Web4WeB e-Collaboration tools

Valentina Janev[1], Jovan Duduković[2], Jelena Jovanović[2],
and Sanja Vraneš[2]

The Mihajlo Pupin Institute,
11000 Belgrade, Serbia and Montenegro
[1]Valentina.Janev@impcs.com
[2]{Jovan.Dudukovic, Jelena.Jovanovic, Sanja.Vranes}@institutepupin.com

Abstract. This paper introduces and compares two innovative tools aimed at enhancing the collaborative work of the researchers and professionals from the West Balkan countries in the Semantic Web field. The first tool, the Alfresco system, is a version controlled content repository designed in REST architectural style and based on a set of document management and collaboration services. The Alfresco workflow engine enables organization and automation of the process activities. The second tool, the Semantic MediaWiki portal, is a semantic collaboration tool and knowledge management system that leverages the latest Semantic Web (SW) technologies and standards. By using semantic annotations, it enriches the Web contents with semantics and makes the knowledge explicit. Thus, it provides structured access to information and accurate and precise knowledge retrieval. This paper discusses the customization of these tools and their utilization in the Web Technologies for West Balkan countries (Web4WeB) project.

Keywords: collaboration tool; Semantic Web technologies; Alfresco; Semantic MediaWiki; assessment.

1. Introduction

The European Union has been financing several projects (e.g, ICT WEB PROMS, WBC-INCO.NET, SCORE) aimed at strengthening the strategic cooperation between the EU and Western Balkan region in the field of information and communication technology (ICT) research and establishing of the European Research Area (ERA) in the Western Balkan Countries (WBC). Because the Semantic Web is one of the fastest developing fields within the ICT sector, while at the same time quite neglected by the West Balkan research body, the EU FP6 funded the Web technologies for West Balkan countries (Web4WeB) project. Within the project, a center of excellence for Semantic Web and other modern Web technologies has been established at the Mihailo Pupin Institute from Belgrade that serves as an information and communications focal point as well as knowledge base for the whole West Balkan region. Herewith, in order to enhance the collaboration of the existing

Valentina Janev, Jovan Duduković, Jelena Jovanović, and Sanja Vraneš

and the new Semantic Web professionals and provide technical support for the Web4WeB network, we propose two collaboration tools: the Alfresco open source Social Computing Platform and Enterprise Content Management (ECM) system and the Semantic MediaWiki open source semantic collaboration tool and knowledge management system (CoKM).

In this paper we will asses the usability of the Alfresco ECM system and the Semantic MediaWiki for enhancing the collaborative work of the researchers and professionals from the West Balkan countries in the Semantic Web field. The paper is organized as follows. First, in Section 2 we discussed about the needs for collaboration in the Semantic Web field, and next, in Section 3 we presented current status of Internet development. Sections 4 and 5 described, respectively, an application of the Alfresco ECM system and an application of the Semantic MediaWiki for the Web4WeB project. At the end, we compared main features of both approaches with relation to the needs of the Web4WeB project.

2. Needs for collaboration in Semantic Web field

The Internet evolution from Web towards Semantic Web consists of several phases commonly known as Web 1.0, Web 2.0, Web 3.0, and beyond [21]. Web 1.0, that lasted from 1990 to 2000, was mainly devoted to establishing a backend of the Web. While Web 1.0 defines the early Internet days with static webpages and passive consumption, Web 2.0 (Social Web) aims to facilitate creativity, collaboration, and sharing among users. The next stage, Web 3.0 (Web of data, Semantic Web) is starting now and it is about representing meanings, connecting knowledge, and putting these to work in ways that make our experience of the Internet more relevant, useful, and enjoyable [5].

In order to support sustainable Semantic Web research development in West Balkan countries and thus better integration in European Research Area, it is necessary to have:

− Research centers that provide an infrastructure and tools for the access, sharing, and distribution of Semantic Web literature,
− discussion forums on relevant open issues in the SW field,
− Semantic Web learning packages in local languages,
− organization of different SW events (workshops, training courses, meetings, etc.).

Adoption of the Semantic Web technologies by the West Balkan research body will raise the awareness and will facilitate a faster take-up of the Semantic Web and other advanced Web technologies and standards, as well as increase the level of competence of all the stakeholders in the region (ICT industry, SMEs, service providers, etc.). Therefore, to speed up the adoption process, we need content management and collaboration tools that will support both the internal activities of the Web4Web network (e.g., preparation of Workshops, collaborative work on internal documents and training

materials), as well as publication and promotion activities. Herewith, in order to enhance the collaboration of the existing and the new Semantic Web professionals, we propose two collaboration tools: the Alfresco[1] open source Social Computing Platform and Enterprise Content Management (ECM) system and the Semantic MediaWiki[2] open source semantic collaboration tool and knowledge management system (CoKM). Both tools have been launched at the same time (e.g., Semantic MediaWiki extensions 0.1 of MediaWiki were released on SourceForge on 29 September 2005; Alfresco launched the Alfresco Product in October 2005) and with the similar goal to serve as foundation for collaborative content management, i.e., "for organizing and facilitating collaborative creation of documents and other content". Alfresco is mainly used for developing in-house ECM systems, while Semantic MediaWiki is widely accepted as a platform for building Semantic Web CoKM applications.

3.   Social Semantic Web

With the Web 2.0, new social computing technologies have appeared that enable users to socialize or interact with each other throughout the World Wide Web thus forming the Social Web. In contrast with Web 1.0, where most of the content was published in the form of static or dynamic pages, today a great part of the content on the Web is created by users in a form of wiki pages, blogs, and folksonomies. The basic features of social networking sites are profiles, friends listings and commenting, messaging, discussion forums, blogging, media uploading and sharing. A general characteristic of the Web 2.0 social computing technologies is that the more collective knowledge systems are used (users collaborate to add contents, semantics, models), they learn and get better. Thus, Web 2.0 has often been referred to as a platform enabling "wisdom of crowds" or "collective intelligence". However, from the Semantic Web perspective, Web 2.0 has some limitations. First, current Web 2.0 tools offer poor query possibilities apart from searching by keywords or tags. Further more, the metadata that is added to the Web content is based on freely chosen keywords (folksonomies) instead of a controlled vocabulary and thus produce ambiguity in the meaning of words or phrases on the Web.

In parallel with the Social Web, Semantic Web technologies are maturing and are finding their way into applications [2],[8]. Semantic technologies include software tools, standards and methodologies that are aimed at providing explicit meaning for the information separately from data and content files, and separately from the application code. Semantic Web technologies provide an abstraction layer above the existing Internet that enables bridging and interconnection of data, content, and processes. This

---

[1] Alfresco ECM, http://www.alfresco.com/.
[2] Semantic MediaWiki, http://semantic-mediawiki.org/wiki/Semantic_MediaWiki.

