

Drupal as an important part of a knowledge management system

Abstract. This paper is concerned with the use of the content management system Drupal to develop prototype system for supporting knowledge management in plant production. Assumptions of knowledge management system for plant production, together with general architecture of such system, are presented. Moreover, methods how to use existing Semantic Web data to enrich site content are demonstrated.

Streszczenie. Artykuł skupia się na wykorzystaniu system zarządzania treścią Drupal do zbudowania systemu wspierającego zarządzanie wiedzą w produkcji roślinnej. Zostały przedstawione założenia systemu zarządzania wiedzą wraz z ogólną architekturą system. Ponadto przedstawiono metody jak wykorzystać dane istniejące w Internecie semantycznym do wzbogacenia zawartości serwisu internetowego. (**Drupal jako ważny element systemu zarządzania wiedzą**).

Keywords: knowledge management, CMS, Semantic Web.

Słowa kluczowe: zarządzanie wiedzą, CMS, Internet semantyczny.

doi:10.12915/pe.2014.02.24

Introduction

Production of a variety of goods, business activities, scientific research and many other areas require the skilled use of knowledge. It is not sufficient to access the modern machines for competent result, companies need to know when and how to use them. Similarly, the financial resources are not enough to gain a competitive advantage, companies must have knowledge of how to use them. On the other hand, in today's world we have too much information, but this information is in general inaccurate and unordered. Currently, many of the needed information can be found on the Internet because the global network is used for publication. The mentioned activities require precise searching and collecting a variety of information. The proper use of the information is important for decision-making. In such situation, we need tools to search information and to request, depending on the context, which will make it possible to find more precise answers to our questions easier. This also applies to issues in the field of agriculture, in particular, the plant production.

This paper describes how to implement a system for knowledge management which utilizes one of the most popular content management systems – Drupal, and is organized as follows. First we discuss different types of available knowledge management systems. Next we present general assumptions of a designed system, its general architecture and process of choosing proper technology to implement it. Then it is explained of how to use Drupal RDF. Finally some conclusions and proposals how to develop the system in the future are made.

Software tools for knowledge management

The concepts of data, information and knowledge are often used inaccurately, and sometimes they are treated as equivalent. Following Ackoff [1] information is data processed in a manner useful to the user giving an answer to the question: "what?", "who?", "where?", "when?", but knowledge is the application of information corresponding to the question "how?". According to this definition we understand knowledge as a set of information to draw conclusions that are important theoretically and practically. We can define knowledge management as: "efficient handling of information and resources within a commercial organization" [14]. Davenport and Prusak [4] gave more precision definition "knowledge management is managing the corporation's knowledge through a systematically and organizationally specified process for acquiring, organizing, sustaining, applying, sharing and renewing both the tacit and explicit knowledge of employees to enhance organizational performance and create value." We not only

apply knowledge management to companies, we understand this activity more widely also for government and educational institutions.

Now we will discuss the various types of software tools that may be used for knowledge management. To be able to work effectively we need appropriate tools. In the process of knowledge management humans can be greatly supported by proper software tools. Traditional tools like books, paper notebooks and file folders are not currently sufficient to collect, organize and use information and, above all, to solve problems and make decisions. Sometimes knowledge management is characterized as delivering the right information to the right people at the right time, through use of the right IT tools. It is important to find or create the appropriate tools. Software vendors often call their products knowledge management systems for marketing purposes, companies and organizations argue that they have implemented systems for knowledge management to emphasize their prestige. This is not always true. Very often a simple database or structured routing of electronic mail is called knowledge management system, what is an obvious misuse. On the other hand, even the best tools may be improperly used and do not provide real knowledge management.

Generally speaking knowledge management tools should support different activities in knowledge life cycle. They are typically knowledge identification, creation, capturing, selection, validation, structuring, organization, distribution, visualization, maintenance, evolution and other tasks like searching or making decisions. We can classify tools according to functionality: tools supporting location and codifying explicit knowledge like databases and data warehouses [8, 9]; tools for gathering explicit knowledge; tools supporting the development of new knowledge from explicit knowledge; tools for sharing knowledge generated in repositories of explicit knowledge, for example vertical web portals [10] or Web 2.0 sites [11]; tools supporting location and codification of implicit knowledge; tools for distributed and collaborative creation of knowledge; and tools for the integration and sharing knowledge. Davenport and Prusak [4] focused on organizational aspects of knowledge management tools. They consider that three aims are the most important: to make knowledge visible and show the role of knowledge in an organization; to develop a knowledge-intensive culture by encouraging and aggregating behaviors such as knowledge sharing and proactively seeking and offering knowledge; to build a knowledge infrastructure: a web of connections among people given space, time, tools, and encouragement to interact and collaborate. Liao [13] understands knowledge

