From Information Wiki to Knowledge Wiki via Semantic Web technologies

Sergiu Dumitriu, Marta Gîrdea, Sabin C. Buraga
Faculty of Computer Science, "A.I.Cuza" University of Iaşi
16 Berthelot, 700483 Iaşi, Romania – Email: {sergiu.dumitriu, marta, busaco}@infoiasi.ro

***Abstract* — The paper presents an enhanced version of an existing wiki platform (XWiki), in order to integrate knowledge, based on various Semantic Web technologies. In particular, we describe how metadata, microformats, ontologies are meaningfully used in this context, and we show the utility of our approach via two use cases.**

I. Introduction

The World Wide Web space is primarily compounded by pages (documents that contain markup) with information in the form of natural language text and multimedia, intended for humans to read and to understand. Computers are principally used to render this hypermedia information, not to reason about it. However in the next stages of Web evolution, information is no longer intended only for human readers, but also for machine processing, enabling intelligent information services, personalized Web sites, and semantically empowered search engines – this is the seminal idea of the *Semantic Web* [4, 7]. The Semantic Web is viewed as "an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation" [4].

When advancing towards Semantic Web, the main obstacle is the effort that the creator of hypermedia information must put into organizing the knowledge and metadata, into tagging entities and relations, using vocabularies he must be familiar with, in order to make it comprehensible not only for humans, but also for machines. A possible answer to this problem could be a platform which helps *transparently* organize data and metadata for machine-comprehensibility, so that the contributors do not need to know Semantic Web vocabularies, as the system automatically generates metadata based on the creator's actions and on the progress of the information manipulated within the platform.

In order to create the knowledge base on a certain domain, one can use the power of online communities. The predominant type of collaborative Web application is represented by wikis [10]. For example, major successes of the wiki concept are the well-known Wikipedia [29] and its related initiatives.

The established wiki systems provide a basic Web interface for (collaborative) content editing by using a simplified syntax. Main facilities are unrestricted editing, the rollback mechanism, several search functions, the support for uploading content. Wiki platforms are currently used for different purposes, such as encyclopedias, collaborative writing, personal or public information management, project management, and many others [14].

To achieve significant knowledge acquisition, as opposed to simple data acquisition, a modern wiki system must provide a means of adding metadata about the concepts and relations established between the concepts within a given wiki.

The goal of this paper is to detail a semantic Web-based extension of the XWiki platform, which can be used to present, create and manage knowledge by both experienced and inexperienced users.

After a short presentation of the current initiatives towards Semantic Wikis, in section III we detail the general architecture of the XWiki system. Section IV is dedicated to the semantic Web-based enhancements of the XWiki platform, followed by two case studies. In the final section, the paper concludes and anticipates further directions of research and development.

II. Context, Related Work and Initiatives

A. Short Presentation of Semantic Web Technologies

Semantic Web technologies are based on the XML [5] family of markup languages and are structured on three main layers [7]:

- *Metadata* layer offers an extensible framework in order to express simple semantic assertions. This conceptual model can be use to attach metadata to resources. Using metadata, we can formulate statements about certain Web resources in the form of <*subject*, *predicate*, *object*> triples. There are available several metadata vocabularies such as DCMI (Dublin Core Metadata Initiative) [17], FOAF (Friend of a Friend) [27], or XFiles [6].
- *Ontology* layer defines a hierarchical description of the concepts and properties of a given resource, from

T. Sobh (ed.), Innovations and Advanced Techniques in Computer and Information Sciences and Engineering, 443–448.

simple taxonomies to complex upper-level ontologies.
- *Inference* layer offers the possibility to reason about knowledge by using rules.

Several XML-based languages specialized in the modeling of knowledge were developed in the context of Semantic Web: RDF (Resource Description Framework) [11] and OWL (Web Ontology Language) [8]. Ontological classes are denoted by OWL assertions, and their instances (individuals) are expressed by RDF statements.

