EKONOMICKÁ UNIVERZITA V BRATISLAVE
FAKULTA HOSPODÁRSKEJ INFORMATIKY

Evidenčné číslo: 17300/B/2011/2723917834

PROJEKTY SÉMANTICKÉHO WEBU

Bakalárska práca

2011 **Matej Hajko**

EKONOMICKÁ UNIVERZITA V BRATISLAVE
FAKULTA HOSPODÁRSKEJ INFORMATIKY

PROJEKTY SÉMANICKÉHO WEBU

Bakalárska práca

Študijný program: Hospodárska informatika

Študijný odbor: 9.2.10 Hospodárska informatika

Školiace pracovisko: Katedra aplikovanej informatiky

Školiteľ: RNDr. Eva Rakovská

Bratislava, 2011 **Matej Hajko**

Ekonomická univerzita v Bratislave
Fakulta hospodárskej informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Matej Hajko
Študijný program: Hospodárska informatika (Jednoodborové štúdium, bakalársky I. st., denná forma)
Študijný odbor: 9.2.10 Hospodárska informatika
Typ záverečnej práce: Bakalárska záverečná práca
Jazyk záverečnej práce: slovenský

Názov: Projekty sémantického webu.

Anotácia: Práca je teoretického charakteru. Vysvetľuje základné pojmy sémantického webu. Orientuje sa na oboznámenie sa s projektmi sémantického webu a ich využiteľnosťou v praxi. Uvádza niektoré Európske projekty postavené na sémantickom webe.

Vedúci: RNDr. Eva Rakovská
Katedra: KAI FHI – Katedra aplikovanej informatiky FHI
Vedúci katedry: doc. Ing. Gabriela Kristová, CSc.

Dátum zadania: 30. 09. 2010

Dátum schválenia: 19. 10. 2010

doc. Ing. Gabriela Kristová, CSc.
vedúci katedry

student

Vedúci

Čestné vyhlásenie

Čestne vyhlasujem, že záverečnú prácu som vypracoval samostatne a že som uviedol všetku použitú literatúru.

Dátum: **…….**

ABSTRAKT

HAJKO, Matej: *Projekty sémantického webu* – Ekonomická univerzita v Bratislave. Fakulta hospodárskej informatiky; Katedra aplikovanej informatiky. – Vedúci záverečnej práce: RNDr. Eva Rakovská, PhD. – Bratislava: FHI EU, 2011, 48s.

Cieľom záverečnej práce bolo poskytnúť základný náhľad na sémantický web a jeho technológie a bližšie opísať projekty, ktoré sú zamerané na podporu rozvoja sémantického webu.
Práca je rozdelená do troch kapitol. Obsahuje 3 obrázky a 2 tabuľky.
Prvá kapitola je venovaná súčasnému stavu v oblasti sémantického webu doma a v zahraničí so zameraním na v minulosti realizované projekty.
V ďalšej časti sa charakterizujú ciele a metodika práce, ako aj teória na pozadí sémantického webu.
Záverečná kapitola sa zaoberá aktuálne bežiacimi projektmi.
Výsledkom riešenia danej problematiky je porovnanie vybraných projektov a návrh ich praktického uplatnenia v ostrom nasadení.

Kľúčové slová:
sémantický web, projekt sémantického webu, ontológia, URI, RDF, schéma

ABSTRACT

HAJKO, Matej: *Semantic Web projects* – The University of Economics in Bratislava. Faculty of Economic Informatics; Department of Applied Informatics. – Thesis supervisor: RNDr. Eva Rakovská, PhD. – Bratislava: FHI EU, 2011, 48 pages.

Purpose of this thesis was to provide a basic insight into the semantic web and its technology and more to describe projects that are aimed at promoting the development of the Semantic Web.
The work is divided into three chapters. It contains 3 figures and 2 tables.
The first chapter is devoted to the current state of the Semantic Web at home and abroad, focusing on previously completed projects.
In the next chapter we describe the objectives and methodology of work and the theory behind the Semantic Web.
The final chapter deals with current projects running on.
The result of solving the issue is the comparison of selected projects and a proposal for their practical application in stark deployment.

Key words:
Semantic Web, Semantic Web project, ontology, URI, RDF, schema

Obsah

Úvod

Ekonomika v súčasnosti pozná okrem troch základných výrobných faktorov (pôda, práca, kapitál) aj štvrtý výrobný faktor, ktorým sú informácie. Spoločnosť sa vyvinula do stavu, keď bez efektívneho spracovania informácií majú organizácie problémy uplatniť sa na trhu. Jedným z hlavných médií, čo sa týka prenosu a spracovania informácií, je internet. Toto médium je organizované do formy celosvetovej siete, ktorá sa nazýva World Wide Web. V súčasnosti sa používa web druhej generácie, tzv. Web 2.0. V prostredí tohto webu pracujú ľudia, ktorí používajú počítače len ako platformu na prístup k potrebným informáciám alebo nástrojom. V súčasnosti sa pracuje na novej generácii webu, ktorá by mala zabezpečiť odstránenie nedostatkov Webu 2.0 (o nedostatkoch budeme hovoriť bližšie v kap. 1) a rozšírenie možností a použiteľnosti webu. Táto generácia je označovaná ako Web 3.0, ale známejšie je pomenovanie sémantický web. Myšlienka novej generácie webu je známa už pomerne dlho. Sformuloval ju v roku 1999 vtedajší riaditeľ W3C, Tim Berners-Lee. Jeho víziou bolo vytvorenie prostredia, v ktorom by mohli pracovať inteligentné agenty.

Pod pojmom sémantický web rozumieme virtuálny priestor, ktorý umožňuje strojom porozumieť významu informácií na webe. Rozširuje sieť stránok prepojených hyperlinkami (ktorým rozumejú iba ľudia) o strojovo čitateľné metadáta s informáciami o obsahu stránok a ako sú navzájom prepojené umožňujúc tak automatickým agentom pristupovať k webu inteligentnejšie a plniť úlohy, ktoré im zadajú používatelia. Výraz sémantický web sa v užšom slova zmysle viaže k formátom a technológiám umožňujúcim vyššie uvedené rozšírenie funkcionality webu. Medzi takéto technológie patria RDF, RDF/XML, N3, RDF Schema, OWL (všetky budú vysvetlené ďalej) a mnohé ďalšie, ktoré majú za úlohu poskytnúť formálne opisy konceptov, výrazov a vzťahov v rámci danej znalostnej domény.

Na začiatku práce sa pokúsime charakterizovať súčasný stav na poli sémantického webu. Zameriame sa na uskutočnené projekty, ktoré mali za cieľ vyvinúť alebo prakticky aplikovať niektoré z technológií sémantického webu. Bude nás zaujímať aj participácia Slovenska na týchto aktivitách. Mnohé zo spomenutých technológií už existujú a používajú sa v rôznych oblastiach, najmä v oblasti spracovania informácií z určitej domény. Týmto technológiám sa budeme venovať pri stručnom načrtnutí teórie za sémantickým webom. K projektom sémantického webu sa vrátime v závere práce. Tu sa ale sústredíme na stále bežiace projekty. Spomedzi uvedených projektov si vyberieme niekoľko predstaviteľov, ktoré si podrobnejšie rozoberieme a pokúsime sa pre ne navrhnúť praktické uplatnenie.

1. Súčasný stav skúmanej problematiky doma a v zahraničí

1.1 Web2.0

V súčasnosti sa vo svete používa Web 2.0. Je to vlastne ďalšie v súčasnosti masívne rozšírené médium, ktoré sa efektívne presadzuje popri rádiu a televízii. Jeho hlavnou črtou je interaktivita. Táto vlastnosť je vyzdvihovaná najmä v porovnaní s Webom 1.0, ktorý slúžil hlavne na získavanie informácií, ale neumožňoval priamu spätnú väzbu. Web 2.0 je veľmi často charakterizovaný participáciou, čo znamená, že užívatelia už nie len pasívne prijímajú, čo im je podávané, ale sa aj aktívne zúčastňujú na tvorbe obsahu webu. K tomu užívateľom web2.0 poskytuje rozhranie, software a aj úložné kapacity. Koncept participácie je ďalej podporovaný aj poskytnutím dát užívateľom a zároveň aj kontroly nad týmito dátami. Za základné črty Webu 2.0 môžeme považovať široké možnosti užívateľov, participáciu, dynamický obsah, metadáta, štandardy a rozšíriteľnosť v zmysle zvládania rastúcich nárokov na internet. K vlastnostiam Webu 2.0 môžu byť pridané aj otvorenosť, voľnosť a kolektívna inteligencia. Z týchto vlastností má mimoriadny význam voľnosť. Prejavuje sa nie len v tom, čo môžu ľudia vytvoriť, ale aj v tom, aké stránky ľudia chcú navštíviť. Pekným príkladom tohto je Wikipedia [2]. Experti jej dávajú slabé hodnotenia, ale stránka je aj napriek tomu veľmi populárna. Táto popularita je zabezpečená možno nie celkom kvalitou obsahu, ale stránka nie je spoplatnená a preto je navštevovaná. Princíp voľnosti ďalej spočíva v tom, že ľudia vytvárajúci obsah nie sú nikým obmedzovaní a obsah ich diel nie je ďalej upravovaný vyššou autoritou, ako je to napríklad v tlačených médiách.

V súčasnosti sa Web2.0 rozdeľuje na niekoľko oblastí, ktoré tvoria jeho kompletnú štruktúru a zároveň sa pomocou nich odlišuje od predchádzajúcich verzií (ktorými rozumieme Web 1.0 a jeho predchodcov)[1]. Týmito oblasťami sú internetové aplikácie, ktoré sú porovnateľné s klasickými desktopovými aplikáciami (k ich reprezentantom patria hlavne aplikácie vytvorené v jazykoch AJAX a Flash), architektúra orientovaná na služby, ktoré vyjadruje, že aplikácie neskrývajú svoju funkcionalitu, aby mohli byť použité integrovane s inými aplikáciami a poskytnúť tak podstatne komplexnejšie služby a oblasť nazývaná sociálny web, ktorá vyjadruje, že web sa snaží komunikovať s koncovými používateľmi a tým ich integrovať do siete. Tieto oblasti sa navzájom prekrývajú a ich funkcie sú zabezpečené pomocou mnohých vlastností a techník. Medzi najznámejšie by sme mohli uviesť vyhľadávanie na základe kľúčových slov, prepájanie obsahu pomocou odkazov (linkov), možnosť vytvárania a udržiavania obsahu stránok najmä na báze spolupráce väčších skupín, možnosť označiť časti obsahu pomocou značiek (tagov), ktoré sú zväčša jednoslovné popisy na zlepšenie vyhľadávania, prípadne vytváranie celých taxonómií, používanie softwaru, ktorý

beží na internete a robí tak z neho platformu pre aplikácie a schopnosť upozorniť ostatných používateľov na dôležité udalosti alebo zmeny v obsahu. Veľmi populárnou oblasťou sa stali blogy a rôzne wiki stránky, ktoré reprezentujú možnosť vytvárania obsahu alebo spoluúčasť pri vytváraní obsahu. Tvorba obsahu je v týchto prípadoch mierne odlišná, keď v prípade blogu jeden používateľ verejne vyjadruje svoje myšlienky a v prípade wiki veľké množstvo používateľov spoločne prispieva do jednej encyklopedickej formy prezentácie poznatkov alebo udalostí. Inou formou vytvárania obsahu sú sociálne siete, ktoré umožňujú zdieľať obrázky, videá a nadväzovať nové kontakty.

Odvetvím, ktoré vo veľkej miere využíva prostredie webu, je marketing. Pomocou rôznych nástrojov sa uplatňuje cielená reklama. Inou masívne využívanou technikou je spolupráca so zákazníkmi na zlepšovaní produktov a služieb. Vo svete výrobcovia dokonca vytvárajú svoje wiki stránky, do ktorých potom používatelia ich produktov prispievajú značnou časťou[4]. Spoločnosti využívajú sociálne siete, pomocou ktorých sa dostávajú bližšie k zákazníkom na skúmanie preferencií zákazníkov, na adresovanie reklamy alebo na zverejňovanie reklamných kampaní spolu s umiestňovaním videí na YouTube. Využívanie sociálnych sietí tiež zvyšuje konkurencieschopnosť malých firiem.

1.2 Nedostatky webu 2.0

Ako každá vec vo svete, aj Web 2.0 má svoje nedostatky. V súčasnosti nastupuje Web 3.0, ktorý sa nazýva aj sémantický web. Zameriava sa na nový prínos do oblasti internetu, ale aj odstránenie niektorých nedostatkov súčasného webu druhej generácie. Jednou z nevýhod Webu 2.0, ktorú uvádzajú viaceré pramene je problém zdieľania dát verzus copyright. Keďže každý môže vytvárať obsah webu, je jasné, že sa na web dostanú aj dáta, ktoré sú niečím majetkom. Najvypuklejší je tento problém v oblasti umenia, najmä film a hudba. Problém je v tom, že ľudia zdieľajú tieto dáta na webe, čo znamená, že ostatní sa k týmto dátam, resp. dielam môžu dostať bez zaplatenie za ne. Ako iste dobre vieme, takzvané „pirátstvo“ je veľmi rozšírené a je podporované širokou komunitou. S touto témou súvisí aj udalosť na stránke Digg, kde je demonštrovaná aj nevýhoda vyplývajúca z prenechania príliš veľkej časti kontroly používateľom ako aj problém pirátstva ako je uvedené na [9]. Ďalšou črtou, ktorá je ostro kritizovaná najmä v akademických kruhoch, je spoľahlivosť informácií dostupných na webe. Jedným z prípadov tohto typu je už vyššie spomenutá Wikipedia. Mnohí odborníci ju kritizujú za nespoľahlivosť informácií. Po príklad nemusíme chodiť ďaleko. Vo vysokoškolských, ale aj iných kruhoch je vnímané výrazne negatívne ak študent vo svojej práci uvedie ako jeden zo zdrojov Wikipediu. Je to tak preto, lebo pri písaní vedeckých prác je nutné, aby bol známy autor použitých zdrojov, čo nie je zabezpečené pri všetkých

stránkach. Táto skutočnosť vyplýva z možnosti participácie pri vytváraní obsahu webu, konkrétne Wikipedie. Toto ale nie je jediný prípad nespoľahlivosti webu. Možno ešte menej spoľahlivé sú informácie, ktoré zverejňujú používatelia na svojich osobných stránkach a blogoch. Toto nás privádza k ďalšej slabej stránke Webu 2.0 a tou je internetový vandalizmus. Tento jav sa dá opísať ako činnosť, pri ktorej niekto zámerne poškodí stránku, poprípade do nej vloží škodlivý kód. Ako príklad poslúži prípad pána Seigenthalera, ktorý je uvedený na [8]. Internetový vandalizmus je spojený s ďalšou slabou stránkou Webu 2.0, ktorou je anonymita. Tá poskytuje relatívnu ochranu ľuďom, ktorí dávajú na web škodlivý alebo nepravdivý obsah. V dôsledku toho, že na stránky môže prispievať ktokoľvek, je prakticky nemožné vystopovať pôvodcu škodlivého obsahu. S vandalizmom súvisia aj útoky hackerov na webové služby. Týmto je načrtnutá širšia problematika bezpečnosti webu. Odborníci z danej oblasti tvrdia, že síce sú rozšírené a podporované rôzne služby Webu 2.0, ale stránka bezpečnosti je značne zanedbaná, dokonca až do takej miery, že nie je odlišná od bezpečnosti používanej vo Webe 1.0, kde boli možnosti používateľov značne obmedzené. Bezpečnostná hrozba vzniká najmä pri dynamicky generovanom obsahu alebo odkazoch (príkladom môže byť RSS-feed), pri ktorých klasické nástroje ako firewally a antivírové programy nie sú schopné vykonávať svoju činnosť dostatočne efektívne. Pri tomto probléme je veľmi dôležitá identifikácia posielajúcej stránky a jej dôveryhodnosť. Riešením by mohol byť Web of Trust [29], čo je termín spájaný so sémantickým webom a bude podrobnejšie opísaný ďalej. Problematika bezpečnosti a príklad zraniteľnosti na stránke MySpace sú širšie prezentované na [7]. Je známy aj prípad keď útočník získal dôverné údaje o používateľoch Facebooku (čo s nimi urobil teraz nie je až také dôležité). Z toho vyplýva aj nutnosť venovať pozornosť otázkam súkromia. Podobné je aj zverejňovanie videí napríklad na YouTube. Každý môže nahrať na server video, ale ak bolo natočené bez súhlasu alebo dokonca vedomia zobrazovanej osoby, ide o porušenie súkromia. K nedostatkom webu, ktorý sa stal aj spoločenským problémom patrí aj závislosť. Síce je dosť ťažké predstaviť si závislosť na internete, ale závislosti na jeho rôznych prvkoch alebo možnostiach sú verejne diskutované (napríklad prístup ku konkrétnym stránkam, o online hrách už ani nehovoriac).