management more widely. He classified knowledge management technologies and applications into six categories: knowledge-based systems, data mining, information and communication technology, artificial intelligence and expert systems, database technology, and modeling. Others try to classify these technologies in historical aspects. For example Goluchowski [6] proposed breakdown into four generations, taking into account primarily the technology with which they are created. These are: documents oriented generation; generation targeted at supporting interpersonal communication using Web 2.0 technologies; generation based on corporate portals that combine document management with opportunities for communication between people; fourth generation using Web 3.0 semantic technologies, advanced search, and automatic processing and creation of new knowledge. We have to note that tools capable of using semantics and context of the problem are becoming increasingly important in a world of excess information. Standards and methods associated with the idea of the Semantic Web, like RDF or OWL, are becoming increasingly important also in other non-network applications. Genesereth and Dar [4] claimed that we can treat the Semantic Web as an instantiation of a document-centric knowledge management system intended to manage the entire Internet's knowledge. At the end we need to note that the most of the tools combine several listed functionalities and both old and new technologies.

Portal for system of knowledge management in plant production

Presented short survey of knowledge management tools constitutes the starting point for the selection of proper tools for the project "Knowledge management tools in plant production" implemented at the Department of Informatics at the Warsaw University of Life Sciences – SGGW.

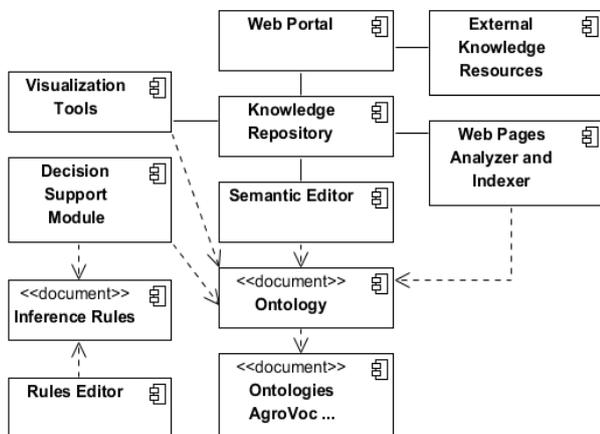


Fig. 1. Architecture of knowledge management system for plant production (own preparation)

Project realization consists of four phases. First – preliminary analysis of existing XML standards used for plant production, agribusiness and related areas, especially analysis of existing thesauri and vocabularies for plant production and agriculture together with semantic analysis of the relationships between the concepts [7]. Second phase – modeling of a preliminary version of the ontology for plant production in Poland and designing first version of the system architecture [12]. Third phase – verification and validation of the ontology for plant production, preparation of proposals of unified standards for data storage and exchange based on ontology for agriculture and agribusiness (in plant production area) in Poland, and preparing first version of web portal. And last phase –

design and implementation of the demonstration software for semantic search agriculture resources in plant production area based on the ontology, integration of all results at web portal. The design of the system and the individual components was prepared using UML. The design strategy was described in [12]. Initial architecture of the system is presented in Figure 1.

The central part of the system is the web portal for users. Some additional components of the system will be subsequently incorporated. General features of the other parts of the system are described in [12]. To create the portal, the authors decided to use an existing CMS system. Because of the academic nature of the project only systems free of charge are taken into account. Most popular free available open source systems in this class are Joomla!, Wordpress and Drupal. All three are written in PHP and use MySQL database. Drupal now has more than 175 000 installations on the Web. Outside the kernel a number of configurable modules are available. Modules to support RDF decided about selection of Drupal 7. An additional advantage was the existence of the project AgriDrupal which is described as: "both a "suite of solutions" for agricultural information management and dissemination, built on the Drupal CMS, and the community of practice around these solutions" [15]. RDF module gives possibility to import core and external vocabularies in RDF and XML format. Core vocabularies are imported automatically: content (<http://purl.org/rss/1.0/modules/content/>), dbp (<http://dbpedia.org/property/>), dbpedia-owl (<http://dbpedia.org/ontology/>), dc (<http://purl.org/dc/terms/>), foaf (<http://xmlns.com/foaf/0.1/>), og (<http://ogp.me/ns#>), rdfs (<http://www.w3.org/2000/01/rdf-schema#>), relationship (<http://purl.org/vocab/relationship/>), sioc (<http://rdfs.org/sioc/ns#>), sioc (<http://rdfs.org/sioc/types#>), skos (<http://www.w3.org/2004/02/skos/core#>), and xsd (<http://www.w3.org/2001/XMLSchema#>). Every content type can be described using vocabularies; we can define RDF type, RDF predicates and data type values. As a result we have described content automatically in the RDFa format (Fig. 2).

```

<div class="content">
<div id="node-8" class="node node-page node-
promoted node-teaser clearfix" about="/node/8"
typeof="foaf:Document">
<h2 property="dc:title" datatype="">
<a href="/node/8">Home page</a>
  