Valentina Janev, Jovan Duduković, Jelena Jovanović, and Sanja Vraneš

enables machines and not only people to understand, share and reason at execution time. The Semantic Web vision [4] promises a generic infrastructure for interchange, integration and creative reuse of structured data, which can help to overcome the most obvious limitations of the Web 2.0. Yet, in 1998 in the "Semantic Web Road map"[3] document, Tim Berners-Lee foresees transformation of the current Web into "Web of Data", where heterogeneous data sources will be exposed on the Web in a form of "Linked Data". Although a lot was done, the vision of ubiquitous Semantic Web still did not become a reality. The Semantic Web vision is under constant critics of the Web 2.0 professionals arguing that the approach is not applicable in large scale domains and that common standards still do not exist that would enable knowledge and information sharing and interoperation. On the contrary, the Semantic Web people deny any intelligence in the Web 2.0 "collective wisdom".
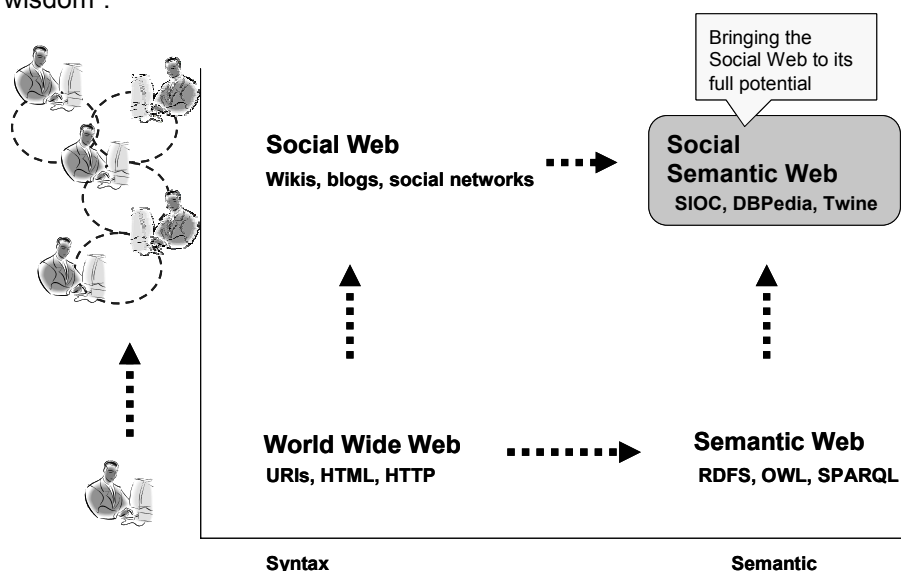

Fig. 1. The Social Semantic Web

In contrast to the above mentioned opinions that treat the Web 2.0 and Semantic Web as competitive visions of the future of the Web, there is growing opinion that the two ideas actually complement each other and that in fact both communities need elements from others to overcome their own drawbacks and limitations (see Fig. 1). Thus, the concept of the Social Semantic Web is proposed [12] where technologies (RDF [7],[15],[17], OWL [11],[18], SPARQL [19]), strategies and methodologies from the Semantic Web and Web 2.0 are combined. The goal of the Social Semantic Web is, on one hand, to use semantic technologies and with the help of standard vocabularies/ontologies (SIOC, FOAF, Dublin Core) interconnect the isolated

---

[3] http://www.w3.org/DesignIssues/Semantic.html

islands of social networking sites such as LinkedIn, MySpace, Friendster, YouTube, Flicker, etc. On the other hand, Semantic Web applications could be enhanced with the wealth of knowledge inherent in user-generated content.

4.   Alfresco enterprise social software

In order to choose the most suitable collaboration and content management system (CCMS) that will provide information storage, sharing and exchange infrastructure for the researchers and practitioners from the Western Balkan countries, we consulted the CMS Matrix[4]. The CMS Matrix rates hundreds of open source content management systems based on roughly one hundred criteria. Although this tool, in our view, is not completely reliable for actually selecting a CCMS, it is useful to narrow down the available options to few that could be evaluated in depth. Thus, we compared four open source Java based CCMS: OpenCMS[5], Apache Lenya[6], Magnolia[7] and Alfresco [25], and have selected Alfresco ECM due to the following reasons:

−   Apart from content management functions, Alfresco supports groupware / collaboration and thus offers additional channels (discussion / forums, chat, mail forum, file distribution) for expertise exchange.
−   Alfresco has built in Events calendar that will help all users to be informed on time about all important events, training courses, and workshops.
−   Alfresco integrates the Apache Lucene search engine which is the leading open source full-text search engine. Alfresco has extended Lucene to not only understand the text within a content object, but also its metadata and categories, and allows several repositories to be searched simultaneously.
−   Alfresco supports many of the latest standards including the JSR-170 file access API, the JSF tag based interface, the Spring framework, JSR-168 portlets and WebDAV file transfers.
−   Alfresco is one of the leading open source enterprise content management system developed by ECM professionals (former employees of Documentum, the leading commercial DMS).

In the following subsection, we give a detail description of the system architecture and the process of design, customization and usage of the Alfresco ECM for the Web4WeB project purposes.

---

[4] www.cmsmatrix.org.
[5] http://www.opencms.org/en/.
[6] http://lenya.apache.org/.
[7] http://www.magnolia.info.

4.1. The system architecture

The proposed Web4WeB e-collaboration solution system architecture that is already installed at the Mihajlo Pupin Institute is depicted in Fig. 2. It aims to serve the SW researchers from the West Balkan countries, as well as the collaborators from the EU countries that will mainly contribute written documents or provide consultancy in the forum discussions.

Alfresco ECM ver. 2.9 was installed on the Windows OS platform (Intel Core 2 Duo, CPU E4500 @ 2.20GHz, 3,25GB of RAM, 240 GB disc space). It runs on the Apache Tomcat application server with the MySQL open source database as a backend for storing the metadata. The metadata describes the components of the Web4WeB e-collaboration solution including spaces, documents, rules, services, while the content itself is stored as a binary object on the file system.
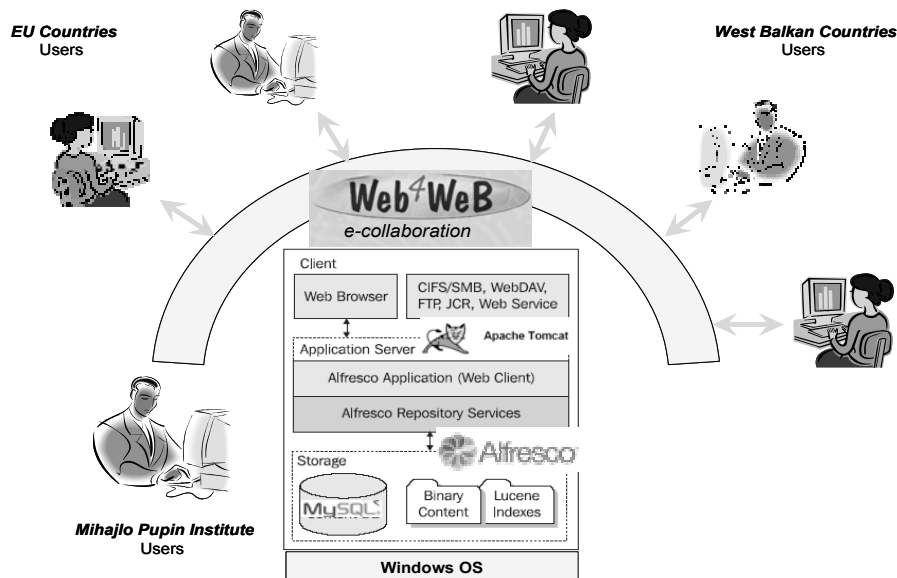
Fig. 2. The Alfresco system architecture

The Alfresco system has been organized into service layers to take maximum advantage of the used Aspect-Oriented programming approach. The Alfresco system uses the Spring Framework as the foundation of its Aspect-Oriented approach. Aspect-Oriented Programming (AOP) was invented in the mid-1990s at Xerox PARC[8], the same place where Object-Oriented Programming (OOP) was invented. Aspects modularized behaviors and made reuse across different classes (applications) much easier.

