

ONTOLOGY-BASED CONTENT MANAGEMENT SYSTEM

Ondřej Čekan

Master Degree Programme (2), FIT BUT

E-mail: xcekan00@stud.fit.vutbr.cz

Supervised by: Radek Burget

E-mail: burgetr@fit.vutbr.cz

Abstract: This paper deals with a content management system based on semantic technologies and ontology languages. It describes the semantic web technology and its concepts and the ways of publishing semantically annotated content on the web. Another part of this work deals with the description and design of the implemented system.

Keywords: Content management system, semantic web, semantic technology, ontology, database, XML, HTML, XHTML, RDF, RDFa, OWL, GRDDL, MVC, Sesame, SPARQL.

1. ÚVOD

Internet je stále rozšířenějším médiem umožňující rychlý přístup k řadě informací z různých oblastí zájmů a to prakticky odkudkoliv. Není proto divu, že každý den přibývá enormní množství nových stránek, článků, dokumentů a jiných informačních položek. Nastává ale problém v této změti informací nalézt relativní odpověď. Současné používané technologie v prostředí internetu se zaměřují na formátování textu a vizuálního vzhledu stránky, ale nic už neříkají o významu jednotlivých prvků na stránce. Tento způsob je nevhodný ke strojovému zpracování, jelikož nevyjadřuje sémantiku. Řešením by měl být tzv. sémantický web, který obsahuje data včetně jejich sémantiky. Sémantický web předpokládá provázanost informací s cílem jejich pochopení. Této roli se věnují tzv. ontologie.

Cílem práce je za pomoci sémantického webu a sémantických technologií vytvořit systém, který zpracuje libovolnou ontologii a umožní publikování anotovaných informací o konceptech této ontologie na WWW.

2. SÉMANTICKÝ WEB A ONTOLOGIE

Samotná data na webových stránkách nemají definovaný význam. Myšlenka sémantického webu spočívá v přidání dodatečných dat tzv. metadat k současné reprezentaci informací na webu. [1] Tímto by bylo možno přistupovat k jakémukoliv zdroji (dokument, obrázek, video apod.) na internetu jednotně a provádět dotazy nad jeho obsahem. Výhodou je přesnost výsledku dotazů. Data se rovněž nemusí nijak analyzovat a transformovat, protože sémantika je již v nich obsažena. Sémantický web staví především na technologiích RDF a OWL.

Pojem ontologie představuje v informatice formalizovaný popis metadat. [2] Popisuje koncepty a vztahy mezi nimi. Ontologie se používají v procesu vývoje znalostních aplikací. Ontologie by se dala chápat rovněž jako slovník, který uchovává a dále poskytuje své znalosti o určité zkoumané oblasti.

3. PUBLIKOVÁNÍ SÉMANTICKY ANOTOVANÉHO OBSAHU NA WWW

Pro vyjádření sémanticky anotovaného obsahu na WWW se používají tři základní přístupy, které rozšiřují standardní HTML či XHTML jazyk za pomoci existujících nebo speciálních dodefinovaných značek a atributů. [3]

První možností je využití tzv. mikroformátů. [4] Mikroformáty umožňují vkládat strojově čitelné informace přímo do atributů ke standardním značkám. Pro tyto účely se využívají atributy *class* a *rel*, popřípadě *rev*.

Druhá metoda je označována jako RDFa [5] a umožňuje zobrazit RDF (vychází ze standardu XML) dokument pomocí XHTML jazyka. Tento způsob je podobný mikroformátům, ale oproti nim přidává do stránky speciální nestandardní značky a atributy (slovník termů). Z tohoto důvodu se stane výsledný dokument nevalidní.

Třetím přístupem pro vyjádření sémantiky je GRDDL. [6] Jedná se o mechanismus, který pomocí speciálního algoritmu definovaného v XSLT dokáže převést XML či XHTML dokument do RDF sémantického formátu. U tohoto přístupu je zajištěna validita, jelikož dokument obsahuje atribut se speciálním prostorem jmen.

Náš systém bude využívat RDFa, jelikož se jedná o nejrozšířenější variantu, která má větší vyjadřovací sílu než použití mikroformátů a zároveň je implementačně jednodušší než GRDDL. Použití atributů nad rámec standardu HTML respektive XHTML nepředstavuje pro současné webové prohlížeče problém.

4. SYSTÉM PRO SPRÁVU OBSAHU S PODPOROU ONTOLOGIÍ

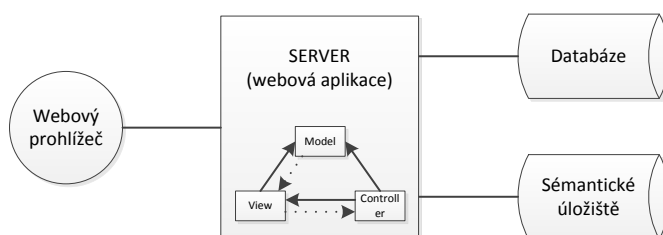
Navrhovaný systém se oproti běžným CMS systémům, které umožňují publikovat pouze obsah, zaměřuje na publikování anotovaných dat. Anotovaná data mají značnou výhodu v budoucím zpracování, jelikož obsahují význam jednotlivých dat na stránce a tak mohou být dále jednoduše zpracována strojově.