1.3 Realizované projekty

V rámci snahy o odstránenie týchto slabých stránok webu boli vytvorené mnohé projekty, ktoré sa venovali rôznym oblastiam uplatnenia technológií sémantického webu. Spoločným zmyslom týchto projektov, hlavne projektov pod záštitou Európskej únie, bolo priviesť sémantický web a jeho možnosti do širšieho povedomia. Európska únia sa zamerala na zavedenie spomínaných technológií najmä do oblasti priemyslu a výskumu, ale aj do iných

oblastí. Značná časť z ďalej uvedených projektov bola rozvinutá v rámci šiesteho rámcového programu EU. Tu si uvedieme iba niekoľko z nich.

Jedným z projektov šiesteho rámcového programu EU je projekt SEKT (Semantically-Enabled Knowledge Technologies alebo Semantic Knowledge Technologies). Tento projekt sa zameriaval na podporu znalostného manažmentu novej generácie. Cieľom projektu bolo uľahčiť všetky operácie súvisiace so znalostným manažmentom, aby sa stal činnosťou vykonávanou bez zbytočnej námahy. Realizácia tohto cieľa sa mala dosiahnuť skombinovaním troch technológií a vytvorením synergií medzi nimi. Tými technológiami sú Ontology and Metadata Technology (OMT)[11], Knowledge Discovery (KD)[11] a Human Language Technology (HLT)[11]. OMT technológia sa zameriava na vytvorenie a používanie metadát založených na ontológiách a tým vytvorenie strojovo spracovateľných dát, nad ktorými bude možné uvažovať, bude ich možné generovať a vyvíjať. To vyústi do zníženia objemu údržby, čím sa podporí pokročilé uvažovanie (strojové) a počítač bude schopný poskytnúť zmysluplné výsledky aj v prípade, že ontológia má v sebe konflikt. Technológia HLT má dva významné aspekty. Prvým je schopnosť sémanticky anotovať neformálne alebo neštruktúrované dáta. Tak sa dosiahne automatická alebo poloautomatická extrakcia metadát. Druhým aspektom je schopnosť generovať prirodzený jazyk na základe formálnych znalostí, čiže ontológií a metadát. Týmto sa podporí viacjazyčnosť. Zahrnuté sú všetky hlavné európske jazyky a aj niektoré ďalšie, od čínštiny po bulharčinu. KD technológia je úzko previazaná s HLT v oblasti generovania metadát a ontológií. KD poskytuje pracovníkom znalosti, ktoré práve potrebujú z veľkých objemov dát, čím je podobná HLT.

Ďalším zaujímavým projektom zo šiesteho rámcového programu je projekt DIP(Data, Information and Process Integration with Semantic Web Services)[12]. Jeho poslaním bolo rozvinúť a rozšíriť sémantický web a webové služby a skombinovať ich do novej technologickej infraštruktúry, nazývanej Semantic Web Services. Táto nová infraštruktúra sa mala uplatniť v oblastiach e-Work a e-Commerce [21]. Poslanie projektu sa delilo na viaceré kľúčové ciele. Prvým z týchto odvodených cieľov bolo aplikovať sémantický web do reality. Tento cieľ sa zakladal na strojovo spracovateľných sémanticky obohatených dátach, ktoré uľahčujú komunikáciu a kooperáciu. Druhým cieľom bolo skombinovať technológiu sémantického webu s webovými službami do podoby už spomenutých Semantic Web Services. DIP dúfal, že Semantic Web Services prinesú prelomovú aplikáciu pre sémantický web a zmenia nie len spracovanie informácií, ale aj spôsob, akým pristupujeme k výpočtovým kapacitám. Výsledkom malo byť poskytnutie novej infraštruktúry pre efektívnejší elektronický biznis a lepšiu spoluprácu. Tretím cieľom projektu bolo aplikovať vyvinuté

technológie v reálnych podmienkach. DIP vytváral riešenia na výzvy z reálneho sveta a tieto aplikoval v jednotlivých organizáciách alebo skupinách organizácií pracujúcich v klasických reťazcoch. DIP teda vyvinul praktické technológie použiteľné v eWork, eGoverment a eCommerce. Tieto technológie zahŕňajú inteligentný manažment informácií, integráciu podnikových aplikácií a dynamickú a rozumnú eCommerce. Všetky tieto ciele boli označené za veľmi ambiciózne. Na záverečnom hodnotení projektu bol projekt vyhodnotený ako úspešný s dobrým splnením stanovených cieľov bez väčších odchýlok.

Na projekt DIP nadväzuje projekt SUPER, ktorý je tiež vedený pod záštitou Európskej únie, konkrétne je financovaný opäť zo šiesteho rámcového programu. SUPER (skratka pre Semantics Utilised for Process Management within and between Enterprises)[18] nadväzuje aj na iné projekty Európskej únie, napríklad SEKT, SWWS a ďalšie. Projekt vychádza z neúspechu reinžinieringu podnikových procesov a adresuje nové smerovanie v tejto oblasti, čo je Business Process Management (BPM). Pojmom BPM sa dá rozumieť väčšie množstvo významov a viac sa o tejto problematike môžeme dočítať na stránke projektu. SUPER má dva hlavné ciele, ktoré sú: 1. posunúť kontrolu procesov od IT špecialistov k manažmentu a 2. preniesť BPM na novú úroveň komplexnosti. K dosiahnutiu týchto cieľov SUPER značne vychádza z výsledkov predchádzajúcich projektov, najmä DIPu. Snahou je skombinovať Semantic Web Services a BPM a vytvoriť tak novú technológiu, ktorá bude schopná opísať podnikové procesy a vertikálne ontológie orientované na telekomunikácie na podporu doménovo špecifickej anotácie. To by sa malo dosiahnuť poskytnutím sémantickej a kontextovej štruktúry, ktorá vie získavať, organizovať a používať znalosti z podnikových procesov dostupné vo firemnom softvéri a v nápadoch zamestnancov. Táto celková úloha projektu je transformovaná do viacerých vedeckých a technických podcieľov, ktoré sú uvedené na stránke projektu. V tabuľke ďalej sú uvedené viaceré projekty, z ktorých veľká väčšina bola vedená pod záštitou Európskej únie ako súčasť šiesteho rámcového programu. Samozrejme, tabuľka neposkytuje úplný výpočet všetkých uskutočnených projektov (ani projektov EU). V tabuľke nie sú uvedené ani projekty, ktoré boli uskutočnené pred rokom 2002 z dôvodu, že tieto projekty sa už nedajú zaradiť do skúmania súčasného stavu problematiky.

Projekt	Ciele	Web
Knowledge Web	podporiť proces osvojovania si a používania technológií sémantického webu	http://knowledgeweb.semanticweb.org/semanticportal/sewView/frames.html

SEKT (Semantic Knowledge Technologies)	vyvinúť technológie na podporu znalostného manažmentu novej generácie	http://www.sekt-project.com/
DIP Integrated Project	rozvoj sémantického webu a jeho technológií a zavedenie služieb sémantického webu	http://dip.semanticweb.org/index.html
aceMedia	objaviť a využiť znalosti v obsahu, uľahčiť prácu s obsahom, automatizovať anotáciu	http://www.acemedia.org/aceMedia/index.html
AIM@SHAPE	vyvinúť systém na spracovanie znalostí v multidimenzionálnych dátových objektoch	http://www.aim-at-shape.net/
COCOON	podpora zdravotnej starostlivosti vybudovaním znaloťami riadených a adaptívnych komunít vrámci európskych zdravotníckych systémov s použitím technológií sémantického webu	×
METOKIS	použitie technológií sémantického webu pre elektronickú publikáciu	http://metokis.salzburgresearch.at/
NeOn	zlepšiť používanie ontológií vo veľkých sémantických aplikáciách a distribuovaných organizáciách	http://www.neon-project.org/nw/Welcome_to_the_NeOn_Project
REWERSE	ustanoviť Európu (EU) za vedúcu silu v oblasti usudzovacích jazykov pre web	http://rewerse.net/
SUPER Integrated Project (nástupník DIP)	dostať Business Process Management z IT úrovne na business úroveň	http://www.ip-super.org/component/option,com_frontpage/Itemid,1/
SWAD - Europe	podporiť iniciatívu W3C ohľadom sémantického webu	http://www.w3.org/2001/sw/Europe/
WS2	zvýšiť povedomie o službách sémantického webu a dostať ich do oblastí výskumu a priemyslu	http://www.w3.org/2004/WS2/

Všetky uvedené projekty sú v súčasnosti skončené. Momentálne bežiacim projektom, ktorých je tiež veľké množstvo sa budeme venovať v kapitole Výsledky práce.

1.4 Realizované projekty v domácich podmienkach

Aj v našom domácom prostredí bolo uskutočnených niekoľko projektov. Tieto projekty boli tiež súčasťou šiesteho rámcového programu EU. Jedným z projektov, na ktorých sa

podieľali aj zástupcovia zo Slovenska bol projekt Access-eGov (Access to e-Government Services Employing Semantic Technologies). Na projekte sa podieľalo 11 partnerov z 5 krajín, pričom hlavým koordinátorom bola Technická Univerzita v Košiciach. „Hlavným cieľom projektu priblížiť administratívu k občanom a podnikateľským subjektom využitím Internetu, pričom celý systém je navrhovaný z hľadiska používateľa.“ [22] Detailnejší pohľad na ciele projektu je na stránke [23]. Projekt pracoval na zlepšení schopnosti spolupráce medzi už existujúcimi službami verejnej správy (elektronickými a tradičnými). Jeho hlavnými prínosmi bol „pohodlnejší a rýchlejší prístup k verejným službám, väčšia dostupnosť verejných služieb a zníženie prevádzkových nákladov.“ [22] Systém pracoval na základe „scenárov“, keď najprv identifikoval životnú situáciu (problém, žiadosť) používateľa, zistil dostupné služby a na základe nich vypracoval scenár postupu vyriešenia danej situácie. Používateľ bol celým procesom sprevádzaný tzv. „osobným asistentom“. Projekt bol nasadený vo viacerých krajinách (Nemecko, Poľsko, Slovensko). Na Slovensku bol použitý v Košickom samosprávnom kraji (mesto Michalovce) na zlepšenie získavania stavebných povolení.

Ďalším významným projektom s účasťou zástupcov aj zo Slovenska bol projekt SAKE (Semantic-enabled Agile Knowledge-based E-Government) [24]. Na tomto projekte sa okrem zahraničných partnerov podieľali Technická Univerzita v Košiciach a Miestny úrad Košice – sídlisko Ťahanovce. Cieľ projektu bol na [22] špecifikovaný takto: „Cieľom projektu je vyšpecifikovať, vyvinúť a nasadiť holistický rámec a podporné nástroje pre agilný znalostný eGovernment dostatočne flexibilný adaptovať sa rozdielnym a meniacim sa potrebám a prostrediam. SAKE zaistí kontinuálne zlepšovanie kvality rozhodovacieho procesu cez aplikáciu sémantických technológií pri konzistentnej propagácii zmien, zvyšovaním produktivity zamestnancov alebo celej organizácie. Vybaví zamestnancov verejnej správy efektívnym prístupom k znalostiam potrebným k riešeniu problémov rýchlo a presne.“ Na Slovensku bol implementovaný v Košiciach, v mestskej časti Ťahanovce v procese tvorby lokálnych legislatívnych noriem [22]. Vo svete bolo uskutočnených ešte mnoho projektov, ktoré so SAKE súvisia. Sú to:

- **Webocracy** (Web Technologies Supporting Direct Participation in Democratic Processes)
  http://www.webocrat.sk/
- **E-Power** (Program for an Ontology-based Working Environment for Rules and regulations)
  www.belastingdienst.nl/epower

- **Estrella** (The European project for Standardized Transparent Representations in order to Extend Legal Accessibility)
http://www.estrellaproject.org/
- **EDEN Project** (Electronic Democracy European Network)
http://www.edentool.org/
- **ACKNOWNET** (Active Knowledge Manager Using Dynamic Self-Modifying Knowledge Models
http://www.tupaisystems.co.il/AcknownetSite/
- **DECOR** (Delivery of Context-sensitive Organisational Knowledge)
http://www.dfki.uni-kl.de/decor/
- **KIWI** (Building Innovative Knowledge Management Infrastructures Within European Public Administrations)
http://www.ist-kiwi.org
- **TERREGOV** (Impact of eGovernment on Territorial Government Service)
http://www.terregov.eupm.net/my_spip/index.php
- **OntoGov** (Ontology-enabled e-Gov Service Configuration)
http://www.ontogov.com
- **EU-PUBLI.com** (Facilitating Co-operation amongst European Public Administration employees through a Unitary European Network Architecture and the use of Interoperable Middleware Components)
www.eu-publi.com
- **Acces-eGov** (Access to e-Government Services Employing Semantic Technologies)
http://www.accessegov.org/
- **MAP** (Mobile Adaptive Procedure)
http://www.map-project.net/
- **One Stop Gov** (A life-event oriented framework and platform for one-stop government)
http://www.onestopgov-project.org/

2. Ciele a metodika práce

2.1 Ciele

Cieľom bakalárskej práce bolo sprehľadniť projekty, ktoré si kladú za cieľ podporiť rozvoj sémantického webu. Tieto projekty potom rozdeliť do skupín odrážajúcich ich súčasný stav (bežiace v kapitole Výsledky práce a skončené v kapitole Súčasný stav skúmanej problematiky). Ďalším cieľom práce bolo poukázanie na perspektívne aktivity medzi prebiehajúcimi projektmi a návrh ich možného uplatnenia v domácich podmienkach.