Alfresco ECM is built around a content repository that is a set of services used to import, classify, store, search, access and control content. Alfresco

---

[8] http://www.parc.com/.

provides SOAP and, with the latest versions, REST-based services as well. SOAP (*Simple Object Access Protocol* [6]) is a lightweight protocol for exchange of information in a decentralized, distributed environment recommended in 2000 by W3C. REST (*Representational State Transfer* [10]) is not a standard, it is an architecture style of networked systems. An important concept in REST is the existence of resources (sources of specific information), each of which is referenced with a global identifier (e.g., a URI in HTTP). In order to manipulate these resources, *components* of the network (clients and servers) communicate via a standardized interface (e.g., HTTP) and exchange *representations* of these resources (the actual documents conveying the information).

In addition, Alfresco integrates the WebDAV server and Apache Lucene[9] full text indexing and searching service. Lucene is a text mining service that offers a wide range of the keyword based searching options. With Lucene, users can explore the document's contents as well as the automatically extracted and user defined metadata. Lucene indexes are stored as folders on the file system.
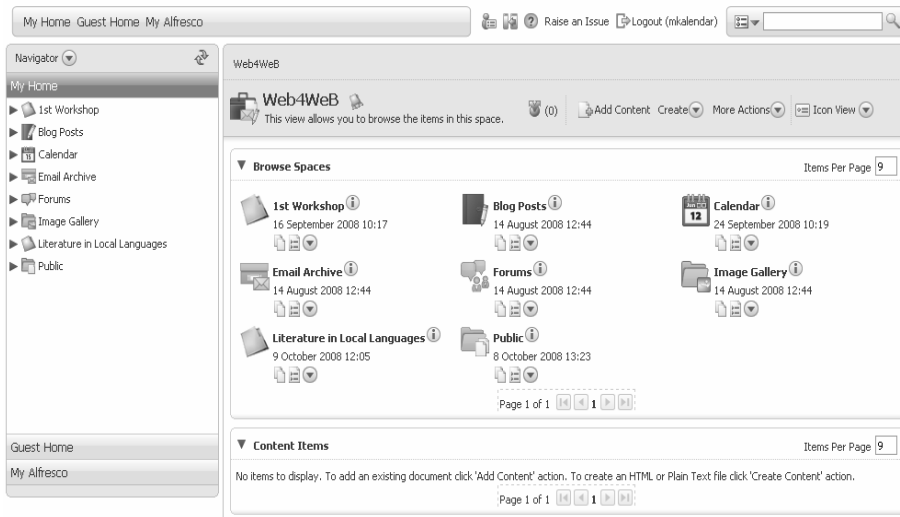

Fig. 3. Web4WeB project user interface

---

[9] http://lucene.apache.org/.

Valentina Janev, Jovan Duduković, Jelena Jovanović, and Sanja Vraneš

4.2.   Customizing Alfresco for the Web4WeB project

As is explained in Table 1, the Alfresco customization consists of defining user spaces, users, user groups, user defined aspects and rules, as well as configuration of the graphical user interfaces (see Fig. 3).

Table 1. Customizing Alfresco for the Web4WeB project

Step 1:	Administrating team spaces
	Collaboration between users in the Alfresco system is going through the special smart folders named spaces. A space is defined by a security policy (a set of actions that can be performed by a user or a group of users on that space) and business rules (a set of actions that is applied on content in that space). Main spaces within the Web4Web project are: the *Web4WeB* space, that is accessible to everyone in the Web4WeB network and four *Translation* spaces aimed to facilitate the process of translation of e-learning materials into languages of nationalities from the West Balkan area.
Step 2:	Administrating users and user groups
	Once all spaces were defined, the next step was to define users, to organize users in groups and to define their roles. Currently the Web4WeB network includes over twenty users organized into the following groups: Web4WeB_Tutor, Web4WeB_Presenters, Web4WeB_Contributors, IMP_Administration, and IMP.
Step 3:	Defining categories
	The Alfresco predefined categories *Tags* and *Languages* were extended with subcategories that are necessary for tagging the Web4WeB documents. Thus, the *Tags* category contains subcategories such as *Semantic Web, Social Web, Query languages, Applications, W3C standards* that are meaningfully used to describe the uploaded contents. In addition to predefined languages in the *Languages* category, local languages that are used in West Balkan Area were added: *Serbian*, *Croatian*, *Macedonian* and *Albanian*.
Step 4:	Defining e-mail templates
	A way to inform participants of the Web4WeB network about a new event or content in the system is via sending e-mail as a notification or an invitation. Alfresco offers possibilities to define different e-mail templates that could be used ad hoc in different occasions or to be used in rules that are linked to spaces.
Step 5:	Defining rules

	In Web4WeB e-collaboration solution business rules are used: to add an aspect that extends the document properties with Dublin Core metadata, to trigger sending e-mail to a specific group of users, to add additional aspects to documents such as: allow categorization, versioning, etc. When a rule is defined on a space, it could be applied to all subspaces, thus avoiding defining the same rule many times.
Step 6:	Defining the user interface
	Alfresco supports various ways of customizing the web-client user interface, e.g., by using *FreeMaker* templates, by editing the configuration (XML) files, by adding user specific icons, by using the *Personal Dashboard Wizard,* by using the *User Profile* option, etc. Regarding the customization of Alfresco for use in local languages of the West Balkan countries, Alfresco ECM supports the *Croatian,* as well as two variants (Latin and Cyrillic) of the *Serbian* language.
Step 7:	Defining the workflows
	Workflow is a process during which documents are going from one user to another according to predefined rules, called workflow definition. Alfresco application offers two types of workflows: simple workflow and advanced workflow. Simple workflow is content oriented, represents the movement of documents from one space to another and each workflow definition needs to be related to a single state. Embedding the JBPM (*JBoss Business Process Management*) engine into its core, Alfresco also supports defining task-oriented workflows.

4.3.   Using the Web4WeB e-collaboration solution

Once the password validation has been successful, a user can start using the functionalities of the solution granted to his/her role. The user interface for the users from the West Balkan region, as well as users from the EU countries is presented in Fig. 3. After its establishment in August 2008, the solution is often used for:

− organization and storage of the project results (papers from Web4WeB workshop[10] and other papers, e-Learning materials on Semantic Web in different formats),
− document exchange between co-workers in the Web4WeB network, automation and tracking of the e-Learning materials preparation process.

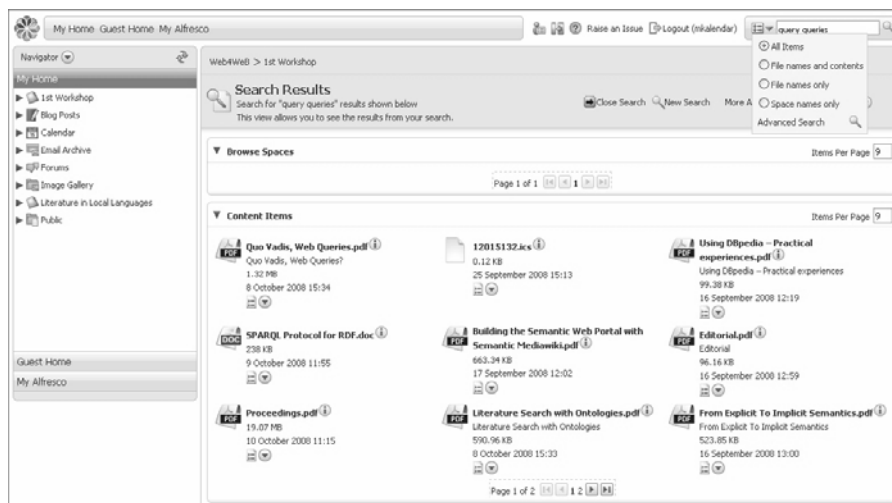[10] The first *Web4WeB Workshop* took place in Belgrade, Serbia, 29-30 September 2008 [26].