A different direction is the use of microformats [22], with the goal of embedding semantic data directly within documents conforming to a presentational XML format (especially XHTML). For example, we can embed complex taxonomies and schemas on the basis of the markup of regular Web pages (e.g., blog entries, wiki documents, etc.). With the help of microformats and appropriate vocabularies we can denote personal virtual cards, tags, relations between persons, events and calendars, etc. This approach is currently used in the creation of the Social Web, one of the facets of "Web 2.0" [13].

B. Semantic wikis

Several initiatives of wiki platforms which provide support for expressing knowledge via current Semantic Web technologies are already available. Most of them are simply ontological editors in the wiki style, which can not be used by people without a prior knowledge of the (Semantic) Web.

The *Rhizome* [15] system offers support for capturing and representing informal, human-authored content and is based on Semantic Web languages. Rhizome Wiki uses custom text formats to make writing semantic content easier, which can be managed by a RDF data access engine. Unfortunately, the rendered content does not contain embedded rich metadata or microformats and final users do not fully benefit of RDF-like stored knowledge. Also, Rhizome's support for expressing ontological constructs is very basic, and it requires the users to be familiar with XML and related technologies.

One of the promising initiatives is *Semantic MediaWiki* [9, 24], a prototype built on the MediaWiki platform, using an external RDF repository. A possible use of such a system is to facilitate knowledge acquisition [12].

Another Semantic Wiki for collaborative knowledge management is *IkeWiki* [14], which supports different levels of formalization (from informal texts to formal ontologies expressed in OWL) and a rich user interaction – e.g., SVG (Scalable Vector Graphics) and AJAX (Asynchronous JavaScript And XML). IkeWiki is developed in Java and uses the well-known Jena [21] semantic platform. IkeWiki presents to users links to (semantically) related pages and provides a tool for automatic generation of content from existing semantic annotations (DCMI metadata or RDF statements).

In order to ease the ontological content creation in a collaborative manner, an interesting platform is *OntoWiki*, based on pOWL [2]. According to [3], *OntoWiki* is a tool that integrates RDF triples into the wiki textual content by using a special syntax in order to simplify the presentation and acquisition of instance data from and for end users in a generic way.

We noticed the majority of available semantic wiki systems can be used by specialists only and can not address the needs of regular communities to exploit the power of semantic Web technologies in a pragmatic way.

III. Architecture of the XWiki Platform

A. Description

We based our implementation on the XWiki platform [28], a powerful open source wiki engine written using J2EE technologies. XWiki is a professional wiki, having many features needed for enterprise usage. It is also an application wiki by allowing development of applications with structured data and scripting right from the Wiki interface. It can be used in a large variety of scenarios, from a simple presentation site, a personal wiki or blog, to a complex structured content repository or advanced collaborative application.

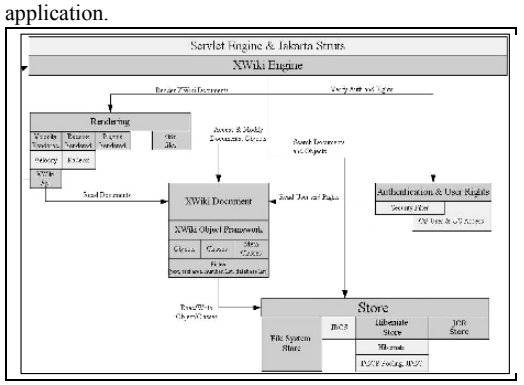

Fig. 1. Overall XWiki architecture.

XWiki is inspired by TWiki [25], a wiki engine written in Perl. The initiator and maintainer of XWiki is Ludovic Dubost, leading a growing team of developers. The first version of XWiki was made public in January 2003, and the official 1.0 version is planned for early 2007.

B. General Structure

The power of XWiki lies in the multi-layered architecture, separating the platform into distinct, loosely coupled modules, based on the MVC (Model-View-Controller) design pattern. The most important modules are depicted in Figure 1, and are described below:

- XWiki is deployed as a *servlet*-based web application, taking full advantage of the underlying server API, handling multithreading, HTTP requests and responses, thread pooling, etc.
- The *XWiki engine* ties all the other modules together, serving as an *intermediary*. It offers APIs that can be