2.2 Metodika

V priebehu vypracovania bakalárskej práce boli použité viaceré vedecké metódy. Pred charakterizáciou použitých metód je dôležité definovať, čo vedecká metóda je. Vedecká metóda je „súhrn pravidiel, ktorými sa treba riadiť v procese poznania alebo súbor pravidiel, vyjadrujúci účelný a objektívne zvolený spôsob ako skúmať jav a dosiahnuť vedecké poznanie“ [26]. Na uvedenej stránke sa nachádza rozsiahlejšia definícia pojmu, ako aj postup získavania poznania a bližšia špecifikácia niektorých metód. Podstatnými vlastnosťami vedeckej metódy sú:

1. „vedecká metóda je vedome kritická,
2. objektívne oznámiteľná,
3. objektívne overiteľná“

Bližšie vysvetlenie uvedených vlastností, spolu s ďalšími prvkami potrebnými k pochopeniu vedeckých metód, je na stránke [27]. Vyčerpávajúci výpočet vedeckých metód poskytuje [28]. Vzhľadom na teoretický charakter práce boli pri jej vypracovaní v hlavnej miere použité nasledujúce vedecké metódy a myšlienkové postupy:

- Pozorovanie – realizované najmä systematickým zberom informácií, ktoré boli následne vyhodnotené a použité pri tvorbe práce
- Klasifikácia – skúmané projekty boli rozdelené do skupín podľa toho, či sú bežiace alebo skončené a podľa miesta aplikácie, kde bol dôraz kladený na odlíšenie projektov aplikovaných na Slovensku
- Korelácia – viaceré projekty na seba nadväzujú, čím je vyjadrená závislosť medzi skúmanými projektmi
- Komparácia – na zistenie miery nadväznosti, resp. závislosti projektov bolo nutné porovnať ich ciele a použité technológie
- Analýza – pri porovnávaní cieľov a použitých technológií

2.3 Teoretické základy sémantického webu

2.3.1 Čo je sémantický web

Sémantický web je sieť informácií pospájaných tak, aby boli ľahko spracovateľné strojmi v globálnom meradle. Môžeme si ho predstaviť ako efektívny spôsob reprezentovania dát na World Wide Web alebo ako globálne prepojenú databázu. Sémantický web bol vymyslený Timom Berners-Leeom, ktorý vynašiel WWW, URI a iné. Vo World Wide Web Consortium (W3C) je odhodlaná skupina ľudí pracujúca na vylepšení, rozšírení a štandardizovaní systému a do teraz už bolo vyvinutých mnoho jazykov, publikácií a nástrojov. Technológie sémantického webu sú zatiaľ len v plienkach a hoci budúcnosť projektu sa zdá vo všeobecnosti jasná, momentálne sú značné nezrovnalosti ohľadom smerovania a charakteristík raného sémantického webu. Aké je odôvodnenie pre takýto systém? Dáta, ktoré sú skryté v HTML dokumentoch sú často užitočné v určitom kontexte, ale nie v žiadnom inom. Problém s dátami, ktoré sú na webe v takejto forme v súčasnosti je, že je ich ťažké použiť vo veľkom rozsahu, pretože neexistuje systém na publikovanie dát v takej forme, aby boli jednoducho spracovateľné pre každého (hocikoho). Napríklad informácie o športových podujatiach, počasí, TV programoch, ... sú všetky prezentované početnými stránkami vo forme HTML. Ich problém teda je, že v niektorých kontextoch je ťažké použiť tieto dáta tak, ako by niekto možno chcel. Takže na sémantický web môže byť nazerané ako na veľké technické riešenie, ale je viac než len to. Uvidíme, že ako bude čoraz jednoduchšie publikovať dáta viacúčelovo, viac ľudí bude chcieť publikovať dáta a to bude začiatok dominového efektu. Môžeme zistiť, že rôzne aplikácie sémantického webu sa dajú použiť na rôznorodé úlohy, zvyšujúc tak modularitu aplikácií na webe. A teraz ako sa to všetko dosiahne. Sémantický web je vo všeobecnosti postavený na syntaxách, ktoré používajú URI na reprezentáciu dát, obyčajne v štruktúrach založených na trojiciach, t. j. mnoho trojíc URI dát môže byť uložených v databázach alebo vymieňaných na WWW s použitím sady čiastočných syntaxí vyvinutých špeciálne na tento účel. Tieto syntaxe sa volajú „Resource Descritpion Framework“.

2.3.2 URI

Na identifikáciu predmetov na webe používame identifikátory. Pretože používame jednotný systém identifikátorov a pretože každý identifikovaný predmet je považovaný za „zdroj“, voláme tieto identifikátory „Uniform Resource Identifier“ alebo URI. URI je teda webový identifikátor ako reťazce začínajúce na „http:“ alebo „ftp:“, ktoré sa často nachádzajú ne webe. URI môžeme dať hocičomu a o všetkom, čo má URI môže byť povedané, že je na

webe, napríklad: osoba, kniha, živočích... URI je základom webu. Zatiaľ, čo takmer každá ďalšia časť webu môže byť vymenená, URI nie – drží zvyšok webu pohromade. Jeden zo známych URI je URL (Uniform Resource Locator). Je to adresa, ktorá nám umožňuje navštíviť webovú stránku. URL hovorí počítaču, kde má nájsť špecifický zdroj (webovú stránku). Na rozdiel od väčšiny ostatných URI, URL identifikuje aj lokalizuje. Pretože web je príliš rozsiahly pre akúkoľvek organizáciu, aby ho kontrolovala, URI sú decentralizované. Žiadna osoba ani organizácia nekontroluje kto URI vytvára alebo ako môžu byť použité. Hocikto môže vytvoriť URI a jeho vlastníctvo je jasne delegované, takže sa stáva ideálnou základnou technológiou na vybudovanie globálneho webu (napr. WWW je takou technológiou). Na syntax URI, však starostlivo dohliada IETF (Internet Engineering Task Force), ktorá vydala všeobecnú špecifikáciu pre URI, RCF 2396. W3C uchováva zoznam URI schém. Zatiaľ, čo niektoré URI schémy (napr. http:)sú závislé na centralizovaných systémoch (ako napr. DNS), iné schémy (napr. freenet:) sú kompletne decentralizované. To znamená, že nepotrebujete od nikoho povolenie na vytvorenie URI. Môžete dokonca vytvoriť URI aj pre zdroj, ktorý nevlastníte. Táto flexibilita robí URI silným, zároveň so sebou prináša aj niekoľko problémov. Keďže hocikto môže vytvoriť URI, nevyhnutne skončíme s viacerými URI reprezentujúcimi tú istú vec. Čo je ešte horšie, neexistuje spôsob, ako určiť, či dve URI odkazujú na presne ten istý zdroj alebo nie. Takto nebudeme nikdy schopní s určitosťou povedať, čo dané URI znamená. To je ale daň, ktorú musíme zaplatiť, ak chceme vytvoriť niečo tak obrovské ako sémantický web. Bežná prax pri vytváraní URI je začať s webovou stránkou. Stránka opisuje objekt, ktorý má byť identifikovaný a zároveň hovorí, že jej URL je URI pre ten objekt. Táto URL adresa potom plní dve úlohy: reprezentuje skutočný predmet a webovú stránku, ktorá ho opisuje. Toto sa volá „The semantic web identification problem“ a opakovane sa o tomto probléme diskutuje medzi pracovníkmi sémantického webu. Toto je dôležitý fakt na porozumenie. URI nie je súbor inštrukcií, ktoré povedia počítaču ako získať špecifický súbor na webe (hoci to robiť môžu). Je to iba meno pre zdroj (nejakú vec). Tento zdroj môže, ale nemusí byť dostupný cez internet. URI môže, ale nemusí poskytnúť spôsob pre počítač, ako získať ďalšie informácie o danom zdroji. URL je URI, ktoré poskytuje spôsob, ako získať informácie o zdroji alebo možno aj ako získať k nemu prístup a ďalšie metódy na poskytovanie informácií o URI a zdrojoch, ktoré identifikujú sú vo vývoji. Je tiež pravda, že schopnosť povedať niečo o URI je dôležitou časťou sémantického webu. Ale netreba predpokladať, že URI poskytuje viac ako identifikátor pre zdroj.

2.3.3 Extensible Markup Language (XML)

XML bol navrhnutý, aby bol jednoduchý spôsob posielania dokumentov cez web. Dovoľuje každému navrhnúť si vlastný formát dokumentov a potom napísať dokument v tomto formáte. Tieto dokumenty môžu obsahovať značky na posilnenie významu obsahu dokumentu. Tieto značky sú strojovo čitateľné, to znamená, že programy ich dokážu prečítať a porozumieť im. Zahrnutím strojovo čitateľných dát v našom dokumente ho robíme podstatne silnejším. Napríklad: ak dokument obsahuje slová označené ako zdôraznené, spôsob zobrazenia týchto slov sa môže prispôsobiť kontextu. Webový prehliadač ich môže zobraziť kurzívou a hlasový prehliadač, ktorý číta stránky nahlas ich môže prečítať silnejším hlasom. Keby boli slová označené iba ako kurzíva, počítač nevie prečo sú zobrazené kurzívou a hlasový prehliadač nevie ako ich má interpretovať. Pri vytváraní vlastných značiek v XML sa môžu vyskytnúť určité problémy. Napríklad niekto iný použije rovnaký názov pre svoju značku, ale sleduje ňou úplne iný význam. Aby sme predišli zmätku, musíme svoje značky jednoznačne identifikovať. Identifikovať treba každý z prvkov a atribútov. Na ich identifikáciu poslúži URI. Deje sa to pomocou menných priestorov (XML Namesapces). Takto môže každý vytvoriť svoje vlastné značky a zmiešať ich so značkami ostatných. Menný priestor je spôsob na identifikáciu časti webu (priestor), z ktorého odvodzujeme význam použitých mien. Menný priestor pre naše značky vytvoríme tak, že preň vytvoríme URI. Na to sa môže použiť webová stránka s opisom nášho jazyka. Keďže všetky značky majú svoje URI, nemusíme sa báť konfliktov medzi ich menami.

2.3.4 RDF

RDF poskytuje možnosť vytvárať oznámenia (tvrdenia), ktoré sú strojovo spracovateľné. Počítač v skutočnosti nerozumie o čom sa ľudia bavia, ale už s tým vie narábať ako by tomu rozumel. Napríklad by sme mohli vyhľadať na webe všetky recenzie na všetky knihy a urobiť pre ne priemerné hodnotenia, ktoré by sme potom publikovali naspäť na web. A nejaká iná stránka by mohla tieto informácie zobrať a vytvoriť „Top ten list“. RDF je dosť jednoduché. RDF oznámenie vyzerá ako normálna veta až na to, že takmer všetky slová sú URI. Každá takáto veta má tri časti: podmet, prísudok a predmet. Veta by mohla vyzerať napríklad takto:

```
<http://aaronsw.com/> <http://love.example.org/terms/reallyLikes>
<http://www.w3.org/People/Berners-Lee/Weaving/> .
```

V tomto príklade, prvý URI je podmet. Je to nejaká osoba „Aaron“. Druhý URI je prísudok. Spája podmet a predmet. V tomto prípade vyjadruje vzťah „má naozaj rád“. Tretí

URI je predmet. Tu je to kniha od Tima Bernersa-Leeho „Weaving the Web“. Takže toto oznámenie tvrdí, že Aaron má naozaj rád Weaving the Web.

RDF oznámenia môžu tvrdiť prakticky hocičo a nezáleží na tom, kto ich hovorí. To nás privádza k dôležitému princípu RDF, ktorý hovorí „čokoľvek môže povedať čokoľvek o čomkoľvek“. Informácia sa šíri webom a dokonca môžu dvaja ľudia povedať úplne protichodné veci. To je sloboda, ktorú nám web poskytuje. Vyššie uvedené tvrdenie je napísané vo forme tzv. „tripletov“, jazyku, ktorý umožňuje písanie jednoduchých RDF tvrdení. Trojica (triplet) môže byť jednoducho opísaná ako tri URI. Jazyk, ktorý používa tri URI takýmto spôsobom sa volá RDF. W3C vyvinulo XML serializáciu RDF vo svojej RDF Model and Syntax recommendation. RDF XML je považovaný za štandardný výmenný formát pre RDF na sémantickom webe, hoci nie je jediný. Napríklad Notation3 je excelentná alternatívna serializácia v jednoduchom texte. Keď sú informácie vo forme RDF je jednoduché ich spracovávať, pretože RDF je generický formát a má už mnoho syntaktických analyzátorov. XML RDF je mnohoslovná špecifikácia a môže trvať dlhšiu dobu, kým sa s ňou oboznámime, lebo je potrebné poznať aspoň niečo z XML a menných priestorov. Príklad XML RDF vyzerá nasledovne:

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:dc="http://purl.org/dc/elements/1.1/"
    xmlns:foaf="http://xmlns.com/0.1/foaf/" >
    <rdf:Description rdf:about="">
        <dc:creator rdf:parseType="Resource">
            <foaf:name>Sean B. Palmer</foaf:name>
        </dc:creator>
        <dc:title>The Semantic Web: An Introduction</dc:title>
    </rdf:Description>
</rdf:RDF>
```

Tento kus RDF hovorí, že článok má titul „The Semantic Web: An Introduction“, a že ho napísal niekto s menom Sean B. Palmer. Toto sú triplety (trojice), ktoré RDF ponúka:

```
<> <http://purl.org/dc/elements/1.1/creator> _:x0 .
this <http://purl.org/dc/elements/1.1/title> "The Semantic Web: An
Introduction" .
_:x0 <http://xmlns.com/0.1/foaf/name> "Sean B. Palmer" .
```

Tento formát je v skutočnosti serializácia v jednoduchom texte Notation3. Niektorí ľudia uprednostňujú XML RDF pred Notation3, ale vo všeobecnosti je Notation3 považovaná za jednoduchšiu na použitie a samozrejme je konvertovateľná do XML RDF.

Písanie RDF, ktoré vyzerá ako je ukázané vyššie (v XML RDF) nie je jednoduché a zdá sa nepravdepodobné, že všetci sa začnú vyjadrovať v tomto novom jazyku v dohľadnej dobe. Tak odkiaľ očakávame, že všetky tieto RDF informácie prídu? Najpravdepodobnejší zdroj sú databázy. Vo svete existujú tisícky databáz, ktoré obsahujú strojovo spracovateľné informácie. Keď sú informácie uložené v databáze, je veľmi jednoduché opýtať sa počítača niektoré otázky o dátach. RDF je ideálne pre publikovanie týchto databáz na webe. Keď dávame databázu ne web, dávame každému jej záznamu URI, takže ostatní ľudia sa o nich môžu tiež rozprávať. Tak môžu inteligentné programy začať spájať dáta do hromady a tým sa môžeme pýtať viac otázok týchto databáz naraz.

2.3.5 Prečo RDF?

Keď sa ľudia po prvý krát stretnú s XML RDF, zvyčajne sa pýtajú:

1. Prečo používať RDF radšej ako XML?
2. Používa sa XML Schema v spolupráci s RDF?

Odpoveď na prvú otázku je jednoduchá a skladá sa z dvoch častí. Prvá časť je prospech, ktorý sa získava z pretiahnutia jazyka do RDF je ten, že informácie sú mapované priamo a nedvojzmyselne na model, ktorý je decentralizovaný, a pre ktorý existuje veľa všeobecných syntaktických analyzátorov. To znamená, že keď máme RDF aplikáciu, vieme ktoré dáta sú sémantika aplikácie a ktoré časti sú iba syntaktické konštrukcie. Tieto veci sú často zrejmé implicitne, bez študovania špecifikácie, pretože RDF je už dobre známe. Druhá časť odpovede je tá, že tvorcovia dúfajú, že RDF sa stane súčasťou sémantického webu, a tak sa prospech z konvertovania dát do RDF vykresľuje porovnateľne s konvertovaním dát do HTML v čase, keď web iba začínal. Odpoveď na druhú otázku je takmer rovnako jasná. XML Schema je jazyk na obmedzenie syntaxi XML aplikácií. RDF už má v sebe zabudovaný BNF, ktorý nastavuje ako má byť jazyk používaný, takže ohľadom tohto je odpoveď tvrdé nie. Avšak používanie XML Schema spolu s RDF môže byť užitočné na vytváranie dátových typov a podobne. Z toho vyplýva odpoveď možno s námietkou, že XML Schema nie je v skutočnosti používaná na kontrolovanie syntaxi RDF. Toto je bežné nedorozumenie, ktoré pretrváva už dlho.

2.3.6 Screen Scraping a Forms

Na dosiahnutie plného potenciálu sémantického webu je potrebné, aby veľa ľudí začalo publikovať dáta vo forme RDF. Skade majú tieto informácie pochádzať? Mnoho ich môže byť odvodených z množstva publikácií, ktoré existujú dnes s použitím procesu, ktorý sa volá

„screen scraping“. Screen scraping je proces, kde sa dáta zo zdroja pretransformujú do lepšie manažovateľnej formy (t. j. RDF) s použitím akýchkoľvek prostriedkov na to potrebných. Dva užitočné nástroje na screen scraping sú XSLT (jazyk na transformáciu XML) a RegExps. Napriek tomu je screen scraping často zdĺhavý proces, takže iný spôsob prístupu je vybudovať správne RDF systémy, ktoré zoberú vstup od používateľa a uložia ho hneď vo forme RDF. Dáta, ktoré môžu byť vložené sú napr. vytvorenie nového emailového účtu, kúpa tovaru cez internet alebo vyhľadanie ojazdeného auta. Všetky tieto dáta môžu byť uložené ako RDF a potom použité na sémantickom webe.

2.3.6 Notation3

XML RDF je relatívne náročné, ale existujú aj iné, jednoduchšie formy RDF. Jednou z nich je Notation3, ktorá bola vytvorená Timom Berners-Leeom. Kritériá pri vytváraní Notation3 boli značne jednoduché – vytvoriť jednoduchý ľahko naučiteľný písateľný RDF formát, ktorý je ľahko analyzovateľný a dajú sa na ňom jednoducho vystavať väčšie aplikácie. V Notation3 sú URI jednoducho písané v tripletoch, kde sú označené < a >. Napríklad jednoduchý triplet vyzerá nasledovne:

```
<http://xyz.org/#a> <http://xyz.org/#b> <http://xyz.org/#c>
```

Pre použitie reťazcov sú tieto ohraničené úvodzovkami:

```
<http://xyz.org/#Sean> <http://xyz.org/#name> "Sean"
```

Pokiaľ nechceme uviesť URI niečoho, o čom hovoríme v triplete, existuje na to spôsob. Namiesto URI sa napíše podtržník, dvojbodka a značka:

```
_:a1 <http://xyz.org/#name> "Sean"
```

Toto môže byť čítané ako „niečo, čo má meno Sean“ alebo „a1 má meno Sean pre niektoré hodnoty a1“. Takýto zápis sa volá anonymný uzol, lebo nemá URI a občas sa im hovorí existenčne kvalifikované uzly. V predchádzajúcich príkladoch bolo použité URI http://xyz.org/# trikrát iba so zmenou písmena za #. Notation3 ponúka pekný spôsob na skrátenie podobných reťazcov tým, že priradíme častiam URI aliasy a budeme používať tie. Alias sa priraďuje takto:

```
@prefix xyz: <http://xyz.org/#>
```

Pred použitím aliasu musí byť alias vždy deklarovaný. Na použitie aliasu napíšeme „xyz:“ namiesto URI a neobaľujeme ho do <>. Čiže to vyzerá nasledovne:

```
xyz:a xyz:b xyz:c
```

Nezáleží na tom aký alias použijeme, pokiaľ ho používame v celom dokumente. Tiež môže byť deklarovaných viac aliasov pre to isté URI. Nasledujúce fragmenty sú ekvivalentné s fragmentom vyššie:

```
@prefix blargh: <http://xyz.org/#> .
blargh:a blargh:b blargh:c .
@prefix blargh: <http://xyz.org/#> .
@prefix xyz: <http://xyz.org/#> .
blargh:a xyz:b blargh:c .
```

Je dôležité poznamenať, že niektoré aliasy sa používajú celkom bežne, takže keď vývojári sémantického webu vymieňajú kód v čistom texte, môžu vynechať prefixy a ľudia sú schopní porozumieť, čo sa v texte píše. Správny kód by ale nemal používať túto vlastnosť. Niektoré zo štandardných aliasov sú:

```
@prefix : <#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix daml: <http://www.daml.org/2001/03/daml+oil#> .
@prefix log: <http://www.w3.org/2000/10/swap/log#> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
```

Prázdny alias „:“ je používaný na označenie nového menného priestoru, pre ktorý autor ešte nevytvoril URI. Notation3 zahŕňa ešte mnoho ďalších konštrukcií vrátane kontextov, DAML listov a alternatívnych spôsobov reprezentovania anonymných uzlov. Bola vynájdená aj syntax, ktorá je zjednodušenou podmnožinou Notation3 a volá sa N-Triples, ale nepoužíva prefixy, a preto mnohé príklady tu použité nie sú platné v N-Triples, ale sú platné v Notation3. Názov Notation3 je odvodený od toho, že RDF M&S bolo prvé, RDF strawman bolo druhé a toto je tretie.

2.3.7 CWM: XML RDF a Notation3 odvodzovací engine

Je vhodné poznamenať, že v súčasnosti je mnoho zo spracovania RDF a sémantického webu (aj keď len na experimentálne a demonštratívne účely) je robené pomocou programu napísaného v Pythone, CWM (Closed World Machine). CWM je veľmi silným balíkom nástrojov sémantického webu. Najlepším príkladom jeho použitia je to, ako je konvertovaný

súbor z XML RDF do Notation3 alebo naopak. Na konverziu súboru „a.n3“ do „a.rdf“ sa použije takýto príkaz:

```
python cwm.py a.n3 -rdf > a.rdf
```

2.3.8 Schémy a ontológie: RDF Schema, DAML+OIL, WebOnt

Všetka práca na databázach predpokladá, že dáta sú takmer perfektné. Veľmi málo databázových systémov je pripravených na neusporiadanosť webu. Akýkoľvek systém, ktorý je tvrdo naprogramovaný, aby rozumel určitým výrazom bude pravdepodobne zastaraný alebo jeho užitočnosť bude značne limitovaná, keď budú vynájdené a definované nové výrazy. Čo ak napríklad niekto príde s novým systémom hodnotenia kníh stupnicou od 1 do 10 namiesto povedania, že ich niekto „má veľmi rád“? Programy postavené na starom systéme nebudú schopné spracovať nové informácie. Navyše neexistuje spôsob, akým by počítač alebo človek zistil, čo nejaký špecifický výraz znamená alebo ako by sa mal používať. Používanie URI je bezvýznamné, ak neopíšeme, čo znamenajú. Práve túto časť sémantického webu riešia schémy a ontológie. Schéma a ontológia sú spôsoby, ako opísať význam a vzťahy medzi výrazmi. Tento opis (v RDF) pomáha počítačovým systémom používať výrazy jednoduchšie a rozhodovať ako medzi nimi konvertovať. Ako riešenie tohto problému boli vyvinuté dva úzko previazané systémy: RDF Schema a DARPA Agent Markup Language with Ontology Interface Layer (DAML+OIL). Schéma by mohla vyzerať napríklad takto:

```
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

# A creator is a type of contributor:
dc:creator rdfs:subClassOf dc:contributor .
```

Tu sa používa Notation3 ako set nadradený tripletom, ktorý nám umožňuje používať viac skratiek. Táto schéma hovorí, že tvorca je podtriedou prispievateľa. Na čo je to dobré? Napríklad by sme napísali program, ktorý zbiera tvorcov a prispievateľov rôznych dokumentov. Tento program používa túto slovnú zásobu na porozumenie informáciám, ktoré získava. Zrazu mnoho ľudí začne vytvárať RDF dokumenty a nikto z nich nevie o našej slovnej zásobe, tak si vymyslia vlastný termín ed:hasAuthor. Starý spôsob:

```
<http://aaronsw.com/> is dc:creator of
<http://logicerror.com/semanticWeb-long>
```

Nový spôsob:

```
<http://logicerror.com/semanticWeb-long> ed:hasAuthor
<http://aaronsw.com/>
```

Normálne by náš program jednoducho ignoroval tieto tvrdenia, pretože im nerozumie. Našťastie niekto prepojil tieto slovné zásoby tým, že poskytol informáciu ako medzi nimi konvertovať.

```
# [X dc:creator Y] is the same as [Y ed:hasAuthor X]
    dc:creator daml:inverse ed:hasAuthor .
```

Toto nášmu programu hovorí, že ed:hasAuthor je opakom dc:creator. To znamená, že všetko, čo musí náš program urobiť je vymeniť podmet a predmet a vymeniť ed:hasAuthor za dc:creator. A keďže náš program pozná DAML ontológie, môže teraz zobrať túto informáciu a použiť ju na spracovanie všetkých hasAuthor záznamov, ktorým predtým nerozumel.

2.3.8.1 Schémy

Prvá vrstva sémantického webu opísaná vyššie je jednoduchý dátový model. Schéma je dokument alebo časť kódu, ktorá kontroluje sústavu výrazov v inom dokumente alebo inej časti kódu.

2.3.8.2 RDF Schema

RDF Schema bola navrhnutá ako jednoduchý dátový model pre RDF. S jej použitím vieme povedať, že Fido je typ pes, a že pes je podtrieda zvieraťa. Tiež vieme vytvoriť vlastnosti a triedy ako aj trochu rozvinutejšie definície, napríklad intervaly a obory pre vlastnosti. Všetky výrazy pre RDF Schema začínajú reťazcom http://www.w3.org/2000/01/rdf-schema#, ktorý sa nachádza medzi štandardnými aliasmi. Často sa používa alias „rdfs:“ a bude použitý aj v tomto texte. Tri najdôležitejšie koncepty, ktoré nám RDF a RDF Schema ponúkajú sú: „Resource“ (rdfs:Resource), „Class“ (rdfs:Class)a „Property“ (rdf:Property). Tieto všetky koncepty sú „triedy“, čo znamená, že všetky výrazy môžu patriť pod niektorú z nich. Napríklad všetky výrazy v RDF sú typu Resource. Na deklarovanie, že niečo je typom niečoho iného sa používa vlastnosť rdf:type.

```
rdfs:Resource rdf:type rdfs:Class .
rdfs:Class rdf:type rdfs:Class .
rdf:Property rdf:type rdfs:Class .
rdf:type rdf:type rdf:Property .
```

Toto hovorí, že Resource je typom Class, Class je typom Class, Property je typom Class a typ je typom Property. Toto všetko sú pravdivé tvrdenia. Je celkom jednoduché vytvoriť si vlastnú triedu. Vytvoríme napríklad triedu pes, v ktorej budú všetky psy.

```
:Dog rdf:type rdfs:Class
```

Teraz môžeme povedať, že Fido je typom psa.

```
:Fido rdf:type :Dog
```

Tiež môžeme vytvárať vlastnosti pomerne jednoducho. Iba povieme, že výraz je typom rdf:Property a potom použijeme naše vlastnosti v RDF.

```
:name rdf:type rdf:Property
:Fido :name "Fido"
```

Prečo sme povedali, že meno Fida je Fido? Pretože výraz Fido je URI a my by sme mohli vybrať pre Fida akékoľvek URI, aj také, ktoré by nedávalo pre ľudí zmysel. Používame URI Fido, lebo sa ľahšie pamätá. Napriek tomu stále musíme stroju povedať, že jeho meno je Fido, lebo hoci to ľudia dokážu uhádnuť z URI, stroje nie. RDf Schema má aj ďalšie vlastnosti, ktoré by sme mohli použiť. Sú to rdfs:subClassOf a rdfs:subPropertyOf a hovoria, že výraz je podtriedou alebo podvlastnosťou niečoho iného. Napríklad by sme mohli povedať, že pes je podtriedou triedy zviera.

```
:Dog rdfs:subClassOf :Animal
```

Takže keď teraz povieme, že Fido je pes, zároveň povieme aj to, že Fido je zviera. Takto by sme mohli vytvoriť nové podtriedy triedy zviera a potom aj inštancie týchto tried. A potom aj nové vlastnosti a použiť ich na vytvorenie ďalších informácií.

```
:Human rdfs:subClassOf :Animal
:Duck rdfs:subClassOf :Animal
:Bob rdf:type :Human
:Quakcy rdf:type :Duck
:owns rdf:type rdf:Property
:Bob :owns :Fido
:Bob :owns :Quacky
:Bob :name "Bob Fleming"
:Quacky :name "Quakcy"
```

RDF Schema je veľmi jednoduchá a umožňuje nám vybudovať poznatkové základne dát v RDF veľmi rýchlo. Ďalšie koncepty, ktoré nám RDF Schema ponúka a je dôležité, aby boli spomenuté, sú intervaly a obory. Sú to intervaly a obory, do ktorých musia patriť všetky objekty tried a objekty všetkých vlastností. Napríklad chceme povedať, že vlastnosť :bookTitle sa môže vzťahovať iba na knihu a musí mať hodnotu literálu.

```
:Book rdf:type rdfs:Class
:bookTitle rdf:type rdf:Property
:bookTitle rdfs:domain :Book
:bookTitle rdfs:range rdfs:Literal
:MyBook rdf:type :Book
```

```
:MyBook :bookTitle "My Book"
```

V tomto fragmente kódu rdfs:domain vždy označuje ktorej triede patrí objekt tripletu a rdfs:range označuje ktorej triede patrí objekt tripletu majúci danú vlastnosť. RDF Schema tiež obsahuje súbor vlastností pre anotáciu schém, poskytovanie komentárov, značiek a podobne. Vlastnosti pre dve zo spomenutých sú rdfs:label a rdfs:comment. Príklad ich použitia je nasledovný:

```
:bookTitle rdfs:label "bookTitle";
   rdfs:comment "the title of a book"
```

Je dobré vždy označiť a okomentovať nové triedy, vlastnosti a aj iné výrazy.

2.3.9 Ontológie, odvodzovanie a DAML

DAML je jazyk vytvorený DARPA ako jazyk pre ontológie a odvodzovanie založený na RDF. DAML posúva RDF Schemu o krok ďalej tým, že poskytuje viac hlbších vlastností a tried.

2.3.9.1 DAML+OIL

DAML ponúka metódu na vyjadrenie takých vzťahov ako sú inverzia, jednoznačné vlastnosti, unikátne vlastnosti, zoznamy, obmedzenia, kardinalita, párovosť v zoznamoch, dátové typy atď. Jedna z konštrukcií, ktoré DAML ponúka je vlastnosť daml:inverseOf. Touto vlastnosťou sa dá povedať, že jedna vlastnosť je opakom inej. Hodnoty rdfs:range a rdfs:domain sú vlastnosťami (rdf:Property) tejto konštrukcie. Môže vyzerať takto:

```
:hasName daml:inverseOf :isNameOf
:Sean :hasName "Sean"
"Sean" :isNameOf :Sean
```

Druhou užitočnou konštrukciou je trieda daml:UnambiguousProperty. Hovorí, že ak predmet (vlastnosti) je ten istý, podmety sú ekvivalentné. Napríklad:

```
foaf:mbox rdf:type daml:UnambiguousProperty
:x foaf:mbox
:y foaf:mbox
```

naznačuje, že :x daml:equivalentTo :y. Veľa aplikácií sémantického webu zahŕňa DAML, hoci nie je to jediný jazyk, ktorý sa používa.

2.3.9.2 Odvodzovanie

Princíp odvodzovania je celkom jednoduchý – byť schopný vyvodiť nové dáta z dát, ktoré už sú známe. V matematike je dotazovanie formou odvodzovania (napríklad schopnosť získať výsledky hľadania z masy dát). Odvodzovanie je jedným z hybných princípov

sémantického webu, lebo nám umožní vytvárať softvérové aplikácie dosť jednoducho. Na ilustráciu použijeme jednoduchý príklad s autom.