Fig. 4. Lucene searching options

Searching for documents

Using the integrated Apache Lucene full text indexing and searching service, Alfresco provides a wide range of search options for different types of contents (attached documents, e-mails, forum discussions, etc.). In the simple search mode, the user enters the keywords in the search field in the right top corner of the screen, and gets the results in the central window as is shown in Fig. 4. In order to refine the search, the user may prefer to use the advanced search option that offers different filtering methods, e.g., the user could search for only those documents that have been previously tagged with keyword "Application".

Using predefined workflows

For the needs of Web4WeB project, predefined advanced workflows are currently used to automate processes of translation of Semantic Web literatures into local Balkan languages (Fig. 5). Initially, documents ready for translation are saved in the *Translation* space and for each document the *Adhoc* workflow process is started. In this process, a new task for interpreters is generated. At the same time, interpreters can see the new assigned task on the Alfresco Dashboard, Fig. 6.

On the Alfresco dashboard, we can notice that each task is described with: short description, start date, due date, status and priority. Following the link, *manage task* interpreters can access the document, edit it, update or attach a new version of the document. In the case that more than one person works

on document, check-out is available to prevent that more persons change the document at the same time.

Once the interpreter finishes translation, he/she saves the translated document in the space *For Approval* and initiates another *Review & Approve* workflow process that aims to ask for confirmation by a Web4WeB researcher. When the Web4WeB researcher approves the task, the translated document is ready for publishing in the Web4WeB space.
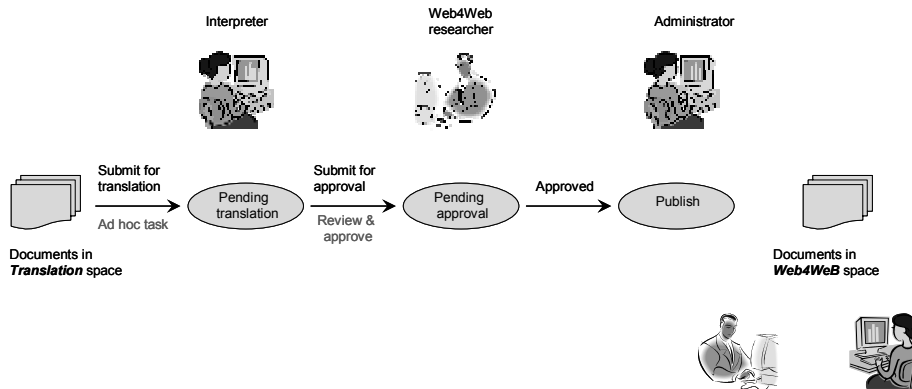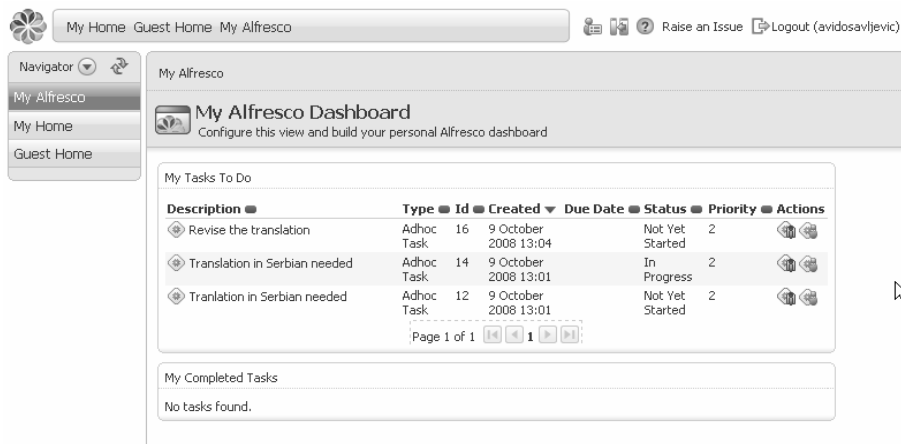

Fig. 5. Workflow process

Fig. 6. My Alfresco dashboard

Using social networking tools

To intensify collaboration between users of the Web4WeB project, we created a forum for discussing different questions about the semantic web. All registered experts and researches have access to this forum. They are

allowed to create new topics and to take part in the already existing discussions.

Beside *Forum*, all users whose home space is Web4WeB can use other features like a blog post where the user can write his/her blogs about the Semantic Web, or to use the image gallery for exchanging photos.

5. Semantic Media Wiki portal

Wikis have become a very popular tool for online collaboration, collecting information on various topics, and for sharing the collected information. The Wiki paradigm is that information should be "world readable and world writable", which conforms nicely to Web 2.0 definitions and thus enables large base of content providers. The information in the Wikis is (hyper) text based, which means that the knowledge is in the text form which requires reading of pages by humans to use it. There is no explicit, machine readable information. Some of the problems with storing of information in plain text format are: the content can be inconsistent because the same information can be placed in different places, finding and combining information is cumbersome in large Wikis as it must be done manually, the information cannot be easily reused in other applications beside user web browsers [16], etc. But knowledge is more than just text, as it also includes how pieces of information relate to each other, in other words, it is structured.

To address the problems with classical Wikis, Semantic Wikis provide additional semantic technologies features. One of the most popular and mature implementations of this concept is the Semantic MediaWiki. On top of the collaborative features for content management of MediaWiki, the Semantic MediaWiki extension provides the seamless integration of semantic features for entering, editing and presentation of the semantic data into the existing user interfaces [24].

The acquisition of the semantic data is achieved via annotations to the wiki-text with the special markup. The annotations in the Semantic MediaWiki are page centric, which means that the information in the annotations refers to the abstract concept represented by the given Wiki page. Thus, unlike RDF statements (subject-predicate-object) the annotations have only predicate and object, as the subject is implicitly given by the location of the annotation (i.e., by the page they are entered).
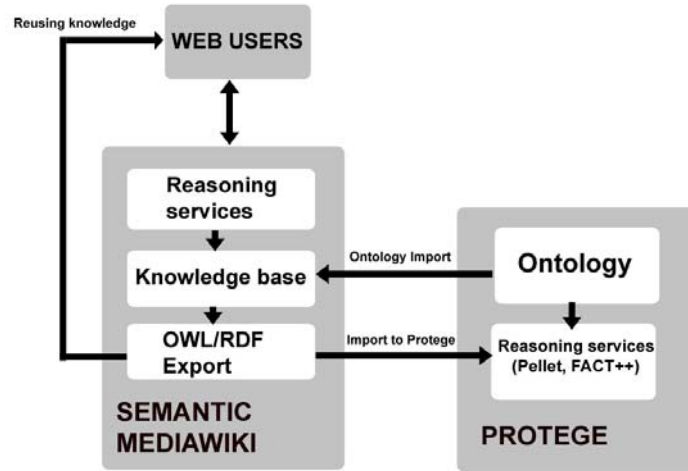
Fig. 7. Infrastructure of Semantic Web portal based on Semantic MediaWiki

Table 2. Description Logic entities to Semantic MediaWiki maping

	OWL DL	Semantic MediaWiki	Example
TBOX	Class	Page in Category namespace	www.web4web.org/portal/Category:Project
	Property	Page in Property namespace	www.web4web.org/portal/Property:homepage
ABOX	Individual	Page in Main namespace	www.web4web.org/portal/Web4WeB
	Class membership	Category annotation on normal page	[[Category:Project]] (annotation on Wiki page)
	Relations	Property annotation on the subject page	[[Homepage:http://www.web4web.org]] (annotation on Wiki page)

5.1.   The system architecture

The schema of the overall Semantic portal is shown in the Fig. 7. The core of the system is the Semantic MediaWiki, which is supported by the ontology system Protégé, for using the ontological knowledge. The two systems are loosely coupled via import/export functionalities. The Semantic MediaWiki provides the semantic knowledge base around the wiki pages, its own reasoning services for the knowledge base, which mostly deal with query

answering, and the OWL/RDF export for reusing the knowledge present in the wiki. Semantic MediaWiki is based on the LAMP (Linux-Apache-MySQL-PHP) development paradigm.