```

Fig. 2. RDFa description of node in Drupal

Such a solution allows for the description of the entire node, although it does not allow for the description of only certain elements of the content. Additionally creators of Drupal RDF module have used the concept of Linked Data [3], to integrate semantic data in RDF format from the outside. To perform this we have to register SPARQL endpoint. Drupal has a very popular module –Views, which is used to build listings of content in various formats. A view typically displays such kind of data as members of a group, and so on. Views can connect not only to local database, but to any data source and pull in data from anywhere. SPARQL Views extends Views and makes it a very powerful tool for selecting and presenting data stored in RDF format. SPARQL Views adds the ability to query SPARQL endpoints and to display the results. Creating a query in SPARQL using Drupal is rather simple and does not require knowledge of the syntax of this query language. It is sufficient that the user indicates the items that he wants to view from the RDF file together with filter criteria. SPARQL query is generated automatically (Fig. 3). The

presented solution gave us a possibility to import data related to plant production from DBPedia and AgroVoc.

```
PREFIX dbpedia-owl:
<http://dbpedia.org/ontology/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-
schema#>
SELECT DISTINCT
?plant_resource_field_plant_thumbnail
?plant_resource_field_plant_label
?plant_resource_field_plant_abstract
?plant_resource
WHERE {
?plant_resource dbpedia-owl:family
<http://dbpedia.org/resource/Poaceae>; dbpedia-
owl:kingdom
<http://dbpedia.org/resource/Plant>; dbpedia-
owl:thumbnail
?plant_resource_field_plant_thumbnail;
rdfs:label ?plant_resource_field_plant_label;
dbpedia-owl:abstract
?plant_resource_field_plant_abstract.
FILTER
langMatches(lang(?plant_resource_field_plant_ab
stract), "EN")
FILTER
langMatches(lang(?plant_resource_field_plant_la
bel), "EN")
}
```

Fig.3. Sample SPARQL query

Customization of Drupal RDFa annotations

RDFa annotations made automatically are not always sufficient. Drupal is an open source tool and give us a possibility to define custom module in PHP. For example, we would like to add elements of our website to be used by the Google's rich snippets. Snippets are the few lines of text that appear under Google search result. If Google understands the semantic description in a particular page, it can create information intended to help users to recognize if this site is relevant to their search. One of the markups, acceptable by Google search engine, is RDFa. Google supports rich snippets only for most searched types of content: Reviews, People, Products, Businesses and organizations, Recipes, Events and Music. Unfortunately among these topics there are no agricultural issues. However, in the project are also important personal data of implementers and it was decided to prepare the RDFa data for participating workers.

A good starting point to create one's own module is the analysis of the Drupal's RDF mapping API [2, 16]. This API consists of few functions and methods: `hook_rdf_mapping`, `hook_rdf_namespaces`, `rdf_get_namespaces`, `rdf_mapping_delete`, `rdf_mapping_load`, `rdf_mapping_save`, `rdf_rdfa_attributes`, `theme_rdf_metadata`, `rdf_rdf_namespaces` `theme_rdf_template_variable_wrapper` and others. For example the method `hook_rdf_mapping` allows modules to define RDF mappings for field bundles. Modules defining their own field bundles can specify which RDF semantics should be used to annotate these bundles. These mappings are then used for automatic RDFa output in the HTML code. While the method `hook_rdf_namespaces` allows modules to define namespaces for RDF mappings. Core vocabularies mentioned earlier are defined in `rdf_rdf_namespaces`. However, if a module implements `hook_rdf_mapping` and uses a prefix that is not defined in `rdf_rdf_namespaces`, this hook should be used to define the new namespace prefix.

