowledge Acquisition and Modeling of the Process of the Consensus, EU IST KnowledgeWeb NoE deliverable (D2.3.2).

[Zhdanova, 2005] Zhdanova, A.V., 2005. Towards Overcoming Limitations of Community Web Portals: a Classmates' Example. In *Proceedings of the ESWC 2005 Workshop on End User Aspects of the Semantic Web*, 29 May 2005, Heraklion, Greece, CEUR Workshop Proceedings, Vol-137, pp. 111-124.

[Zhdanova, 2005a] Zhdanova, A.V., 2005. Towards Community-Driven Ontology Matching. In *Proceedings of the Third International Conference on Knowledge Capture*, 2-5 October 2005, Banff, Canada, ACM Press, pp. 221-222.

[Zhdanova and Fensel, 2005] Zhdanova, A.V., Fensel, D., 2005. Limitations of Community Web Portals: A Classmates' Case Study. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, 19-22 September 2005, Compiegne, France, IEEE Computer Society Press, pp. 101-104.

[Zhdanova and Keller, 2005] Zhdanova, A.V., Keller, U., 2005. Choosing an Ontology Language. In *Proceedings of the Second World Enformatika Congress*, 25-27 February 2005, Istanbul, Turkey, ISBN 975-98458-3-0, pp. 47-50.

[Zhdanova et al., 2005] Zhdanova, A.V., Krummenacher, R., Henke, J., Fensel, D., 2005. Community-Driven Ontology Management: DERI Case Study. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, 19-22 September 2005, Compiegne, France, IEEE Computer Society Press, pp. 73-79.

[Zhdanova and Martín-Recuerda, 2005] Zhdanova, A.V., Martín-Recuerda, F., 2005. Consensus Making on the Semantic Web: Personalization and Community Support. In *Proceedings of the 6th International Conference on Web Information Systems Engineering*, 20-22 November 2005, New York City, New York; Springer-Verlag, LNCS 3806, pp. 599-600.