5.2.   The Web4WeB Portal Knowledgebase

The Semantic web portal is based on Description logics and a portal ontology.

Description logics (DL) are a family of logic-based knowledge representation formalisms, seen as sub-languages of predicate logic (first order logic), designed to represent and reason about the knowledge of an application domain in a structured and well understood way [20]. Description Logics are currently the most used formalisms for building ontologies, and have been proposed as standard languages for the specification of ontologies in the Semantic Web. The basic building blocks are concepts, roles and individuals. Concepts describe the common properties of a collection of individuals and can be considered as unary predicates which are interpreted as sets of objects. Roles are interpreted as binary relations between objects. Each description logic defines also a number of language constructs (such as intersection, union, role quantification, etc.) that can be used to define new concepts and roles. Depends on the number of the applied constructs and axioms, different levels of expressivity are achieved.

DL ontologies have two main parts for storing formalized knowledge (see Description Logic Handbook [3]). The TBOX part is for schematic information (i.e. information on classes of things), and the ABOX for knowledge about the individual things. Hence, the reasoning services in the Description logics can be separated into two classes, TBOX reasoning, and ABOX reasoning. TBOX reasoning deals with such problems as computing the inferred superclasses of a class (classification), determining whether or not a class is consistent (a class is inconsistent if it cannot possibly have any instances) or deciding whether or not one class is subsumed by another (subsumption). ABOX reasoning deals with checking the consistency of the facts stated about individuals with the given TBOX, or with query answering based on facts stated in the ABOX (instance checking).

In the Semantic MediaWiki the ontological elements of the TBOX are represented by pages in the wiki, while the ABOX assertions are represented by annotations on the appropriate pages which represent given individuals.

The Description Logic to Semantic MediaWiki mappings are presented in Table 2.

Portal ontology

Following the best practices of reusing existing ontologies, the Web4Web.org Portal ontology (see Fig.8) includes concepts and properties from the the SWRC ontology [22], the Dublin Core vocabulary and DCMI Type Vocabulary.

- Organization
    - Company
    - University
    - Research Institute
- Document
    - Deliverable
    - News Item
    - Journal
- Project
- Tool
- Event
    - Conference
    - Workshop
- Collection
    - Repository
    - Initiative
    - Web News
- Semantic Web topic

Fig. 8. The Web4WeB ontology tree

The SWRC ontology[11] (Semantic Web Research Community Ontology) generically models key entities relevant for typical research communities and the relations between them. The ontology has found it's use in numerous applications and projects including the AIFB portal, Bibster and the SemIPort project. Currently the Web4WeB ontology imports 13 concepts and 20 properties from the SWRC ontology (see http://www.web4web.org/portal/MediaWiki:Smw_import_swrc).

The Dublin Core Metadata Element Set[12] is a vocabulary of fifteen properties for use in resource description. The Dublin Core Metadata Element Set is part of a larger set of metadata vocabularies and technical specifications maintained by the Dublin Core Metadata Initiative (DCMI)[13]. The Web4WeB ontology imports all 15 properties from the Dublin Core vocabulary (see http://www.web4web.org/portal/MediaWiki:Smw_import_dc).

The DCMI Type Vocabulary[14] provides a general, cross-domain list of approved terms that may be used as values for the Resource Type element to identify a resource. Currently the Web4WeB ontology imports 8 concepts from the DCMI Type Vocabulary (see http://www.web4web.org/portal/MediaWiki:Smw_import_dcmi).

Currently the Web4Web ontology consists of 19 classes and 30 properties imported from the SWRC and the DCMI Type Vocabulary. Each Web4Web category is defined in a form of a Semantic MediaWiki (see http://www.web4web.org/portal/Special:Categories). The Web4Web ontology tree looks as follows.

---

[11] http://ontoware.org/projects/swrc/.

[12] http://purl.org/dc/elements/1.1/.

[13] http://dublincore.org/.

[14] http://dublincore.org/documents/dcmi-type-vocabulary/.

Each page on the Web4WeB portal is described with semantic properties that can be found at the bottom of the page (see Fig. 9).

Fig. 9. Web4Web Repository of e-learning modules described with semantic properties

Ontology import

The portal ontology built in Protégé or coming from some other OWL knowledge base can be imported to Semantic MediaWiki. As of version 1.2 of the Semantic MediaWiki there is no functionality to import ontologies in the Wiki knowledge base automatically. When dealing with shallow ontologies (small TBOX) with (almost) empty ABOX the import can be easily done manually by an administrator. To import a bigger ontology with large TBOX and/or ABOX, from an external ontology building system such as Protege, an automated procedure should be developed based on the mappings from Table 2.

During building (customizing) the Semantic portal, for each Class of the portal ontology a page in the Category namespace should be created, while for every Property we must create a page in the Property namespace. Each ontology individual (instance) is represented as a normal Wiki page. For the class instantiations of individuals from the ontology, Category annotations are created in the wiki, and for properties of the individuals the Property annotations should be made. This completes the mappings, but although the ontological elements are present in the Semantic Wiki knowledge base, their meanings are now local to the Wiki. For example, we would like to say explicitly that the *Category:Project* in our Web4WeB.org Semantic MediaWiki

has the same meaning as *SWRC:Project* concept. Fortunately, the SemanticMedia Wiki has two mechanisms for doing so, the *imported from*, and the *Equivalent URI* properties.

The *imported from* property of the Semantic MediaWiki has a built-in meaning that enables the users to provide the explicit meaning to the ontological concepts in the wiki knowledge base. The pages with this property are considered to represent the same concepts that the *imported from* annotation says. For example, in the Web4WeB portal the Category:Project contains the following annotation *[[imported from::swrc:Project]]* which says that instead of having local meaning to the wiki the class of Projects represent the *SWRC:Project* class. This information is used by all exports of the knowledge contained in the wiki. In order to enable the use of the SWRC ontology with the *imported from* property the wiki administrator must create a page in the wiki which defines the usage of the SWRC ontology in the wiki. For each ontology that is to be used in the Wiki a new special page must be created, and mappings for all ontological elements that can be used with the *imported from* property must be defined. Therefore, the *imported from* property is suitable only for representing the TBOX elements. The advantage of this approach is that wiki annotations do not use full URIs of the ontological concepts, but short versions. In the previous example instead of using the full URL: http://swrc.ontoware.org/ontology# of the ontology the shorthand *swrc:Project* is used.

The other mechanism for giving explicit semantics to local ontological elements of the wiki is the *equivalent URI* property. This property is suitable for individuals also. The *equivalent URI* is a special property in Semantic MediaWiki with a built-in meaning; it marks a page in the wiki as having a well-known meaning beyond this wiki. The meaning is defined by the external URI. In RDF Export the "Equivalent URI" special property exports as *owl:sameAs*. This OWL property indicates that the article in the wiki and the external URI actually refer to the same "identity

For creating an automated script for importing the ontological elements from OWL/RDF to the Wiki we used the Python programming language, with a great help from two external libraries, Pywikibot and RDFlib.