```
:MyCar de:macht "160KW"
```

V nemeckom procesore sémantického webu výraz „:macht“ môže byť zabudovaný a hoci anglický procesor môže mať ekvivalentný výraz zabudovaný v sebe, nebude rozumieť celému kódu s časťou, ktorej nerozumie. A tu je zápis, vďaka ktorému bude systém schopný porozumieť.

```
de:macht daml:equivalentTo en:power
```

Tu sme použili DAML vlastnosť „:equivalentTo“ na vyjadrenie toho, že macht v nemeckom systéme je ekvivalentný výraz k power v anglickom systéme. S použitím odvodzovacieho enginu by klient sémantického webu mohol určiť:

```
:MyCar en:power "160KW"
```

Toto je len jednoduchý príklad, ale pekne ilustruje ako by mohol byť systém zjednodušený. Spájanie databáz bude len otázkou zapísania do RDF, že meno osoby v prvej databáze je ekvivalentné s menom v druhej databáze. Potom stačí už len dať informácie dokopy a nechať procesor nech ich spojí. Toto je možné s nástrojmi sémantického webu, ktoré sú k dispozícii už dnes, napríklad CWM. Bohužiaľ, veľké úrovne odvodzovania môžu byť robené iba s použitím First Order Predicate Logic a DAML nie je úplne FOPL jazyk.

2.3. 10 Logika

Aby bol sémantický web schopný poskytnúť dostatok výrazov na to, aby bol užitočný v širokej palete situácií, bude nutné skonštruovať mocný logický jazyk na odvodzovanie. V súčasnosti prebieha horúca debata o tom, ako a či vôbec je to možné, v ktorej niektorí ľudia poukazujú na to, že RDF nemá schopnosť kvantifikovať a oblasť kvantifikácie nie je dostatočne definovaná. Napriek tomu je v súčasnosti k dispozícii široký výber nástrojov na budovanie sémantického webu: výroky a citácie v RDF, triedy, vlastnosti, intervaly a dokumentácie v RDF Schema, rozpojené triedy, jednoznačné a unikátne vlastnosti, dátové typy, inverzie, ekvivalencie, zoznamy a mnohé ďalšie v DAML+OIL. Notation3 predstavuje konštrukciu „kontext“ a poskytuje schopnosť zhromažďovať tvrdenia dokopy a kvantifikovať ich s použitím špecificky navrhnutej logickej slovnej zásoby. Napríklad:

```
{ { :Joe :loves :TheSimpsons } a log:Falsehood
  { :Joe :is :Nuts } a log:Falsehood
} a log:Falsehood
```

Toto môže byť interpretované ako: nie je pravda, že Joe nemiluje Simpsonovcov a nie je blázon. Tento príklad nevychádza poriadne z XML RDF, pretože XML RDF nemá koncept kontextu ako je označený množinovými zátvorkami vyššie. Avšak podobný efekt môže byť dosiahnutý použitím zhmotnenia a kontajnerov.

2.3.11 Sila jazykov sémantického webu

Najväčšou silou jazykov sémantického webu je to, že každý môže vytvoriť svoj jazyk jednoducho tým, že publikuje nejaký RDF, ktorý popisuje súbor URI, čo robia a ako by mali byť použité. Napríklad RDF Schema a DAML sú veľmi silné jazyky na tvorbu ďalších jazykov. Vďaka tomu, že používame URI pre každý z výrazov nášho jazyka, môžeme zverejňovať jazyky ľahko a bez strachu, že budú zle interpretované alebo ukradnuté a s vedomím, že každý, kto má generický RDF procesor ich môže použiť.

2.3.11.1 Princíp čo najmenej moci

Sémantický web pracuje na princípe čo najmenej moci – čím menej pravidiel, tým lepšie. To znamená, že sémantický web je v zásade takmer neobmedzujúci v tom, čo môžu ľudia povedať a teda uplatňuje, že ktokoľvek môže povedať čokoľvek o čomkoľvek. Keď vezmeme do úvahy, čo je cieľom sémantického webu, bude veľmi jasné prečo je takáto úroveň moci dôležitá. Keby sme začali obmedzovať ľudí, neboli by schopní vytvoriť plný rozsah aplikácií, a tak by sa sémantický web mohol stať neužitočný pre niektorých ľudí.

2.3.11.2 Koľko je už príliš veľa?

Vyskytli sa názory, že táto sila bude príliš veľká. Čo ak sa stane, že niekto bude chcieť urobiť nejakú jednoduchú úlohu na odvodzovacom stroji a namiesto toho príde na plán pre svetový mier alebo niečo podobné? Toto sa ale nestane. Hoci sú sémantický web, RDF a koncepty na pozadí iba minimálne obmedzujúce, aplikácie postavené na sémantickom webe budú navrhnuté na vykonávanie špecifických úloh a ako také budú veľmi dobre definované. Napríklad program severových záznamov: niekto možno chce ukladať serverové záznamy v RDF a urobiť program, ktorý by ich používal na výpočet štatistík, ako napríklad návštevnosť a pod. To ale neznamená, že zmení CD mechaniku na hriankovač, iba bude spracovávať záznamy zo servera. Sila, ktorá sa získa z publikovania informácií v RDF je tá, že keď už sú raz publikované vo verejnej doméne, môžu byť použité na iné účely omnoho ľahšie. Keďže RDF používa URI, je plne decentralizované, takže netreba žiadať nejakú centrálnu autoritu, aby publikovala náš jazyka a dáta za nás, môžeme to spraviť sami. Je to Do It Yourself data management.

2.3.11.3 Pedantný web

Bohužiaľ, v sémantickom webe stále zostáva atmosféra akademického a korporátneho myslenia, čo viedlo k zavedeniu termínu „pedantný web“ a množstvu nesprávnych informácií a nepotrebnej reklame. Napríklad takmer každý začiatočník v RDF si musí prejsť cez tzv. „krízu identít“, kde si mýli osobu s jej menom, dokument s jeho názvom a pod. Príklad uvedeného problému vyzerá takto:

```
<http://example.org/> dc:creator "Bob"
```

Avšak Bob je len znakový reťazec a reťazec nemôže napísať dokument. Autor v skutočnosti znamená:

```
<http://example.org/> dc:creator _:b
_:b foaf:name "Bob"
```

Čiže example.org bol vytvorený niekým, kto sa volá Bob.

2.3.11.4 Vzdelávanie a presah

RDF Schema a DAML+OIL sú jazyky, ktoré je potrebné sa naučiť, ale čo robiť ak ľudia nemajú čas alebo potrebnú trpezlivosť, aby sa ich učili a aj tak chcú vytvárať aplikácie sémantického webu? Našťastie väčšina aplikácií sémantického webu budú menej náročné aplikácie, takže rozsah potrebných vedomostí o RDF nebude väčší ako vedomosti, ktoré vyžaduje Amaya o HTML (XHTML).

2.3.12 Dôvera a dôkaz

Ďalším krokom v architektúre sémantického webu je dôvera a dôkaz. Táto vrstva bude v budúcnosti veľmi užitočná. Najjednoduchším spôsobom ako ju vysvetliť je odpovedať na otázku: ak jeden človek povie, že x je modré a iný človek povie, že x nie je modré, nerozpadne sa celý sémantický web? Odpoveď je, samozrejme, nie a to z dvoch dôvodov. Po prvé: v súčasnosti aplikácie sémantického webu vo všeobecnosti silne závisia na kontexte a po druhé: v budúcnosti budú aplikácie obsahovať mechanizmy na kontrolu dôkazov a digitálne podpisy.

2.3.12.1 Kontext

Aplikácie sémantického webu budú na kontexte závisieť všeobecne preto, aby dali ľuďom vedieť, či dôverujú dátam. Ak napríklad dostanem RDF feed od kamaráta o filmoch, ktoré videl a teraz ich hodnotí, viem, že tejto informácii dôverujem. Navyše môžem použiť tú informáciu a bezpečne predpokladať, že prišla od neho a nechať na svojom vlastnom úsudku ako veľmi dôverujem jeho kritikám filmov, ktoré videl. Taktiež, skupiny ľudí pracujú na zdieľanom kontexte. Ak jedna skupina vyvíja zobrazovaciu službu na sémantickom webe, ktorá ukladá do katalógu kto sú ľudia, aké sú ich mená a kde sú ich fotky, potom moja dôvera

tejto skupine je závislá na tom, ako veľmi dôverujem ľuďom, ktorí sú v nej, že neurobia falošné tvrdenia. Z toho vyplýva, že kontext je dobrý v tom, že nás nechá pracovať na malých a stredných mierach intuitívne, bez nutnosti spoliehať sa na komplexné autentifikačné a kontrolné systémy. Čo ak sa objaví skupina, ktorú poznáme, ale nevieme ako overiť, že istá množina RDF dát pochádza on nej? Vtedy sa používajú digitálne podpisy.

2.3.12.2 Digitálne podpisy

Digitálny podpis je jednoducho malý kúsok kódu, ktorý môžeme použiť na jednoznačné overenie, že niekto napísal daný dokument. Digitálny podpis je založený na matematike a kryptografii a poskytuje dôkaz, že konkrétna osoba napísala (alebo súhlasí s) dokument alebo tvrdenie. Používa tú istú kľúčovo založenú PGP technológiu, ktorá sa používa na šifrovanie a podpisovanie správ. Jednoducho sa táto technológia pridá do RDF. Majme napríklad informáciu v RDF, ktorá obsahuje odkaz na digitálny podpis:

```
this :signature <http://example.org/signature>
:Joe :loves :Mary
```

Aby sme sa presvedčili, či veríme, že Joe naozaj miluje Mary, môžeme poskytnúť toto RDF „trust enginu“, čo je odvodzovací mechanizmus so zabudovaným overovaním digitálnych podpisov, a nechať ho zistiť, či dôverujeme zdroju tejto informácie.

Ďalšou oblasťou, kde sa môžu digitálne podpisy uplatniť je problém, že kvôli princípu čokoľvek môže povedať čokoľvek o čomkoľvek je na mieste otázka dôverovania takémuto webu. Pomocou digitálnych podpisov si môžu byť používatelia istí kto napísal dokument (resp. sa zaručil za jeho autentickosť). Tým si môžu jednoducho nastaviť svoj program ktorým podpisom má dôverovať a ktorým nie. S takto nastavenou úrovňou dôvery počítač sám rozhoduje, čo z toho, čo prečítal je dôveryhodné a čo nie je. Je značne nepravdepodobné, že ľudia budú natoľko dôverčiví, že budú používať veľkú časť webu. A to zabezpečuje Web of Trust. Povedzme, že dôverujeme nášmu priateľovi, ktorý zasa dôveruje ďalším ľuďom a tí zas ďalším atď. Ako sa tieto vzťahy dôvery rozširujú od nás ďalej, formujú Web of Trust a ku každému z týchto vzťahov je priradený istý stupeň dôvery, resp. nedôvery. Nedôvera je v tomto prípade rovnako užitočná ako dôvera. Napríklad ak sa náš počítač dostane k dokumentu, ktorému nikto explicitne nedôveruje, ale ani ho nikto neoznačil za nedôveryhodný, bude mu počítač dôverovať viac ako dokumentu, ktorý bol explicitne označený ako nedôveryhodný. Počítač zohľadňuje všetky tieto faktory keď sa rozhoduje nakoľko sú informácie dôveryhodné. Zároveň môže urobiť tento proces tak transparentný alebo nejasný ako chceme. Niekomu môže stačiť jednoduché thumbs up/thumbs down

a niekto iný môže požadovať komplexné vysvetlenie aj so všetkými faktormi, ktoré sa podieľali na vytvorení hodnotenia. Tim Berners-Lee navrhol „Oh, yeah?“ tlačidlo, ktoré sa po kliknutí pokúsi poskytnúť dôvody na dôveru dátam. Ale bez ohľadu na to, či sa o dôverovaní rozhodneme sami alebo to ponecháme na počítač, informácie potrebné na rozhodnutie budú dostupné cez Web of Trust.

2.3.12.3 Dôkazové jazyky

Dôkazový jazyk je jazyk, ktorý nám umožní dokázať, či je tvrdenie pravdivé alebo nie. Inštancia dôkazového jazyka bude vo všeobecnosti pozostávať zo zoznamu odvodzovacích predmetov, ktoré boli použité na odvodenie skúmanej informácie a informáciu o dôvere týmto predmetom, ktoré môžu byť overené. Napríklad chceme dokázať, že Joe miluje Mary. Na túto informáciu sme prišli tak, že sme našli na dôveryhodnej stránke dva dokumenty, z ktorých jeden tvrdí: ":Joe :loves :MJS" a druhý tvrdí: ":MJS daml:equivalentTo :Mary". Predpokladajme tiež, že máme kontrolné súčty týchto dokumentov osobne, od správcu stránky. Na overenie tejto informácie môžeme spísať tieto kontrolné súčty do lokálneho súboru a nastaviť zopár FOPL pravidiel na: „ak súbor ’a‘ obsahuje informáciu, že Joe miluje MSJ a má kontrolný súčet md5:0qrhf8q3hfh, potom zaznamenaj úspech A“ a „ak súbor ‘b‘ obsahuje informáciu, že MSJ je ekvivalentné Mary a má kontrolný súčet md5:0892t925h, potom zaznamenaj úspech B“ a „ak platí úspech A aj úspech B, tak Joe miluje Mary“. Uvedené pravidlá by mohli vyzerať nasledovne:

```
@prefix : <http://infomesh.net/2001/proofexample/#> .
@prefix p: <http://www.w3.org/2001/07/imaginary-proof-ontology#> .
@prefix log: <http://www.w3.org/2000/10/swap/log#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

p:ProvenTruth rdfs:subClassOf log:Truth .

# Proof

{ { <a.n3> p:checksum <md5:blargh>;
           log:resolvesTo [ log:includes { :Joe :loves :MJS } ] }
log:implies
{ :Step1 a p:Success } } a log:Truth .

