

ZASTOSOWANIE CMS DRUPAL W TWORZENIU SYSTEMU ZARZĄDZANIE WIEDZĄ W PRODUKCJI ROŚLINNEJ*

Waldemar Karwowski

Katedra Informatyki, Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

Streszczenie. Do podejmowania decyzji ważna jest wiedza rozumiana jako umiejętność użytecznego wykorzystania informacji. Dotyczy to także zagadnień z dziedziny rolnictwa w szczególności szeroko pojętej produkcji roślinnej. Obecnie wiele potrzebnych informacji można odnaleźć w Internecie, także globalna sieć wykorzystywana jest do publikacji rezultatów. Aby można było efektywnie wykorzystywać informacje w systemie zarządzania wiedzą, muszą być one precyzyjnie opisane. Sieć semantyczna a w szczególności idea Linked Data stwarza taką możliwość. Realizując projekt „Narzędzia zarządzania wiedzą w produkcji roślinnej” podjęto próbę wykorzystania systemu zarządzania treścią Drupal 7 do stworzenia prototypowego portalu będącego częścią systemu wspierającego zarządzanie wiedzą w produkcji roślinnej. Omówiono funkcjonalność modułu RDF Drupala, oceniono praktyczną ich przydatność do realizacji zadania oraz możliwości rozszerzenia.

Słowa kluczowe: produkcja roślinna, integrowanie źródeł informacji, sieć semantyczna, zarządzanie wiedzą.

Wstęp

Do podejmowania decyzji ważna jest wiedza i umiejętność użytecznego jej wykorzystania. Dotyczy to wykorzystania dostępnych zasobów danych, wyników badań czy opisanych doświadczeń innych osób czy instytucji. Rosnąca ilość informacji z jednej strony a z drugiej coraz większe możliwości technologiczne spowodowały, że systemy informatyczne stały się niezbędne w procesach pozyskiwania, przechowywania, przetwarzania czy udostępniania wiedzy. Coraz szersze wykorzystanie systemów informatycznych ma zastosowanie także w rolnictwie na przykład w zagadnieniach związanych z produkcją roślinną. Prowadzone są rozmaite badania dotyczące takich systemów. Przykładem jest system informatyczny wspomagający gromadzenie i analizę danych pozyskiwanych w procesie kompostowania [Promiński i in. 2011], innym ciekawym przykładem jest system pozyski-

* Praca zrealizowana w ramach projektu badawczego N N310 038538.

wania danych o geometrii produktów rolniczych na przykładzie ziarniaka kukurydzy [Weres 2010]. Jednakże systemy stacjonarne, dostępne lokalnie, nie zawsze są wystarczające. Pojawia się potrzeba aplikacji w których funkcjonalność i dane dostępne są dla wielu użytkowników jednocześnie. Systemy takie na ogół gromadzą i przechowują informacje centralnie i udostępniają interfejs w postaci stron WWW wykorzystując architekturę klient-serwer. Badania nad wykorzystaniem tego typu rozwiązań do zagadnień związanych z produkcją roślinną również są prowadzone. Interesującym przykładem jest system wspierający ochronę upraw rzepaku [Weres i in. 2010]. Rozwijany też jest Krajowy System Wspomagania Decyzji w Produkcji Roślinnej o bardzo rozbudowanej funkcjonalności. System ten integruje informacje pochodzące z różnych źródeł między innymi dane pogodowe [Zaliwski 2010]. Integracja informacji z różnych źródeł staje się coraz ważniejsza. Ciekawa koncepcja semantycznej integracji zasobów sieciowych w celu zbudowania elastycznego systemu sieciowych usług rekomendacyjnych przedstawiona jest w pracy [Marciniak 2010]. Obecnie do publikacji rezultatów badań i wyników zrealizowanych projektów wykorzystywany jest Internet i wiele potrzebnych informacji można odnaleźć w globalnej sieci. Jednakże integracja informacji pochodzących z Internetu napotyka na problemy związane z precyzyjnym ich opisem oraz olbrzymią ilością danych. Rozwiązaniem może być opisanie zasobów internetowych na poziomie semantycznym.

Celem niniejszej pracy było zbadanie jak wykorzystać istniejące obecnie dane opisane semantycznie dostępne w sieci jako źródło danych dla systemu informatycznego. Ponadto jak opisywać i udostępniać dane na poziomie semantycznym. W szczególności podjęto próbę wykorzystania systemu zarządzania treścią Drupal 7 do stworzenia fragmentu prototypowego systemu wspierającego zarządzanie wiedzą w produkcji roślinnej.