Reusing portal knowledge

Besides using the portal for local knowledge management tasks the gathered knowledge can be used for any other purpose. The data in the knowledge base is semantically enriched according to any ontology used by the portal explicitly, but also used by the Semantic MediaWiki *sviwt* ontology. The RDF data for a single page can be retrieved via RDF Feed link present on the fact box (see Fig.9) of every page and the results are in the form of an RDF graph, as is presented in Fig. 10. The exported data could be accessed via the local mounted SPARQL endpoint, provided by using the RAP library, or by any SPARQL endpoint on the Web.

```
<?xml version="1.0" encoding="UTF-8" ?>
<!DOCTYPE rdf:RDF (View Source for full doctype...)>
- <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:smw="http://smw.ontoware.org/2005/smw#"
    xmlns:wiki="http://127.0.0.1/portal/Special:URIResolver/"
    xmlns:property="http://127.0.0.1/portal/Special:URIResolver/Property-3A" xmlns:dcmi="http://purl.org/dc/dcmitype/"
    xmlns:swrc="http://swrc.ontoware.org/ontology#" xmlns:dc="http://purl.org/dc/elements/1.1/">
    <!-- ontology header -->
- <owl:Ontology rdf:about="">
    <smw:exportDate rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2009-04-
      10T15:19:26+02:00</smw:exportDate>
  </owl:Ontology>
  <!-- exported page data -->
- <smw:Thing rdf:about="http://127.0.0.1/portal/Special:URIResolver/ELearning_Modules">
    <rdfs:label>ELearning Modules</rdfs:label>
    <smw:hasArticle rdf:resource="http://www.web4web.org/portal/ELearning_Modules" />
    <rdfs:isDefinedBy rdf:resource="http://www.web4web.org/portal/Special:ExportRDF/ELearning_Modules" />
    <rdf:type rdf:resource="http://purl.org/dc/dcmitype/Collection" />
    <swrc:homepage rdf:resource="http://www.web4web.org/portal/ELearning_Modules" />
    <dc:coverage rdf:datatype="http://www.w3.org/2001/XMLSchema#string">WestBalkan</dc:coverage>
    <dc:title rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Web4WeB Repository of e-learning Semantic Web
      Materials in local languages</dc:title>
    <dc:description rdf:datatype="http://www.w3.org/2001/XMLSchema#string">The objective of the Web4WeB e-learning
      modules is to provide an introduction to the Semantic Web (SW) in languages of nationalities that are part of the West
      Balkan region. Covering the basis of the Semantic Web field including concepts, methods, standard languages, the
      Web4WeB e-learning modules are a good starting point for students and newcomers in the Semantic Web field. Next,
      providing an updated picture of the Semantic Web state of the art research activities and initiatives, the Web4WeB e-
      learning modules are also very useful for scientific researchers already dealing with Semantic Web technologies.
      Furthermore, summarizing the achievements in the Semantic Web field and the contribution of Semantic Web research
      to industry, and surveying the market of commercial and open source SW products, Web4WeB e-learning modules could
      help the IT professionals easily choose the right SW tool and technology. And finally, the analysis of the maturity and
      applicability of SW technologies in practice is intended for IT managers, as well as for top management to help them
      understand the usability of SW technologies and plan the migration of existing systems to Semantic
      Web.</dc:description>
  </smw:Thing>
  <!-- auxiliary definitions -->
```

Fig. 10. RDF Export

5.3.   Customizing the Web4WeB portal

Semantic MediaWiki enables the full freedom in defining the semantics of the page. If this freedom is not controlled in some way it can lead to more concepts and properties than are actually needed. Therefore, sometimes it is good to impose a structure on some pages, especially when the end users have no knowledge about the underlying knowledge model of the Semantic MediaWiki. Besides providing a predefined structure (in the form of an ontology), it is also favorable to allow the end users to enter some predefined semantic annotations without learning the simple annotations language of Semantic MediaWiki. The Semantic Forms is an extension to Semantic MediaWiki that allows users to add and edit semantic data using web forms. When an external ontology is used, like in case of the Web4WeB portal, the Semantic Forms can be used to lead the users to follow the structure defined by the ontology.

The Semantic Forms extension allows declarative definitions of forms, which enables adding of various forms in the Semantic MediaWiki system without any programming. Forms can be created and edited not just by administrators, but by users themselves. Currently, the extension only enables the creation of forms, while editing of forms is not yet possible.

The main components of Semantic Forms functionality are form definition pages, for which a new *Form* namespace is defined. These are pages consisting of the markup code which gets parsed when a user wants to add or edit data.

The Semantic Forms extension uses the templates in creating semantic data, as the semantic markup is stored indirectly through templates. So the first step is to create a template, either by hand or by using the page *Special:Create_Template*. When the template is created it is possible to create a Form for that template. The form contains the fields defined in the template to enable users to enter semantic annotations via populating the simple web forms. If necessary, it is possible to include several templates in a single form. An example of creation of a form is shown in Fig. 11.

Fig. 11. Creating a Form with Semantic Forms

5.4. Browsing the portal with Semantic Drill down

The Semantic Web4WeB portal [9] can be reached via the following address: http://www.web4web.org/. The Semantic MediaWiki integrates an extension

named Semantic Drilldown, that enables "drilling down" through data in the wiki knowledge base by using categories and filters on semantic properties. The Semantic Drilldown "Browse data" shows all the top-level categories in the wiki; i.e., the categories that are not subcategories of another category, and the number of pages within that category. Each category name is a link to a drilldown for the pages in that category. It lets the user select additional constraints to limit the number of results. These constraints come in two types: Subcategories – for navigation through an entire category tree, and Filters – for navigation based on manually added filters based on values of any semantic property. In the Fig. 12 we can see the navigation filter which is defined for the *Project Category* for the *project type* property.

Fig. 12. Browsing the wiki with Semantic Drilldown

5.5. Querying and Reasoning services within the Semantic Web portal

Semantic MediaWiki provides a limited set of reasoning services for its knowledge base. The support exists for querying the knowledge base, with it's own simple query language whose syntax is based on the MediaWiki's mark-up language. The query language corresponds to concepts of description logic EL++, and it has a subset of SPARQL query language capabilities. A rather primitive example of the Semantic MediaWiki query which is used on the Web4WeB.org portal is the following:

```
<ask format="table" limit="3">
```

```
[[Category:Project|Project]]
[[Subtitle::*]]
[[Description::*]]
[[ProjectStatus::Execution]]
</ask>
```

This query can be used either in interactive mode by typing it on the special query page, or in the stored mode, where the queries are saved on the page, by some user who is familiar with queries. The page with the query results can then be read by all users, even by those that don't know the syntax of the language. The results of the query against the Web4WeB.org knowledge base are shown in the Fig 13.
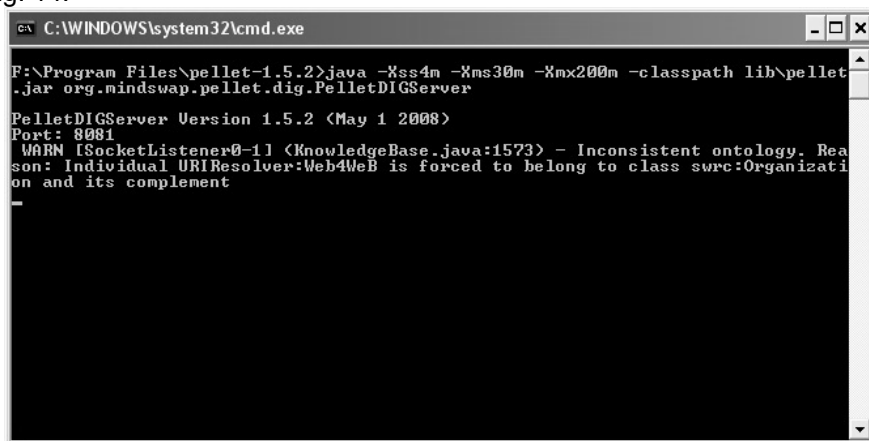

Fig. 13. Querying the Web4WeB knowledge base

The Semantic MediaWiki on which our Semantic Web portal is based currently doesn't provide some very important reasoning features. Besides a simple query language of the Semantic MediaWiki the complex reasoning is left out to external DL reasoners. Although it is possible to import background-knowledge of the domain in the knowledge base of the Semantic MediaWiki, by importing ontology, the ontology itself is not fully used while editing and presenting the content. For example, it is possible to assert relationships that are not consistent with the ontology being used, as there is no consistency checking inside the wiki.