{ { <b.n3> p:checksum <md5:test>;
          log:resolvesTo [ log:includes { :MJS = :Mary } ] }
log:implies
{ :Step2 a p:Success } } a log:Truth .
```

```
{ { :Step1 a p:Success . :Step2 a p:Success }
log:implies
{ { :Joe :loves :Mary } a p:ProvenTruth } } a log:Truth .
```

Súbor hovorí sám za seba a keď spracovanie s použitím CWM skutočne funguje, produkuje mienený výstup. CWM nemá možnosti na automatické overovanie kontrolných súčtov alebo digitálnych podpisov, ale je iba otázkou času kedy sa poriadny trust engine sémantického webu napíše.

2.3.13 Informácie okolia a SEM

O rozsahu informácií sa zatiaľ povedalo dosť málo, tak je dôležité zvážiť, čo to naozaj znamená mať lokálny alebo globálny systém. Vo všeobecnosti sú systémy malého a veľkého rozsahu a interakcie medzi týmito dvoma druhmi systémov bude pravdepodobne tvoriť veľkú časť transakcií, ktoré sa objavia v sémantickom webe.

2.3.13.1 Systémy veľkého rozsahu

Príkladom systémov veľkého rozsahu môžu byť dve spoločnosti, ktoré sa spájajú a potrebujú skombinovať svoje databázy. Ďalším príkladom môže byť vyhľadávací engine, ktorý spracúva výsledky založené na veľkom rozsahu dát. Tieto systémy všeobecne obsahujú rozsiahle databázy a sú potrebné vysoko výkonné pravidlá a procesory na zvládanie týchto databáz.

2.3.13.2 Systémy stredného rozsahu

Systémy stredného rozsahu sa pokúšajú porozumieť systémom veľkého rozsahu alebo môžu byť systémy malého rozsahu, ktoré sa spojili. Príkladom prvého môže byť spoločnosť, ktorá sa snaží čiastočne pochopiť dva veľké faktúrovacie systémy dostatočne na to, aby ich mohla používať spolu. Príkladom druhého zas sú dve skupiny tvoriace jazyky adresárov pokúšajúce sa vytvoriť jeden adresárový superjazyk.

2.3.13.3 Systémy malého rozsahu

O systémoch malého rozsahu sa hovorí menej. Týmito systémami rozumieme jazyky, ktoré sa budú používať primárne offline alebo skupiny dát, ktoré budú vymieňané len v obmedzenej miere, napríklad medzi priateľmi, oddeleniami alebo spoločnosťami.

Zdieľanie dát na lokálnej úrovni je veľmi dobrý príklad na to, ako môže byť sémantický web užitočný v nespočetných situáciách. V ďalšom si uvedieme, ako budú interakcie medzi rôzne veľkými systémami formovať kľúčovú časť sémantického webu.

2.3.13.4 SEM – SEmantic Memory (sémantická pamäť)

Koncept sémantickej pamäte bol prvýkrát navrhnutý Sethom Russelom, ktorý navrhol aby RDF výpisy osobných databáz, ktoré ľudia pozbierajú zo zvyšku sémantického webu (druh sémantického oblaku) boli nutné na udržanie koherentného pohľadu na dáta. SEM by bola pravdepodobne rozdelená na dáta, ktoré sú vlastné celému sémantickému webu (tzn. Schémy pre hlavné jazyky ako XML RDF, RDF Schema, DAML+OIL atď.), lokálne dáta, ktoré sú dôležité pre akékoľvek aplikácie sémantického webu, ktoré môžu byť práve používané (napr. informácie o logike menného priestoru pre CWM, ktorý je práve zabudovaný) a dáta, ktoré ľudia osobne používajú, publikujú alebo sa akokoľvek inak dostali do základného kontextu SEM. Vnútorná štruktúra SEM bude s veľkou pravdepodobnosťou podstatne zložitejšia ako obvyklé triplety v RDF, možno na úrovni štvoríc alebo dokonca pätíc. Polia navyše sú pre kontexty (StID) a možno sekvencie. Inými slovami, existujú rôzne spôsoby zoskupovania informácií v rámci SEM pre ľahkú údržbu a aktualizáciu. Napríklad by malo byť veľmi jednoduché vymazať akýkoľvek triplet, ktorý bol pridaný do konkrétneho kontextu cez zrušenie všetkých tripletov s daným StID. Mnoho prác na sémantickom webe sa sústredilo na to, aby boli dátové sklady (ako SEM) schopné spolupráce, čo je dobre, ale na druhej strane to viedlo k menšiemu rozsahu prác vedených na tom, čo sa vlastne deje v SEM ako takej, čo znamená, že reprezentácia štvoríc a pätíc v RDF je zatiaľ nevyriešená. Samozrejme, výroky môžu byť modelované nasledovne:

```
rdf:Statement rdfs:subClassOf :Pent .
_:s1 rdf:type :Pent .
_:s1 rdf:subject :x .
_:s1 rdf:predicate :y .
_:s1 rdf:object :z .
_:s1 :context :p .
_:s1 :seq "0" .
```

Ale je jasné, že určený formát pätíc bude vždy efektívnejší a predíde rizikám tvaru