Tento systém umožní uživateli registrovat se a následně přihlásit se do systému uživatelskými údaji. Po přihlášení je uživateli nabídnuta možnost importu nové ontologie, nebo zvolení uložené (dříve importované) ontologie. Po načtení zvolené ontologie má uživatel možnost s touto ontologií operovat. Ontologii je rovněž nutno vhodným způsobem prezentovat uživateli.

Systém by měl obsahovat alespoň dvě uživatelské role a to roli správce a běžného uživatele, kde správce má navíc možnost spravovat stránky oproti běžnému uživateli.

4.1. NÁVRH ŘEŠENÍ

Systém lze rozdělit na čtyři části, které mezi sebou vzájemně spolupracují pro vykonání požadované operace. Rozdělení systému znázorňuje obrázek 1.



Obrázek 1: Obecné schéma systému.

Komunikace s uživatelem bude probíhat prostřednictvím bezstavového protokolu HTTP přes webový prohlížeč. Tento způsob je nejrozšířenější, není závislý na platformě, není potřeba distribuovat systém mezi uživatele, stačí mít pouze přístup k internetu.

Jelikož se jedná o webovou aplikaci, architektura systému bude vycházet z návrhového vzoru MVC, který se v prostředí webu nejvíce osvědčil. Základní princip je založen na načtení a porozumění libovolné ontologii reprezentované jazykem OWL. OWL vychází ze standardu XML, načítá-

ni bude tedy probíhat rozložením vstupního dokumentu na jednotlivé značky a atributy. Porozumění jednotlivým značkám bude vycházet ze standardu OWL. Ten vychází ze tří základních komponent, což jsou jedinci, vlastnosti a třídy. [2] Porozumění ontologie a její následné zpracování není vůbec jednoduché, jelikož se jedná o rozsáhlý standard, který je stále ještě ve vývoji. Načtená ontologie bude uložena do sémantického úložiště, nad kterým lze pomocí speciálních jazyků podobných SQL provádět dotazy. Sémantické úložiště Sesame [7] je de-facto standardem pro práci se sémantickými daty. Využívá dotazovací jazyk SPARQL. Pomocí zvolené ontologie systém prezentuje uživateli data anotovaně, jak bylo uvedeno metodou RDFa. Pokud například ontologie definuje třídu osoba a její vlastnosti, uživateli prezentujeme data o konkrétních osobách. Tyto data jsou ve stránce označeny pomocí RDFa a ukazují, jaká část představuje jméno, příjmení apod. Uživatel bude komunikovat prostřednictvím webových formulářů, které se pro aktuálně zpracovávanou ontologii automaticky vygenerují do struktury webové stránky. Upravenou ontologii bude možno exportovat a to zpětným prováděním operací, jako než při importu ontologie.

Oproti sémantickému úložišti, bude systém využívat i databázové úložiště, které bude uchovávat údaje potřebné pro chod webové aplikace. Pro tyto údaje je vhodné využít fixní schéma uložení dat, a proto bude použita relační databáze. Mezi informace, které se do databáze budou ukládat, patří především údaje o uživatelských účtech, údaje o ontologiích, základní nastavení stránek a podobně.

5. ZÁVĚR

Implementace tohoto systému přinese nesmírnou výhodu a to především v univerzálnosti použití. Systém může být použit prakticky na správu čehokoliv, v závislosti na právě používané ontologii. Není problém vést záznamy o studentech, předmětech, vyučujících a jejich vzájemných vztazích nebo například spravovat druhy pečiva včetně jejich surovin, vést záznam o tom, jaká surovina je kde použita a podobně, stačí pouze naimportovat příslušnou ontologii. Systém jistě získá v budoucnu obrovskou komunitu uživatelů, jelikož nalezne uplatnění v řadě oborů. Další výhodu představují anotovaná data pro strojové zpracování, jednoduchost použití a rovněž zálohování exportováním ontologie.

REFERENCE

- [1] DACONTA, Michael C, Leo Joseph OBRST a Kevin T SMITH. *The Semantic Web: a guide to the future of XML, Web services, and knowledge management*. Indianapolis, Ind.: Wiley Pub., c2003, 281 p. ISBN 04-714-3257-1.
- [2] Svátek, V.: *Ontologie a sémantický web* [online]. DATAKON, Brno, 2002 [cit. 2013-02-25]. Dostupné z: <http://nb.vse.cz/~svatek/onto-www.pdf>
- [3] BURGET, Radek. *Ontologie a sémantický web* [online]. 2010 [cit. 2013-02-25]. Dostupné z: <https://www.fit.vutbr.cz/study/courses/PIS/private/cviceni/semweb2010.pdf>
- [4] Mikroformát. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2001-2013 [cit. 2013-02-25]. Dostupné z: <http://cs.wikipedia.org/wiki/Mikroform%C3%A1t>
- [5] RDFa Core 1.1: Syntax and processing rules for embedding RDF through attributes. *World Wide Web Consortium (W3C)* [online]. 2012 [cit. 2013-02-25]. Dostupné z: <http://www.w3.org/TR/rdfa-syntax/>
- [6] Gleaning Resource Descriptions from Dialects of Languages (GRDDL). *World Wide Web Consortium (W3C)* [online]. 2008 [cit. 2013-02-25]. Dostupné z: <http://www.w3.org/TR/rdfa-shttp://www.w3.org/2004/01/rdxh/spec>
- [7] *OpenRDF.org: About* [online]. 1997-2012 [cit. 2013-02-27]. Dostupné z: <http://www.openrdf.org/about.jsp>