Semantyczny opis danych w Internecie

Język HTML i przeglądarki internetowe służą do prezentacji w formie możliwej do czytania przez człowieka, problem pojawia się gdy chcemy efektywnie przejrzeć tysiące stron. Pomocą są wyszukiwarki internetowe ale jakość wyników wyszukiwania jest niestety niezadawalająca w przypadku bardziej skomplikowanych zapytań. Aby rozwiązać ten problem twórca standardu WWW Tim Berners-Lee już wiele lat temu przedstawił wizję Semantic Web [Strona Semantic Web], sieci w której zasoby byłyby opisane w sposób zrozumiały także przez maszyny co w szczególności poprawiło by jakość wyszukiwania. Realizacja Semantic Web wymaga aby dane były opisane przez meta-dane, ponadto takie meta-dane powinny być stworzone w formie standardu akceptowanego przez wszystkich zainteresowanych. Znaczniki <meta> w języku HTML niestety nie mogą pełnić tej roli, opisy powinny być bardziej sformalizowane i wykorzystywać słowniki w formie tezauryusa lub ontologii. Konsorcjum W3C opracowało standardy będące wsparciem dla idei Semantic Web: RDF (*Resource Description Framework*), RDF Schema, OWL (*Web Ontology Language*), które bazują przede wszystkim na języku XML (choć idea RDF opisu wiedzy w postaci trójek *subject-predicate-object* jest ogólna i nie zależy od formy zapisu) [Karwowski 2010]. Do tworzenia zapytań powstał język SPARQL. Przedstawienie zasobów w formacie RDF okazało się dość trudne w szczególności sposób ich powiązania z plikami HTML. W efekcie niewiele z tego co oferuje Internet zostało przedstawione w ten sposób. Aby dołączyć semantyczne adnotacje do znaczników HTML konsorcjum

W3C opracowało standard RDFa (*RDF in attributes*) udostępniający zestaw atrybutów umożliwiający realizację tego zadania. Niezależnie, równolegle do prac W3C, opracowano zestaw tzw. mikro formatów (*microformats*) zdefiniowanych w XML umożliwiających semantyczny opis niewielkich fragmentów stron HTML także przy wykorzystaniu atrybutów znaczników. Wreszcie w powstającym, początkowo niezależnie od konsorcjum W3C, standardzie HTML5 przewidziano mikrodane (*microdata*) zawierające semantykę również przy wykorzystaniu atrybutów znaczników. Mikro dane opracowano głównie pod kątem wyszukiwarek internetowych z myślą o wyszukiwaniu najpopularniejszej tematyki (filmy, osoby, przepisy kulinarne itp.) i takie opracowano słowniki. Z kolei Tim Berners-Lee widząc problemy praktycznej realizacji Semantic Web zaproponował sposób publikacji zasobów semantycznych nazwany Linked Data. Opiera się on na następujących zasadach:

- Do identyfikowania obiektów należy używać tylko identyfikatorów HTTP URI, które mogą być rozumiane także przez maszyny.
- Do opisu obiektów należy używać standardu RDF/XML lub innych o których wiemy, że będą rozumiane przez użytkowników.
- W opisach obiektów należy używać odniesień do innych opisanych semantycznie zasobów co ułatwi odnajdywanie pokrewnych informacji w sieci WWW.

W ciągu ostatnich lat podjęto próby realizacji tej idei - najbardziej znana to Linked Open Data (LOD) [Projekt Linking Open Data]. Centralnym elementem tego projektu jest DBPedia [Strona DBPedia], która zawiera zestrukturalizowane dane z Wikipedii w formacie RDF. Nadal sieć LOD zawiera tylko ułamek zasobów internetowych ale stale rośnie. Wiele danych dotyczących rolnictwa można znaleźć w DBPedia, do sieci LOD włączony jest także AGROVOC, dostępne są też dane z Gene Ontology. Na ogół zasoby sieci LOD oferują możliwość zadawania zapytań poprzez tzw. SPARQL Endpoints co pozwala na integrację zasobów przez niezależne aplikacje. Należy dodać, że w 7 Programie Ramowym idea ta realizowana jest w projekcie LOD2 [Projekt LOD2].

Drupal jako element systemu zarządzania wiedzą

W Katedrze Informatyki SGGW realizowany jest projekt „Narzędzia zarządzania wiedzą w produkcji roślinnej”. W ramach projektu prowadzona jest między innymi analiza jak wyszukiwać oraz integrować dane i informacje pochodzące z zasobów internetowych o dość swobodnie określonej strukturze. Badania obejmują także działanie w drugą stronę a mianowicie jak udostępniać wiedzę związaną z produkcją roślinną w Internecie oraz w jaki sposób tworzyć narzędzia informatyczne pozwalające udostępniać tę wiedzę. Zaprojektowano system, którego założenia, metodyka projektowania i architektura opisana jest w [Karwowski, Orłowski 2011]. Główne elementy systemu to: portal użytkowników, repozytorium wiedzy, prototypowa ontologia i narzędzia zarządzania nią, komponent opisywania semantycznego zasobów w repozytorium, komponent realizujący wyszukiwanie, analizę stron WWW i ich indeksowanie w repozytorium. Ponadto przewidziano komponent do wizualizacji zasobów repozytorium oraz moduły realizujące wsparcie decyzji wraz z edytorem reguł wnioskowania. Do tworzenia portalu zdecydowano wykorzystać gotowy system klasy CMS. Najpopularniejsze dostępne bezpłatnie (open source) systemy tej klasy to Joomla!, Wordpress i Drupal. Wszystkie trzy napisane są w języku PHP i korzystają w pierwszej kolejności z bazy MySQL. Wybrano system Drupal, który ma obecnie ponad