```
:x :y :z :p "0" .
```

Tento jazyk taktiež potrebuje označenie predvoleného kontextu, ktoré indikuje základný kontext dokumentu. Základný kontext je miesto v dokumente, do ktorého sú (nekótované) výroky skompilované, je to konceptuálny informačný priestor, v ktorom sú všetky tieto výroky považované za pravdivé. Akákoľvek kótovaná informácia v dokumente (napr. pomocou kontextovej syntaxe Notation3) bude v inom (zrejme anonymnom) kontexte. Tim Berners-Lee zrejme používa "...#_formula" ako základný kontext pre dokument, čo je dosť

nepríjemný hack – čo ak niekto vytvorí prvok s rovnakým URI? Udržanie schopnosti spolupráce na tejto úrovni sémantického webu je dôležitá vec pre vývojárov a je potrebné sa ňou zaoberať.

2.3.14 Vývoj

Veľmi dôležitou stránkou sémantického webu je vývoj – prechádzanie z jedného systému do ďalšieho. Dve kľúčové časti schopnosti vývoja sú čiastkové porozumenie a premenlivosť. Teraz si povieme ako sa tieto vlastnosti sémantického webu prirodzene prejavujú pri zmene veľkosti systému.

2.3.14.1 Čiastkové porozumenie: z veľkých systémov na stredné

Koncept čiastkového porozumenia je veľmi dôležitý a často sa ho dá nájsť v už starších dokumentoch, ktoré boli vydané v čase, keď sa o sémantickom webe začalo hovoriť po prvýkrát. Príklad čiastkového porozumenia, keď sa prechádza z veľkého systému do stredného, je už spomenutý problém firmy, ktorá sa snaží porozumieť dvom faktúram od dvoch rôznych spoločností. Je dobre známe, že obe spoločnosti používajú podobné polia v ich faktúrach, takže firma snažiaca sa porozumieť im môže ľahko skompilovať nadradený zoznam výdavkov jednoducho vyťažením dát z dvoch jazykov faktúr. Ani jedna z fakturujúcich spoločností nemusí vedieť, že sa to deje. Tento príklad použil aj Tim Berners-Lee v jeho XML 2000 keynote.

2.3.14.2 Premenlivosť: z malých systémov na stredné

Príkladom na prechod z malých systémov na strené je vyššie uvedený príklad s dvoma adresárovými formátmi. Ich spojením vznikne nový, väčší a lepší adresárový formát. Každý, kto používal jeden zo starých formátov by si ho mohol zrejme skonvertovať na nový a tak dosiahnuť väčšiu schopnosť spolupráce. Toto sa deje všeobecne pri prechodoch z malých systémov na stredné, hoci sa často vyskytujú niektoré nevýhody alebo nekompatibility. Sémantický web tieto problémy rieši tým, že automatizuje 99% tohto procesu (dokáže skonvertovať pole A na pole B, ale nedokáže doplniť nové údaje, samozrejme, nové polia môžu byť na istý čas prázdne).

2.3.14.3 Napomáhanie schopnosti vývoja

Ako sa zaznamenáva vývoj jazykov? Toto je veľmi dôležitá a nástojčivá otázka, ktorú Tim Berners-Lee sumarizoval dosť elegantne:

„Where for example a library of congress schema talks of an "author", and a British Library talks of a "creator", a small bit of RDF would be able to say that for any person x and any resource y, if x is the (LoC) author of y, then x is the (BL) creator of y. This is the sort of

rule which solves the evolvability problems. Where would a processor find it? [...]"- *Semantic Web roadmap, Tim Berners-Lee*

Jednou z možností sú databázy tretích strán. Veľmi často je dosť nepraktické nechať LoC alebo BL zaznamenávať, že dve z ich polí sú rovnaké, takže táto informácia bude musieť byť zaznamenaná uznávanou treťou stranou. Jenou takouto skupinou, ktorá má na starosti výskum práve tohto problému je SWAG – Semantic Web Agreement Group. Táto skupina si kladie za cieľ zabezpečiť schopnosť spolupráce v sémantickom webe. Výsledkom je pravdepodobne prvý slovník sémantického webu od tretej strany, WebNS SWAG Dictionary.

2.3.15 Prepletanie, zložité, ale dôležité

Hoci je sémantický web ako celok len na začiatku, ľudia si ho začínajú všímať, začínajú publikovať informácie pomocou RDF a tak ich robia prijateľné pre sémantický web. Avšak málo sa robí pre spojenie informácií dokopy, to znamená, že so spojenia „sémantický web" časť „sémantický" postupuje pekne, ale časť „web" je veľmi slabá. Ľudia nepoužívajú výrazy druhých efektívne a keď ich používajú, robia to často preto, že sa bezcieľne snažia pomôcť, ale iba vytvárajú šum v procese. Pred použitím dát iných ľudí je dobré si zistiť, aké výhody z toho plynú vopred. Napríklad, iba tým, že použijeme výraz "dc:title" v našom RDF namiesto vlastného ":title", to ešte neznamená, že Dublin Core aplikácia je zrazu schopná porozumieť nášmu kódu. Tento tvar však znamená, že ak je náš výraz použitý tak, že ho bude potrebné znovu použiť alebo zmeniť v blízkej budúcnosti, môžeme z toho získať výhodu, lebo "dc:title" je tak bežne používaný výraz, že budeme schopní modifikovať súčasné pravidlá alebo čokoľvek iné. Ďalšou časťou tohto nedostatku môže byť problém podobný tomu, s ktorým sa stretol aj World Wide Web vo svojich začiatkoch. Prečo publikovať informácie na stránke, keď nie je žiadna iná stránka, ktorá by na nás odkázala alebo by sme sa na ňu odkázali my? Načo vytvárať stránku, keď tak málo ľudí má prehliadač? Načo písať prehliadač, keď je tak málo stránok? Niekto musí urobiť prvý krok, aby sa to rozbehlo a to je pomalý proces. Čo sa dá v tejto situácii robiť? Je tu možnosť, že sa vyrieši sama. Jedným z dobre známych princípov aplikácií sémantického webu je, že nemá význam znovu vynachádzať koleso. Ak niekto už vynašiel schému, ktorá obsahuje úplný a dobre porozumený a používaný súbor výrazov, ktoré budeme potrebovať aj v našej aplikácii, nemá zmysel snažiť sa znovu robiť, čo už bolo urobené. Toto je zrejme to, čo Tim Berners-Lee myslel, keď povedal, že výrazy sa objavia zo sémantického oblaku, že keď ľudia začnú používať výraz „zip" namiesto zaznamenávania, že môj výraz „zip" je ekvivalentný tvojmu výrazu „zip" , ktorý je ekvivalentný výrazu „zip" niekoho iného, všetci začnú používať to isté URI a spolupráca bude ohromne zlepšená.

3. Výsledky práce

Sémantický web je veľmi aktuálnou témou a v súčasnosti prebieha mnoho projektov, ktoré buď aplikujú už zavedené technológie, alebo sa snažia vyvinúť nové technológie a presadiť ich v praktických aplikáciách. Veľká časť projektov beží s podporou konzorcia W3C. Konzorcium samé vedie niekoľko projektov vo svojich pracovných skupinách [31].

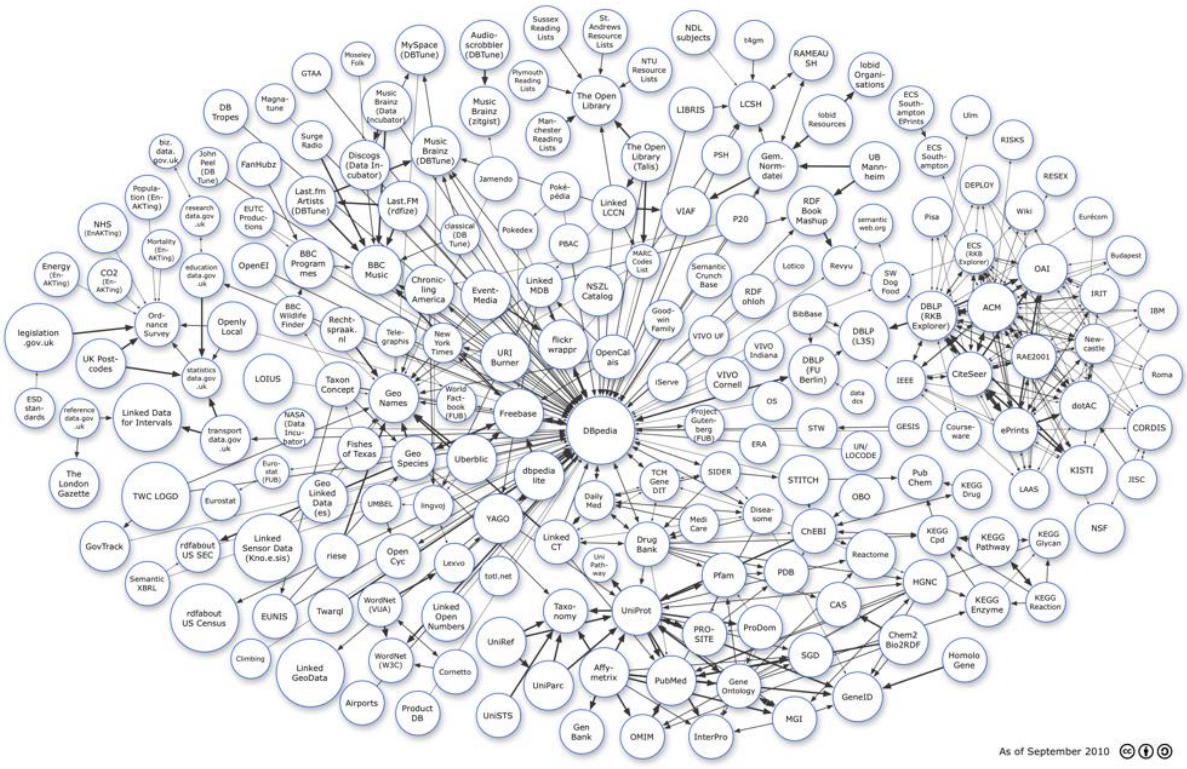
3.1 Stručný prehľad aktuálnych projektov

Počet prebiehajúcich projektov je veľmi veľký a preto ani v tejto práci nie sú zahrnuté všetky. Pre potreby práce sa budeme venovať niekoľkým projektom, ktoré sú pre prehľadnosť uvedené aj s odkazmi na stránky s ich opismi v nasledujúcej tabuľke.

Projekt	Stránka
Dbpedia	http://dbpedia.org/About
FOAF	http://www.foaf-project.org/
GoodRelations	http://www.heppnetz.de/projects/goodrelations/
SIOC	http://sioc-project.org/
SIMILE	http://simile.mit.edu/
NextBio	http://www.nextbio.com/b/nextbio.nb
Linking Open Data	http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData
OpenPSI	http://www.openpsi.org/
Meteor - S	http://lsdis.cs.uga.edu/projects/meteor-s/
SKOS	http://www.w3.org/2004/02/skos/
CNI	http://domino.research.ibm.com/comm/research_projects.nsf/pages/semanticweb.Semantic%20Web%20Projects.html
CRAFT	http://domino.research.ibm.com/comm/research_projects.nsf/pages/semanticweb.Semantic%20Web%20Projects.html
MARIO	http://domino.research.ibm.com/comm/research_projects.nsf/pages/semanticweb.Semantic%20Web%20Projects.html
Metadata Interoperability Framework leveraging Semantic Web	http://domino.research.ibm.com/comm/research_projects.nsf/pages/semanticweb.Semantic%20Web%20Projects.html
Semantic Web Portal Project	http://sw-portal.deri.at/

Projekt DBpedia je jedným z kľúčových projektov v oblasti sémantického webu. Kladie si za cieľ extrahovať informácie z Wikipedie v štruktúrovanej forme a sprístupniť ich na webe. Keďže aj Wikipedia je prístupná cez web (Web 2.0), predstava naplnenia tohto cieľa je schopnosť používateľov pýtať sa Wikipedie sofistikované dotazy (napríklad: „Zoznam všetkých hudobných skupín, ktoré vznikli na Slovensku po roku 1993.“) a prepojenie ostatných dát na internete s dátami Wikipedie. Splnenie uvedených cieľov by malo mať za následok nové spôsoby využitia dát na Wikipedii a nové mechanizmy na zlepšenie používania Wikipedie samej.

DBpedia je jednou časťou oveľa väčšieho projektu (aktivity), ktorý sa realizuje pod záštitou W3C a volá sa Linking Open Data. Okrem DBpedie je v ňom zahrnutých veľa ďalších datasetov, pričom DBpedia zaujíma kľúčové postavenie v rámci projektu. Pre ilustráciu uvedieme niekoľko, napríklad Geonames, MusicBrainz, WordNet, DBLP bibliography a mnoho ďalších. Cieľom projektu je urobiť dáta voľne prístupné pre každého. Tento cieľ sa má naplniť rozšírením webu o dáta publikované na webe ako otvorené datasety v RDF a prepojením týchto dát pomocou RDF linkov medzi objektmi z rôznych dátových zdrojov. RDF linky umožnia navigáciu z jedného dátového objektu v niektorom dátovom zdroji do súvisiacich objektov v iných dátových zdrojoch s použitím prehliadača sémantického webu. Tieto linky môžu byť použité aj webovými vyhľadávačmi (crawlermi), ktoré poskytnú možnosti na sofistikované vyhľadávanie a dotazovanie nad prehliadanými dátami. Ako výsledok dotazovania dostávame štruktúrované dáta a nie len odkazy na HTML stránky a preto môžu byť použité v ďalších aplikáciách. Pre lepšiu predstavu o rozsiahlosti a rozšírenosti projektu uvedieme obrázok, ktorý zobrazuje projektom publikované a prepojené datasety zo septembra 2010.



Prebrané z [38] 19. 3. 2011

V rámci W3C Semantic Web Activity prebieha aj ďalší projekt, ktorý má názov SKOS (Simple Knowledge Organization System). Práce na projekte sa zameriavajú na vývoj špecifikácií a štandardov na podporu používania systémov na organizáciu znalostí (knowledge organization system – KOS) ako napríklad thesauri, kalsifikačné schémy a taxonómie v rámci sémantického webu. SKOS poskytuje štandardný spôsob reprezentácie systémov KOS s použitím RDF. Zakódovanie informácií do RDF umožňuje počítačovým aplikáciám odovzdávať si tieto informácie vo forme, v ktorej sú schopné spolupráce. Použitie RDF tiež umožňuje použitie KOS systémov v distribuovaných a decentralizovaných metadátových systémoch (decentralizované metadáta sa stávajú typickým scenárom, kde poskytovatelia služieb chcú pridať hodnotu do metadát pozbieraných z viacerých zdrojov). V súčasnosti sú SKOS špecifikácie publikované ako W3C odporúčania, čo značí, že sú v stabilnom stave.

Ďalším významným projektom je Friend Of A Friend (FOAF). Tiež pracuje s myšlienkou prepojených dát, ale jeho cieľom je vytvorenie strojovo čitateľných stránok, ktoré by opisovali ľudí a veci a vzťahy, ktoré medzi sebou majú. Tento cieľ by sa mal dosiahnuť technológiou, ktorá uľahčuje zdieľanie informácií, prenos informácií medzi webovými stránkami a automatické rozšírenie, spájanie a znovu použitie týchto informácií na webe. Myšlienka projektu je známa a používaná aj v súčasnosti, hoci nemá podobu technológií sémantického webu. Jedna z takýchto stránok je aj na Slovensku [47].

S projektom FOAF súvisí projekt SIOC (Semantically-Interlinked Online Communities). Jeho cieľom je umožniť integráciu informácií z online komunít (napríklad sociálne siete). SIOC poskytuje ontológiu pre reprezentáciu dát z tzv. „sociálneho webu“ pomocou RDF (všetky SIOC dáta používajú RDF ako základný formát a dá sa s nimi pracovať ako s RDF). V poslednej dobe významne preniká do praxe – ontológia je používaná vo viacerých komerčných aj open-source aplikáciách. SOIC ontológia je často používaná spolu so slovnou zásobou FOAF na vyjadrenie osobných profilov a informácií v sociálnych sieťach. Štandardizovaním SIOC-u pre vyjadrovanie obsahu vytváraného používateľmi týchto stránok, SIOC umožňuje vznik nových scenárov použitia pre dáta online komunít a dovoľuje vybudovanie inovatívnych sémantických aplikácií na báze existujúceho sociálneho webu. Bolo navrhnutých mnoho aplikácií, ktoré používajú SIOC a tu si uvedieme iba niekoľko príkladov pre lepšiu predstavu penetrácie SIOC-u do reálneho používania. SIOC je navrhnutý na zobrazenie dát o obsahu a štruktúre komunitných webstránok v strojovo čitateľnej forme a preto boli vytvorené rôzne nástroje, exportéry a služby na zobrazenie týchto dát z existujúcich online komunít. Tieto nástroje môžeme rozdeliť na:

1. API (Application Programming Interface)
    a. SIOC Export API pre PHP – na podporu tvorby SIOC exportérov a uľahčenie manipulácie s SIOC dátami pomocou PHP objektov a metód
    b. SIOC API pre Javu – pre každý objekt v SIOC ontológii toto API generuje triedy s prepojeniami na iné objekty
    c. SIOC API pre Perl
    d. RDF API pre Durpal
2. Exportéry (z weblogov, fór a CMS [48])
    a. WordPress SIOC Exporter
    b. Dotclear SIOC Exporter
    c. b2evolution SIOC Exporter
    d. Drupal SIOC Exporter
    e. phpBB SIOC Exporter
    f. vBulletin SIOC Exporter
    g. BlogEngine.NET
3. Iné exportéry
    a. OpenLink Data Spaces Modules
    b. Talk Digger
    c. SWAML

d. Twitter2RDF
e. a iné

Po vytvorení SIOC dát má ich použitie viac podôb. Tieto podoby sú tiež súčasťou penertácie SIOC-u (podrobnejšie na [35]):

1. dotazovanie nad SIOC dátami
2. prehliadanie SIOC dát (vo forme crawlingu aj browsingu)
3. používanie SIOC-u pre nové dáta
4. znovupoužitie SIOC dát

Na oblasť obchodu a služieb sa orientuje projekt GoodRelations. Jeho hlavnou problémovou oblasťou je internetový obchod, najmä vyhľadávanie relevantných ponúk na internete. Technológiou, ktorá je výsledkom projektu je jazyk, resp. slovná zásoba (pomenovanie technológie GoodRelations nie je jednotné), ktorá umožňuje veľmi precízne opísať produkty alebo služby ponúkané firmami na internete. Prínosom pre firmy je zlepšenie viditeľnosti ich ponúk pri použití najnovších verzií webových vyhľadávačov a odporúčacích systémov. Vďaka podrobnému opisu ponuky pomocou GoodRelations sa firma dostane vždy na prvé miesto zoznamu výsledkov hľadania pre používateľov, ktorí majú zodpovedajúce potreby (alebo preferencie), a teda hľadajú práve produkt danej firmy.

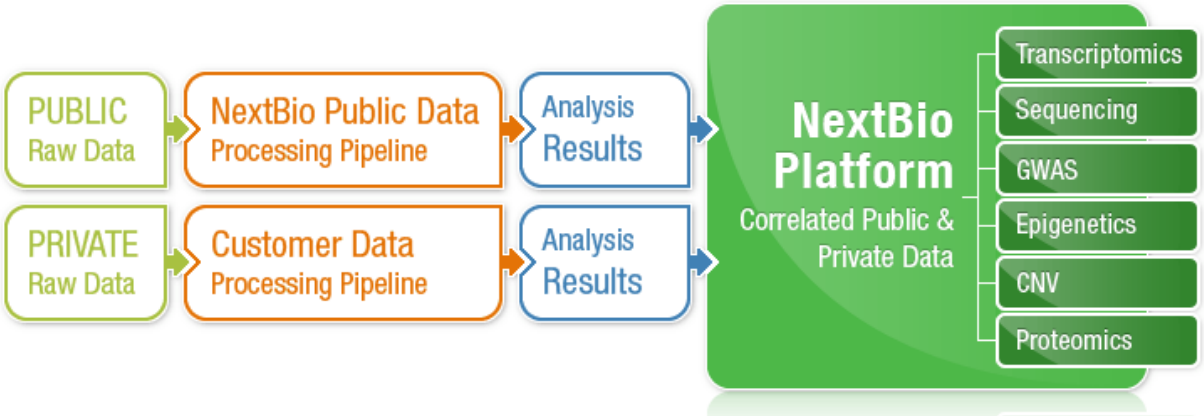
Pozornosť si zaslúži aj projekt z MIT (konkrétne MIT Libraries a MIT CSAIL), SIMILE (Semantic Interoperability of Metadata and Information in unLike Enviroments). Cieľom tohto projektu je zlepšenie schopnosti spolupráce medzi digitálnymi aktívami, schémami, ontológiami, metadátami a službami. Základným problémom v tejto oblasti je fakt, že kolekcie, ktoré majú spolupracovať sú rozptýlené medzi jednotlivcov, komunity a inštitúcie. Prekonanie tejto prekážky spočíva v schopnosti poskytnúť službám koncových používateľov aktíva, schémy, ontológie a metadáta, ktoré sa nachádzajú u spomínaných entít. SIMILE chce posilniť a rozšíriť DSpace [49] a tým zlepšiť jeho podporu pre ľubovoľné schémy a metadáta, najmä apikáciou RDF a techník sémantického webu. Projekt tiež chce implementovať architektúru rozširovania digitálnych aktív založenú na webových štandardoch. Táto architektúra rozširovania bude poskytovať mechanizmus na pridávanie užitočných „pohľadov“ na konkrétny digitálny artefakt (t. j. aktívum, schéma alebo inštancia metadát) a zviazať tieto pohľady so spotrebúvajúcimi službami. Projekt SIMILE sa sústreďuje na dobre definované prípady z reálneho sveta, konkrétne z oblasti knižníc. V súčasnosti prebiehajú paralelné práce na rozmiestnení DSpace do viacerých významných výskumných knižníc, čo by malo viesť k verejnej demonštrácii užitočnosti a pripravenosti nástrojov a techník

sémantického webu. SIMILE vydáva svoje verzie pod licenciou typu BSD (čo znamená open source dostupnosť).