In our case additional dictionary used by Google's rich snippets - `http://rdf.data-vocabulary.org/` - was added (see `person_rdf_namespaces()` in Fig.4). At the same time, the

new nodes and their mapping to elements in the dictionary were defined (see `person_rdf_mapping()` in Fig.4). Due to this module we received a page that provides additional information after the search result in Google. Testing is possible by using the Rich Snippets Testing Tool.

```
/* Implements hook_rdf_namespaces() */
function person_rdf_namespaces() {
return array(
// Google's custom vocabularies.
'v' => 'http://rdf.data-vocabulary.org/#',
);}
/* Implements hook_rdf_mapping() */
function person_rdf_mapping() {
return array(
'person' => array(
'type' => 'node',
'bundle' => 'person',
'mapping' => array(
'rdfstype' => array('v:Person'),
'fn' => array(
'predicates' => array('v:name'),
),
),
...
}
```

Fig.4. Implementation of methods

Conclusions

Experience gained during the project shows that the CMS system Drupal has proven to be a convenient tool for the rapid implementation of the web portal in a knowledge management system. Drupal 7 can import resources in RDF format which means that for example all resources of Linked Open Data are available to use. Sharing the data at the semantic level is slightly less convenient. The presented architecture is the basis for the further development of the system. Drupal together with the RDF module allow for the development of the system according to the standards proposed in the Semantic Web initiative. Based on the portal we can continue the iterative development of the system, together with a group of users – practitioners in plant production. First we plan to prepare a repository, an editor for semantic description of repository items and tools for searching and indexing the web pages. In the longer term, modules for visualization and application will be prepared.

Scientific work financed from funds for science in the years 2010-2011 as research project No. N N310 038538.

REFERENCES

- [1] Ackoff R., From Data to Wisdom, *Journal of Applied Systems Analysis* 16 (1989), 3–9
- [2] Clark L., The Semantic Web, Linked Data and Drupal, Part 1: Expose your data using RDF. *IBM DeveloperWorks* (2011)
- [3] Clark L., Corlosquet S., The Semantic Web, Linked Data and Drupal, Part 2: Combine linked datasets with Drupal 7 and SPARQL Views. *IBM DeveloperWorks* (2011)
- [4] Davenport, T. H., Prusak, L., Working knowledge: How organizations manage what they know. *Boston: Harvard Business School Press* (1998)
- [5] Genesereth M., Dar Z., Knowledge Management Technological Literature Review, *White Paper Media X at Stanford University* (2011),
- [6] Gołuchowski J., Technologie informatyczne w systemach zarządzania wiedzą w organizacji, *AE Katowice* (2007), wyd. 2
- [7] Karwowski W., Ontologies and Agricultural Information Management Standards. *Information systems in management* VI, ed. P. Jałowicki & A. Orłowski, WULS Press, Warszawa (2010)
- [8] Karwowski W., Kwasowicz W., Orłowski A., Hurtownie danych jako ważny element systemów zarządzania wiedzą, *Zarządzanie wiedzą w agrobiznesie w warunkach*

- polskiego członkostwa w Unii Europejskiej, redakcja naukowa: M. Adamowicz, Wydawnictwo SGGW, Warszawa, 2005, 334-347
- [9] Karwowski W., Kwasowiec W., Orłowski A., Różański J., Rusek M., Techniki i narzędzia eksploracji danych, *Metody Ilościowe w Badaniach Ekonomicznych VI*, redakcja naukowa: A. Orłowski, Wydawnictwo SGGW, Warszawa, (2006) 128-138
- [10] Karwowski W., Mazur P., Orłowski A., System "PAKD", *Polskie Stowarzyszenie Zarządzania Wiedzą: Studia i Materiały*, Nr 13, PSZW, Bydgoszcz, (2008) 74-81
- [11] Karwowski W., Orłowski A., New trends in the Internet applications. Practical example of Web 2.0 service, *Studies & Proceedings of Polish Association for Knowledge Management*, Nr 24, PSZW, Bydgoszcz, 2009
- [12] Karwowski, W., Orłowski A., Strategie tworzenia systemów zarządzania wiedzą. *Pomiary Automatyka Kontrola* (2011) Vol. 57, nr 11, 1384-1387
- [13] Liao, S. H., Knowledge management technologies and applications—Literature review from 1995 to 2002. *Expert Systems with Applications* (2003) 25, 59-71
- [14] oxforddictionaries.com
- [15] www.agridrupal.org/
- [16] api.drupal.org/api/drupal/modules%21rdf%21rdf.module/7

Authors: dr Waldemar Karwowski, Szkoła Główna Gospodarstwa Wiejskiego w Warszawie, Katedra Informatyki, ul. Nowoursynowska 159, 02-776 Warszawa, E-mail: waldemar_karwowski@sggw.pl; dr hab. Arkadiusz Orłowski, Szkoła Główna Gospodarstwa Wiejskiego w Warszawie, Katedra Informatyki, ul. Nowoursynowska 159, 02-776 Warszawa, E-mail: arkadiusz_orlowski@sggw.pl.