175 000 instalacji w sieci WWW. Typowa instalacja Drupala składa się z jądra i szeregu konfigurowalnych modułów instalowanych w razie potrzeby. O wyborze Drupal 7 zdecydowały moduły do obsługi RDF, których nie ma w innych systemach tej klasy [Clark 2011; Clark, Corlosquet 2011]. Po zainstalowaniu odpowiednich modułów przede wszystkim możliwy jest import słowników w formacie RDF i XML. Podstawowe moduły importowane są automatycznie, są to między innymi ontologie DBPedia, Dublin Core, Friend of a Friend czy Simple Knowledge Organization System. Badania wykazały, że zasadniczo prosty jest import innych słowników wybranych przez administratora Drupala, w szczególności słownika AGROVOC w formacie SKOS. Każda zawartość portalu może być opisana przy użyciu pojęć z zaimportowanych słowników, możemy zdefiniować typ RDF, predykat RDF oraz typy danych. Zawartość generowanych stron w formacie HTML zostaje opisana poprzez dołączane adnotacje RDFa (rys. 1).

```
<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>
  </h2>
  ...
```

Rys. 1. Przykład opisanego węzła w systemie Drupal

Fig. 1. Example of described node in Drupal CMS

Cały ten proces wymaga jedynie przemyślanego działania administratora portalu, funkcjonalność dostarczona jest automatycznie i nie trzeba pisać kodu w języku PHP. Administrator ma do dyspozycji przyjazny interfejs graficzny (rys. 2). Możliwe jest opisywanie tylko bloków czy typów zawartości stron generowanych w Drupalu. Jest to niestety duże ograniczenie dlatego, że nie możemy opisać fragmentu bloku na przykład pojedynczego wyrazu. Co prawda w Drupalu możemy swobodnie definiować własne bloki ale chcąc zejść na poziom pojedynczego wyrazu należałoby dokonać całkowitej granulacji tekstu co z kolei byłoby bardzo uciążliwe przy wprowadzaniu danych.

Drupal jako system z otwartym kodem umożliwia dopisywanie modułów, ponadto całe API modułu RDF jest dostępne i można z niego korzystać [Dokumentacja modułu RDF Drupal 7]. Dopisano moduły, które pozwalają rozszerzyć funkcjonalność tak aby można było opisywać zawartość w sposób możliwy do wykorzystania przez popularną wyszukiwarkę Google. Wyszukiwarka ta umieszcza w wynikach wyszukiwania tzw. Google's rich snippets rozszerzone informacje o zawartości strony. Testy modułów wypadły pomyślnie, problemem jest natomiast to, że wyszukiwarka ogranicza się tylko do kilku dziedzin z których najbliższą produkcji roślinnej są przepisy kulinarne.

Kolejnym krokiem było zbadanie możliwości integracji danych z repozytoriów LOD. Drupal Views to jeden z modułów, który jest używany do tworzenia listy zawartości w różnych formatach ze źródła danych, najczęściej bazy relacyjnej. W obecnej wersji widoków można podłączyć dowolne dane źródłowe i ściągnąć je z dowolnego miejsca. Możli-

wość tę wykorzystano do utworzenia widoków SPARQL w module RDF. Dzięki temu na podstawie zapytań SPARQL wyświetlane są wyniki w taki sposób, jak z lokalnej bazy danych. W ramach sprawdzania przydatności Drupala zbadano rejestrację SPARQL endpoints. Przebiega ona również bez konieczności pisania kodu. Następnie zdefiniowano widoki SPARQL dla DBPedia (rys. 3) oraz dla AGROVOC. Wygodny interfejs graficzny umożliwia zarówno zdefiniowanie zapytania jak i utworzenie adnotacji RDFa do definiowanych pól.