These reasoning tasks can be performed outside the Wiki. The complete knowledge base of the portal is first exported to the OWL/RDF file using a batch export script. This data is then imported to the Protege ontology editor, which is connected with external DL reasoners like Pellet and Fact++. For this connection, the DIG protocol of the DL Implementation group is used. The

DIG is an XML based representation of ontological entities and is used as standard API for interaction with DL reasoners.

The main reasoning task for our portal is to check the consistency of our knowledge base. To give an example of that, we made the Web4WeB project to belong to two categories which belong to two disjoint classes, Project and Organization. The result of the pellet DL reasoner for this example is shown in Fig. 14.

Fig. 14. Using Pellet to detect inconsistencies in the wiki

Protégé didn't displays full information about the inconsistency but only the fact that the knowledge base was inconsistent, so we had to check the pellets message directly on it's console output.

6.   Comparative analysis of functionalities of Alfresco ECM and Semantic MediaWiki CoKM

Taking into consideration that the Alfresco Enterprise Content Management System and the Semantic MediaWiki Cooperative Knowledge Management System serve the same goal of "enhancing the cooperation of researchers from the West Balkan in Semantic Web field", these two tools could be compared with regard to Social Semantic Web readiness, knowledge representation (expressivity), usability, flexibility, scalability, interoperability, and so on.

6.1.   Semantic Web readiness

Semantic MediaWiki is an extension to MediaWiki that allows for the encoding of semantic data within wiki pages, thus turning a wiki that includes the extension into a semantic wiki. Data that has been encoded can be used in

semantic searches, used for aggregation of pages, and exported to the outside world via RDF. The Semantic MediaWiki extension makes the MediaWiki ready for the Semantic Web. Thus, nowadays, Semantic MediaWiki is widely accepted for building Semantic Web knowledge management applications.

In contrast with the Semantic MediaWiki, Alfresco ECM is a social computing platform that is not Semantic Web ready. This is due to the fact that the metadata and tags that are used to determine the document/content are locked in the relational database. Hence, special services are needed, that do not exist at this moment, to export the metadata in a machine understandable and interpretable (RDF format).

Table 3. Assesment of Web4WeB e-collaboration tools

Functionality	Alfresco ECM	Semantic MediaWiki
SW Readiness	no	yes
Knowledge representation	Database schema	Ontologies, RDF export
Flexibility	- advanced document management system - covers a wide range of collaborative and networking functionalities: e-mail notification, robust user profiles, blogs, discussion boards.	- simple Web document management system - ideal for collaborative document production
Usability	- easy-to-use - provides extensive number of options for personalizing the user interfaces - provides a wide range of search options based on Lucene engine	- easy-to-use - using ontologies, provides a powerful Drill down navigation option
Scalability	could provide acceptable performance with a large content store (greater than 100 million documents)	scalability is a major obstacle when handling large datasets, however the performances of the Semantic MediaWiki version 1.2 satisfy the Web4WeB needs

6.2. Knowledge representation (expressivity)

It is desirable to have on the Web as much knowledge as possible in a machine processable format, but it is well-known that this often conflicts with usability and performance. In Semantic MediaWiki this particularly affects advanced features, such as reasoning with time and space, for which practical solutions are still sought. Semantic MediaWiki provides various means of structuring its content, and such existing structures are a natural choice for formalization. However, it was also pointed out that Semantic MediaWiki by default does not integrate logical inconsistency checking mechanisms and this is left to the developers of the concrete application to deal with the problem.

Alfresco provides possibilities for structuring data similar to Semantic Forms in MediaWiki and does not have problems with logical inconsistency as this is solved at the database level. However, once entered in the Alfresco system, reuse of knowledge is not that simple as it is with the RDF export of MediaWiki. Reuse of knowledge in Alfresco is thus an interoperability problem.

6.3. Flexibility

Both tools can be employed for a great variety of knowledge acquisition applications, and users can adjust the form and content of the collected information in almost unrestricted ways. While Semantic MediaWiki is ideal for collaborative document production, it lacks some social networking functionalities such as sophisticated e-mail notification, robust user profiles, blogs, discussion boards, etc. The Alfresco system could be used both as a social networking platform and a document management solution. The aspect-oriented programming approach allows Alfresco to define different processing/transformation rules on documents and thus be used in complex enterprise content management tasks. The integration of jBPM workflow management system further extends the applicability of Alfresco in business process management field.

6.4. Usability

Although Alfresco ECM and Semantic MediaWiki CoKM are developed using different architectural models, both tools provide extensive number of options for personalizing the user interfaces that make them easy-to-use. In our view users without prior training could use the systems, while little technical background is needed for customizing Alfresco or editing/annotation of the Semantic MediaWiki pages.

6.5. Scalability

The investigation of the practical use of semantic technologies showed that SW technologies are still in an early maturity stage, while scalability is a major obstacle especially when handling large datasets (e.g. Wikipedia stores several million pages + articles, DBpedia stored more than one hundred million triples) or massively concurrent users [2], [13]. In order to increase scalability, semantic technologies had to adopt in practice approved approaches and techniques developed in the database research community. Benchmarking results on the scalability of Alfresco ECM system showed that Alfresco system could provide acceptable performance with a large content store (greater than 100 million documents)[23].

Considering that the research network on the Semantic Web technologies from West Balkan countries will have up to one hundred users and a small number of documents, both tools fulfill the needs of the Web4WeB project.

7. Conclusion

This paper introduces two open source collaboration and content management systems and assesses their applicability for building a research network that aims to contribute to sustainable Semantic Web research development in the West Balkan countries. The analysis of the capabilities of both tools has shown that the Alfresco Enterprise Content Management System and Semantic MediaWiki Collaborative Knowledge Management System are adequate choice for enhancing and strengthening the collaboration of the Semantic Web fellows.

While both tools provide the basic collaboration features (editing of contents and e-mail), they support also complementary functionalities. The strength of Alfresco ECM is in integration of different types of rules that could be used to automate the document processing and document exchange. The integrated jBPM workflow engine could further facilitate the notification process and collaborative work. The strength of the Semantic MediaWiki is in structuring the knowledge in a Semantic Web standard form that allows integration of the knowledge collected/created by the West Balkan researchers in the European research space and beyond.

Future work might include study and assessment of models for e-collaboration, examination of the impact of collaborative technologies on group work, and establishment of additional services for the Web4WeB network members for competence building in the Semantic Web field.

Valentina Janev, Jovan Duduković, Jelena Jovanović, and Sanja Vraneš

8. Acknowledgments

9. References

1. Auer, S., Bizer, C., Lehmann, J., Kobilarov, G., Cyganiak, R., & Ives, Z.: DBpedia: A Nucleus for a Web of Open Data. In: Aberer et al. (eds.): The Semantic Web, 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007, Busan, Korea. Lecture Notes in Computer Science 4825/2007. Springer-Verlag, Berlin Heidelberg New York (November 11–15, 2007)
2. Auer, S.: Methods and Applications of the Social Semantic Web. In: Vraneš S. (ed.): Semantic Web and/or Web 2.0: Competitive or Complementary?, 100-128. Academic mind, Belgrade. (September 29-30, 2008)
3. Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P.: The Description Logic Handbook. Cambridge University Press. (2003)
4. Berners-Lee, T., Hendler, J., & Lassila, O.: The Semantic Web. Scientific American, May 2001. Online: http://www.sciam.com/article.cfm?id=the-semantic-web.
5. Bojars, U., Breslin, J., Peristeras, V., Tummarello, G., Decker, S.: Interlinking the Social Web with Semantics. IEEE Intelligent Systems 23(3), 29-40. (2008)
6. Box, D., et all.: Simple Object Access Protocol (SOAP) 1.1. W3C Recommendation. (2000)
7. Brickley, D., Guha, R.V.: RDF Vocabulary Description Language 1.0: RDF Schema. W3C Recommendation. (2004)
8. Doms, A., Dietze, H., Wachter, T. & Schroeder M.: Literature search with ontologies, a Tutorial. In: Vraneš S. (ed.): Semantic Web and/or Web 2.0: Competitive or Complementary? 81-100. Academic mind, Belgrade. (2008)
9. Duduković, J. and Janev, V.: Building the Semantic Web Portal with Semantic Mediawiki In: Vraneš S. (ed.): Semantic Web and/or Web 2.0: Competitive or Complementary?, 185-191. Academic mind, Belgrade. (September 29-30, 2008)
10. Fielding, R.: Architectural Styles and the Design of Network-based Software Architectures. Ph.D. dissertation, University of California, Irvine. (2000)
11. Grau, B. C., Motik, B.: OWL 2 Web Ontology Language - Model-Theoretic Semantics. W3C Recommendation. (2007)
12. Gruber, T.: Where the Social Web Meets the Semantic Web. Keynote presentation at ISWC The 5th International Semantic Web Conference. (November 7, 2006)
13. Janev V.: A comparison and critical assessment of cutting edge Semantic Web products and technologies. In: Vraneš S. (ed.): Semantic Web and/or Web 2.0: Competitive or Complementary? 148-165. Academic mind, Belgrade. (September 29-30, 2008)
14. Kinsella S., Breslin, J.G., Passant, A., & Decker, S.: Applications of Semantic Web methodologies and techniques to Social Networks and Social Websites. In:

C. Baroglio et all., (eds.): Reasoning Web, 4th International Summer School 2008. Lecture Notes in Computer Science 5224/2008. Springer-Verlag, Berlin Heidelberg New York, 171-199. (2008)

15. Klyne, G., Carroll, J. J., & McBride, B.: Resource Description Framework (RDF): Concepts and Abstract Syntax. W3C Recommendation. (2004)
16. Krotzsch, M., Schaffert, S., Vrandečić, D.: Reasoning in Semantic Wikis. Reasoning Web, Tutorial notes of the Third International Summer School 2007, Springer, pp. 310-329. (September 2007)
17. Manola, F., Miller, E. & McBride, B.: RDF Primer. W3C Recommendation (2004)
18. McGuinness, D. L., van Harmelen F.: OWL Web Ontology Language. W3C Recommendation. (2004)
19. Prud'hommeaux, E. & Seaborne, A.: SPARQL Query Language for RDF. W3C Recommendation (2007)
20. Sattler, U.: Reasoning in Description logics: Basics, Extensions, and Relatives. Reasoning Web, Tutorial notes of the Third International Summer School 2007. Springer, pp. 155-182. (September 2007)
21. Spivack, N.: How the WebOS Evolves? Nova Spivacks WebLog. Radar Networks.
    http://novaspivack.typepad.com/nova_spivacks_weblog/2007/02/steps_towards_a.html. (February 09, 2007)
22. Sure Y. et al.: The Ontology – Semantic Web for Research Communities. Progress in Artificial Intelligence, Lecture notes in Computer Science, pp. 218-231. (November 2005)
23. Benchmark Results: Scalability of Alfresco Content Management System in a Unisys ES7000/one Enterprise Server Environment. White paper, Unisys Corporation. (Decembar, 2007)
24. Volkel, M. , Krotzsch, M. , Vrandecic, D., Haller, H. and Studer R.: Semantic wikipedia. In: Proceedings of the 15th international conference on World Wide Web, pp. 585 – 594 (2006)
25. Vraneš, S.: D5.3 – Implementation of groupware/collaboration features, Web4WeB project deliverable. (November 2008)
26. Vraneš, S. (ed.): Semantic Web and/or Web 2.0: Competitive or Complementary? Academic mind, Belgrade. (2008)
27. Welty, C. (2007). How I was right even when I was wrong. In: In Aberer et al. (Eds.): The Semantic Web, 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007, Busan, Korea, Lecture Notes in Computer Science 4825/2007. Berlin / Heidelberg: Springer. (November 11–15, 2007)

Valentina Janev, MSc, is a Research Associate at the Mihailo Pupin Institute, Belgrade, Serbia. Since 1996 she has participated in design, development and implementation of many business information systems first as a Software Engineer and later as a Consultant. She is actively involved in scientific projects financed by the Serbian Ministry of Science, as well in EU FP6 and FP7 projects. Her resent interests include Semantic Web, applications of semantic technologies and W3C standards in the knowledge management field, and knowledge management architectures for R&D organizations. She is an author of more than 30 published papers, some of them in international journals and books.

Valentina Janev, Jovan Duduković, Jelena Jovanović, and Sanja Vraneš

Jovan Duduković, Dipl. ing., is a Junior researcher at the Mihailo Pupin Institute, Belgrade, Serbia. He has received a Dipl. Ing. degree in 2004 at the School of Electrical Engineering, University of Belgrade, Serbia. While working in scientific projects financed by the Serbian Ministry of Science and EU (FP6 and FP7) he took part in the design and implementation of several decision support systems (SARIB), design and implementation of Semantic Web Portal (Web4WeB), and implementation of enterprise applications (SUSDEV). He has experience in enterprise application software development, GIS application development, web application architecture and database development. His interests include Semantic Web, Knowledge representation and architecture of software systems.

Jelena Jovanović, finished Faculty of Mathematics, group for computer science, in Belgrade, Serbia. Since 2002, she has been employed with the Mihajlo Pupin Institute. She took part in design and implementation of CyberSchool (School management system, 2003), WiseChoice (multi-criteria decision support system, 2004), SARIB (SAva River Basin: Sustainable Use, Management and Protection of Ressources, 2004–2007) and Prometea web portal (project "Empowering women engineers" 2005-2007). She also took part in several projects financed by the Ministry of Science and Technology of Serbia.

Dr. Sanja Vraneš is jointly appointed as a Scientific Director of the Mihailo Pupin Institute and as a Professor of Computer Science at the University of Belgrade. From 1999 she has been engaged as a United Nations Expert for information technologies, and from 2003 as an EU Expert for intelligent content and semantics area. Her research interests include semantic web, knowledge management, intelligent agents, e-collaboration, decision support systems, multicriteria analysis algorithms, etc. In these areas she published over 150 scientific papers. She serves as a reviewer of respectable international journals, like IEEE Transaction on Computer, IEEE Intelligent Systems magazine, and also as a reviewer and evaluator of EU FP6 and FP7 projects. She has also served as a project leader and/or principal architect of more than 20 complex software projects. She is a member of IEEE, ACM and AAAI. She is also a member of Serbian Academy of Engineering Sciences and of National Scientific Council.