Rys. 2. Proces opisywania elementów portalu
 Fig. 2. Process of describing Portal elements

Rys. 3. Definiowanie Widoku SPARQL (fragment)
 Fig. 3. Defining of SPARQL View (fragment)

Podsumowanie

CMS Drupal okazał się wygodnym narzędziem do szybkiej implementacji portalu użytkowników systemu. W zakresie zarządzania wiedzą Drupal w wersji 7 umożliwia import zasobów w zapisanych w formacie RDF jeżeli udostępniają one SPARQL endpoint. Nieco gorzej wypada udostępnianie danych na zewnątrz w formacie RDFa jednakże próby z tworzeniem modułów w języku PHP dają nadzieję na opracowanie narzędzia pozwalającego opisać pojedyncze wyrazy. Prace w tym kierunku są prowadzone. W celu udostępnienia informacji bezpośrednio w formacie RDF należałoby by jednak wykorzystać inne narzędzie. Jednakże Drupal okazał się dobrym rozwiązaniem i obecnie wokół portalu realizowane będą następne moduły systemu zarządzania wiedzą.

Bibliografia

- Clark L.** (2011): The Semantic Web, Linked Data and Drupal, Part 1: Expose your data using RDF, IBM DeveloperWorks [on-line], [dostęp 30.06.2012], Dostępny w Internecie: <http://www.ibm.com/developerworks/web/library/wa-rdf/>.
- Clark L., Corlosquet S.** (2011): The Semantic Web, Linked Data and Drupal, Part 2: Combine linked datasets with Drupal 7 and SPARQL Views, IBM DeveloperWorks [on-line], [dostęp 30.06.2012], Dostępny w Internecie: <http://www.ibm.com/developerworks/library/wa-datasets/>.
- Karwowski W.** (2010): Ontologies and Agricultural Information Management Standards. Information systems in management VI, ed. P. Jałowiecki & A. Orłowski, WULS Press, Warszawa, 49-56.
- Karwowski, W., Orłowski A.** (2011): Strategie tworzenia systemów zarządzania wiedzą. Pomiar Automatyka Kontrola, Vol. 57, 11, 1384-1387.
- Marciniak A.** (2010): Koncepcja elastycznego systemu sieciowych usług rekomendacyjnych. Inżynieria Rolnicza, 5(123), 159-164.
- Zaliwski A.S.** (2010): Pozyskiwanie i wykorzystanie danych pogodowych w krajowym systemie wspomaganie decyzji w produkcji roślinnej. Inżynieria Rolnicza, 5(123), 311-317.
- Weres J., Mueller W., Kozłowski R., Kluza T.** (2010): Advanced technologies in developing Web-based decision support systems for agriculture, 6 pp. Proc. of the CIGR XVIIth World Congress and the 8th World Congress on Computers in Agriculture, 13-17 June, 2010, Québec City, Canada. Quebec: Agriculture and Agri-Food Canada.
- Weres J.** (2010): Informatyczny system pozyskiwania danych o geometrii produktów rolniczych na przykładzie ziarniaka kukurydzy. Inżynieria Rolnicza, 7(125), 229-236.
- Promiński P., Mueller W., Dach J., Kujawa S.** (2011): System informatyczny wspomagający gromadzenie i analizę danych pozyskiwanych w procesie kompostowania. Inżynieria Rolnicza, 6(131), 165-171.
- Dokumentacja modułu RDF Drupala 7 [on-line], [dostęp 30.06.2012], Dostępna w Internecie: <http://api.drupal.org/api/drupal/modules%21rdf%21rdf.module/7>.
- Projekt Linking Open Data [on-line], [dostęp 30.06.2012], Dostępny w Internecie: <http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>.
- Projekt LOD2 [on-line], [dostęp 30.06.2012], Dostępny w Internecie: <http://lod2.eu>
- Strona DBPedia [on-line], [dostęp 30.06.2012], Dostępna w Internecie: <http://dbpedia.org/About>.
- Strona Semantic Web [on-line], [dostęp 30.06.2012], Dostępna w Internecie: <http://www.w3.org/2001/sw/>

USING DRUPAL CMS FOR IMPLEMENTATION OF KNOWLEDGE MANAGEMENT SYSTEM IN PLANT PRODUCTION

Abstract. Knowledge understood as the ability to apply the information in useful way, is significant to make proper decisions. This also applies to agricultural issues, in particular, wider plant production. Currently, a lot of information can be found on the Internet. The global network is used to publish the results. The information must be precisely described to be used effectively in the knowledge management system. The Semantic Web and in particular the idea of Linked Data offers such a possibility. During implementation of the project “Knowledge management tools in plant production” the content management system Drupal 7 has been used to create a prototype portal which is a part of the supporting knowledge management system in plant production. The functionality of Drupal RDF module is presented and its practical usefulness for the task is rated as well as the possibility of its extension is discussed.

Key words: plant production, integration of information sources, semantic web, knowledge management

Adres do korespondencji:

Waldemar Karwowski; e-mail: waldemar_karwowski@sggw.pl
Katedra Informatyki
Szkoła Główna Gospodarstwa Wiejskiego
ul. Nowoursynowska 166
02-787 Warszawa