Medzi významné projekty sa radí aj projekt spoločnosti NextBio. Je to projekt realizovaný komerčne, ako produkt, ktorý je určený na predaj. Jeho základom je technológia založená na ontológiách. Projekt sa zameriava na oblasť medicíny, biológie, genetiky a vied o živote. Poskytuje platformu pre rôzne spoločnosti aj jednotlivcov, ktorí pracujú v niektorej z oblastí zamerania projektu, či už ako výskumné inštitúcie alebo prakticky aplikujú dostupné poznatky v diagnostike a liečení pacientov. Platforma NextBio zhromažďuje dáta od všetkých používateľov do jednej súvislej databázy a sprístupňuje ich pre všetkých (to znamená, že výsledky výskumu niektorej inštitúcie sú okamžite použiteľné aj vo všetkých ostatných). Ďalšou úlohou technológie použitej v NextBio je prekladanie komplexných vysoko priepustných dát do formy informácií, ktoré sú jednoducho použiteľné na vytvorenie a testovanie nových hypotéz. Táto funkcia podporuje pochopenie funkcií génov, priebeh chorôb a terapeutických zákrokov. Platforma NextBio je dostupná v troch produktových verziách:

1. NextBio Basic – webová aplikácia, ktorá je zdarma a poskytuje iba prístup k dátam uloženým v databáze
2. NextBio Professional – pridáva možnosť využívať vnútorne generované dáta
3. NextBio Enterprise – ponúka úplnú funkcionalitu platformy, možnosť spolupráce medzi výskumnými (a inými) skupinami a možnosť vytvárať synergie medzi verejne dostupnými a vlastnými dátami

Produkt spoločnosti NextBio je v súčasnej dobe používaný v mnohých výskumných inštitúciách z oblastí medicíny, vied o živote, farmácie a i. Myšlienku produktu NextBio výstižne zobrazuje nasledujúci obrázok.



Prebrané z [37] 19. 3. 2011

Zo sféry projektov realizovaných súkromnými firmami určite stoja za zmienku projekty, ktoré prebiehajú vo firme IBM. IBM ich robí hneď niekoľko a postupne si ich tu predstavíme. Prvým projektom z dielne IBM je projekt CNI (Connection Network Intelligence). CNI zahŕňa inévačnú technológiu na analýzu vzťahov. Zameriava sa na podporu rozhodovania pomocou odhaľovania menej očividných vzťahov medzi entitami na finančnom trhu. Vzhľadom na to, že projekt sa zameriava na poskytnutie silného a ľahko použiteľného analytického nástroja pre ľudí z oblasti obchodu, CNI pracuje na doménovej ontológii, ktorá by pomohla vybudovať dátová model pre finančné tvrdenia a akciové odhalenia. To by umožnilo týmto ľuďom formalizáciu ich znalosti o detekcii podvodov prostredníctvom pravidiel založených na doménovej ontológii. Odvodzovací mechanizmus CNI bude spracúvať tieto pravidlá a robiť dotazy nad faktami uloženými v relačných databázach. Výsledky dotazovania budú poskytnuté používateľom na podporu rozhodovania. Táto technológia už bola použitá v niektorých ázijských trhoch akcií. IBM pracuje aj na ďalšom projekte na podporu rozhodovania, ale tento nerieši otázky finančného trhu, ale oblasť znalostí. Projekt má názov CRAFT (Collaborative Reasoning and Analysis Framework and Toolkit). Táto technológia je určená pre analytikov. CRAFT umožní analytikom reprezentovať svoje znalosti o situácii, zaznamenávať otázky a hypotézy a vytvárať prieskumy na nové informácie z interných databáz a verejných zdrojov. Na to sa použijú koncepty a prípady vybraté zo stále sa vyvíjajúcej ontológie. CRAFT obsahuje prostriedky na informovanie analytikov o nových informáciách a výsledkoch prieskumov, znovu použitie informácií pridaných ostatnými analytikmi a spoluprácu medzi prieskumami a funkciami. Filozofia projektu CRAFT bola ovplyvnená wiki softvérom, v ktorom používatelia jednoducho pridávajú, upravujú a prepájajú informácie o rôznych témach a neustále zušľachťujú popisy tém a porovnávajú ich s predchádzajúcimi verziami. Na rozdiel od bežných wiki, informácie v CRAFTe sú sémanticky zakódované a pridávané cez kombináciu grafických a formulárových vstupov. Mienenými používateľmi systému sú znalostní pracovníci, ale nie znalostní inžinieri, takže projekt sa zameriava na jazyk na reprezentáciu znalostí, ktorý vyvažuje schopnosť vyjadrovania a jednoduchosť použitia. Ďalším z projektov IBM je projekt MARIO (Mashup Automation with Run-time Invocation and Orchestration). Tento projekt sa snaží odpovedať na výzvu v systémoch založených na komponentoch. Tou výzvou je automatizovaná kompozícia aplikácií zo sady základných komponentov ako odpoveď na vysoko náročné požiadavky. MARIO si kladie za cieľ podporiť rôzne stupne od manuálnej až po automatickú kompozíciu v rôznych druhoch systémov. MARIO sa predovšetkým sústreďuje na prúdovo založené aplikácie spracovania informácií, ktoré sú

súpravy kompnentov zostavené v orientovanom acyklickom grafe komponentov (v štýle čiernej skrinky) pospájaných prepojeniami dátových prúdov. Takéto prúdovo založené aplikácie sú bežné v rôznych oblastiach, napríklad „Service Oriented Architectures, Event-Driven Systems, Data Mashups, Stream Processing Systems, Extract-Transform-Load based systems and the Grid“ [44]. V roku 2008 sa začal projekt Metadata Interoperability Framework leveraging Semantic Web. Projekt pracuje na poprepájaní rôznych druhov IT metadát do Metadata Interoperability Framework. Projekt zatiaľ dosiahol:

- priraďuje URI každej metadátovej entite bez ohľadu na to, kde sa nachádza referenčné kópia entity
- s použitím Linked Open Data http CRUD rozhrania poskytuje RDF pohľad každej metadátovej entity
- zbiera všetky súčasné metadátové entity a pokúša sa identifikovať vzťahy medzi nimi
- ukazuje, že moderné webové technológie (ako napr. Mashup authoring alebo Smart Search) môžu získať miesto popri týchto metadátových entitách
- používa ľahkú OWL odvodzujúcu implicitné vzťahy, ktorá obsahuje flexibilný systém klasifikácie entít

V IBM sa vyvíja aktivita aj smerom k webovým službám, konkrétne sa sústreďuje na služby sémantického webu. Zaoberá sa problémom nájdenia zodpovedajúcej služby, ktorého riešenie závisí od schopnosti poskytovateľov opísať možnosti svojich služieb a od schopnosti žiadajúcich opísať svoje požiadavky v jednoznačnej a strojovo interpretovateľnej forme. Výsledkom tejto aktivity je stroj, ktorý vyhľadáva služby zodpovedajúce požiadavkám používateľov. Ak sa nenájde služba, ktorá by zodpovedala požiadavkám, systém použije plánovacie algoritmy umelej inteligencie na poskladanie služieb tak, aby zodpovedali zadaným požiadavkám. IBM pracuje aj na projekte SHER (Scalable Highly Expressive Reasoner). SHER je „OWL reasoner“ (preklad slova reasoner je logicky uvažujúci človek, namiesto slova človek by sa malo v tomto prípade použiť slovo stroj), ktorý je navrhnutý na poskytovanie sémantického dotazovania nad veľkými relačnými databázami s použitím ontológií OWL. SHER ponúka ontologické analýzy ontológií s vysokou schopnosťou vyjadrovania. Vlastnosti tejto technológie sú uvedené na [42]. SHER bol úspešne použitý v medicíne, na priraďovanie elektronických záznamov pacientov ku kritériám klinických testov, na vytvorenie lepšieho vyhľadávania v biomedicínskej literatúre a na vyčistenie výsledkov analytických nástrojov pre text.

Vo Veľkej Británii prebieha ďalší projekt, na ktorom spolupracujú University of Southampton a vláda Spojeného kráľovstva za výraznej podpory fondu JISC. Projekt sa volá OpenPSI. Práce na projekte vedie National Archive a ich cieľom je vytvoriť a odskúšať novú formu komunitne poskytovanej informačnej služby. Táto služba má používať štandardy sémantického webu na poskytnutie integračného bodu pre informácie pochádzajúce od vlády a vytvárať priestor pre interakciu medzi poskytovateľmi vládnych informácií, komunitou výskumníkov, ktorí potrebujú prístup k týmto informáciám a novou formou informačných sprostredkovateľov, tvorcov mashupov (mashup – [50]).

Medzi univerzitné projekty sa radí aj projekt METEOR-S: Semantic Web Services and Processes. Tento projekt prebieha na University of Georgia, konkrétne v LSDIS Lab (Large Scale Distributed Information Systems). METEOR-S vychádza zo súčasnej situácie, kde si uvedomuje najmä veľký rozmach webových služieb a architektúry orientovanej na služby, ktoré ponúkajú atraktívnu základňu na realizáciu dynamických architektúr, ktoré by odrážali dynamickosť a neustále premeny obchodného prostredia. Prostredie (prostredie vývoja internetových technológií) akceptuje štandardy ako Business Process Execution Language for Web Services (BPEL4WS), Web Service Description Language (WSDL) a Simple Object Access Protocol (SOAP), čo otvára dvere pre webové služby, aby ponúkli nízko-nákladovú a bezprostrednú integráciu s ostatnými aplikáciami. Cieľom projektu je rozšírenie týchto štandardov o technológie sémantického webu, čím by sa dosiahla väčšia dynamika a rozšíriteľnosť. Konkrétne sa METEOR-S sústreďuje na pridanie sémantiky do WSDL a UDDI (z čoho by mala vzniknúť nová verzia WSDL, WSDL-S s podporou sémantickej reprezentácie), pridanie sémantiky do BPEL4WS. Ďalej projekt prináša tému poloautomatického prístupu k anotácii webových služieb, opísaných s použitím WSDL. Projekt sa teda snaží definovať a podporiť kompletný životný cyklus procesov sémantického webu, v ktorom identifikoval viaceré štádiá. Tieto štádiá sú podľa [40]:

- Semantic Annotation and Publicaton of Web Services
- Abstract Process Creation
- Semantic Discovery of Web Services
- Orchestration/Compostion of Web Services

Uvedené štádiá pokrývajú rôzne časti projektu METEOR-S. Ďalšou problémovou oblasťou, ktorou sa projekt zaoberá je Quality of Service (QoS). V obchodných procesoch sa na podporu využíval tzv. „workflow management system“. Ako podniky preberajú nové modely pracovania, vznikajú nové výzvy pre workflow systémy. Takýmito výzvami sú podpora pre adekvátny manažment kvality služieb a umožnenie kompozície workflow

aplikácií zahŕňajúcich webové služby (vzhľadom na to, že dobrý manažment kvality služieb priamo ovplyvňuje úspech firmy). Tieto nové workflow modely sú odlišné od tradične používaných workflow systémov kvôli rozsahu a rozdielnosti dostupných webových služieb. Z toho vyplývajú dva hlavné problémy:

1. ako efektívne objavovať webové služby
2. ako zabezpečiť ich spoluprácu

Na riešenie týchto problémov boli vyvinuté QoS model a matematický model, ktoré boli hodnotené spustením v skupine produkčných workflow systémov v oblasti genetiky.

Na oblasť prístupu k informáciám a ich využitia sa zameriava projekt Semantic Web Portal. Za cieľ si kladie vytvorenie portálu sémantického webu, na ktorom by demonštroval zrelosť technológií sémantického webu v praktickej aplikácii. Projekt spolupracuje s partnermi z odvetvia a snaží sa vyvinúť technológiu, ktorá bude použiteľná na viaceré portály rôznych komunít. Projekt stavia na dvoch základných pilieroch:

1. použiteľnosť portálu pre skúsených aj neskúsených používateľov
2. podpora komunít umožňujúca spoluprácu v rámci aj medzi skupinami ľudí (zoskupených v sieťach)

Portálová podpora komunít nebude zahŕňať len podporu v pasívnom prijímaní informácií, ale aj podporu v aktívnej publikácii a spolupráci medzi členmi komunít, pomocou zhodných slovných zásob. Snaha je v rámci portálu priviesť dokopy siete (v zmysle skupiny ľudí) a posilniť spoluprácu medzi nimi. Prípadová štúdia tejto technológie sa uskutoční na komunitnom portáli komunity sémantického webu, semanticweb.org. Cieľom tejto štúdie je spojiť výskumné skupiny, projekty, softvérových vývojárov a komunity používateľov z oblasti sémantického webu. Preloženie týchto cieľov do technologických termínov je vytvorenie uspokojujúceho prostredia pre manažment ontológií potrebného pre sémanticky rozšírené komunitné portály a zároveň presadiť rozsiahle využitie technológií sémantického webu pre posilnenie prostriedkov na spracovanie informácií a vytvoriť spôsob pre sémantickú spoluprácu medzi rôznymi komunitami a dokonca aj rôznymi portálmi sémantického webu.

3.2 Bližší pohľad na vybrané projekty

V tejto časti sa bližšie pozrieme na niektoré projekty, ktoré boli spomenuté vyššie, ale je vhodné venovať im viac pozornosti.

Zmyslom tvorby sémantického webu je vytvorenie nového štandardu, v ktorom by boli všetky dáta rozumne prepojené a ošetrené tak, aby s nimi vedeli pracovať počítače v štýle inteligentných agentov. Prepájaniu dát sa venuje rozsiahla aktivita vedená W3C, Linking

Open Data, ktorá pozostáva z mnohých menších projektov. Medzi týmito projektmi zaujíma významné postavenie DBpedia. Ako už bolo spomenuté, cieľom DBpedie je získať informácie zo súčasnej Wikipedie a dať ich do formy, s ktorou pracuje sémantický web kvôli zlepšeniu ich použiteľnosti. Základom projektu je báza poznatkov (poznatky organizované do báz poznatkov majú čoraz väčšiu úlohu v rozširovaní inteligencie webu, vyhľadávaní a integrácii informácií). Väčšina súčasných báz poznatkov pokrýva iba špecifické oblasti, je vytvorená malými skupinami znalostných inžinierov a vyžaduje veľké náklady na udržanie si aktuálnosti. Oproti tomu, Wikipedia sa vyvinula do jedného z hlavných zdrojov poznatkov ľudstva a je udržiavaná veľkým počtom prispievateľov. DBpedia vyberá z tohto zdroja štruktúrované informácie a sprístupňuje ich na webe. Báza poznatkov DBpedie v súčasnosti opisuje viac ako 3,5 milióna vecí, z ktorých 1,67 milióna vecí je klasifikovaných v súvislej ontológii a dokopy pozostáva z 672 miliónov RDF tripletov. Okrem tripletov z N-Triples datasetov, v DBpedii sú aj N-Quads datasety, ktoré obsahujú ešte URI pôvodu pre každé tvrdenie (pre každý triplet). URI pôvodu pozostáva z URI článku na Wikipedii, z ktorého tvrdenie pochádza a z parametrov označujúcich presný riadok, na ktorom sa nachádza dané tvrdenie. Tieto parametre sú:

- absolute line – číslo riadku v článku (prvý riadok má číslo 1)
- relative line – číslo riadku v sekcii článku
- section – označenie sekcie článku

Ontológia DBpedie je plytká viacdoménová ontológia, ktorá bola manuálne vytvorená z najčastejšie používaných infoboxov (infobox – tabuľka vpravo hore na stránke Wikipedie) vo Wikipedii. Infoboxy obsahujú špecifické informácie o veciach a preto sú cenným zdrojom štruktúrovaných informácií, ktoré môžu byť použité na dotazy na Wikipediu. DBpedia momentálne používa tieto datasety odvodené z infoboxov Wikipedie (podrobnejšie o datasetoch DBpedie na [32]):

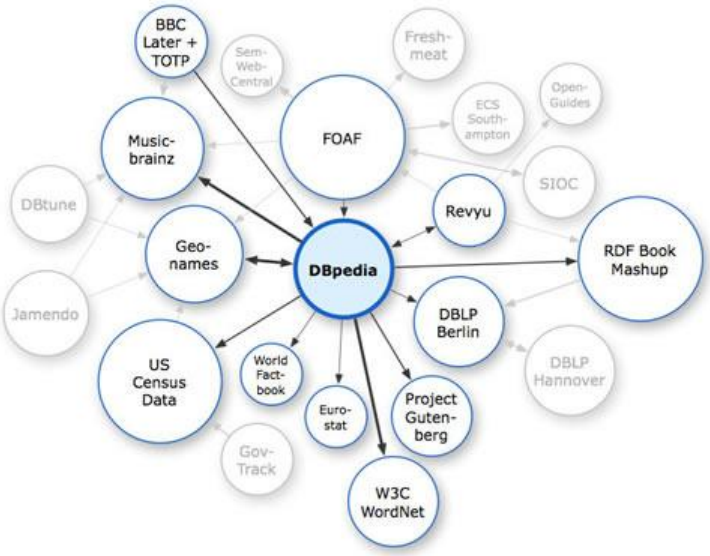
1. Infobox Dataset
2. Infobox Ontology, ktorá sa delí na
    a. Ontology Infobox Types
    b. Ontology Infobox Properties
    c. Ontology Infobox Properties (Specific)

Táto ontológia má v súčasnosti 272 tried usporiadaných do hierarchie a opísaných 1300 rôznymi vlastnosťami. Dataset DBpedie obsahuje značky a abstrakty pre tieto veci (všetkých 3,5 mil.) v 97 jazykoch. Dataset obsahuje tiež HTML a RDF linky na externé stránky, resp.

zdroje. Existujú dva typy HTML linkov. Prvým je dbpedia:reference, ktorý odkazuje na stránky o opisovanej veci. Druhým typom HTML linkov je foaf:homepage, ktorý ukazuje na stránku považovanú za oficiálnu domovskú stránku (ak existuje). RDF linky sú reprezentované pomocou owl:sameAs. Báza poznatkov DBpedie má niekoľko výhod oproti iným bázam poznatkov a to, že pokrýva mnoho oblastí, reprezentuje dohovor komunity, automaticky sa vyvíja so zmenami Wikipedie, je viacjazyčná a umožňuje používateľom pýtať sa sofistikované dotazy (napríklad „Všetky slovenské univerzity s počtom študentov viac ako 5000.“). Dotazy na DBpediu sa robia v jazyku SPARQL. Každá vec v DBpedii je identifikovaná pomocou URI vo forme http://dbpedia.org/resource/Name, kde Name odvodené od URL zdrojového článku na Wikipedii, ktoré má formu http://en.wikipedia.org/wiki/Name. Týmto je každá vec prepojená s anglickou verziou článku na Wikipedii. Všetky zdroje v DBpedii sú opísané skupinou vlastností. Základné vlastnosti sú značka, dlhý a krátky abstrakt v anglickom jazyku, odkaz na korešpondujúci článok na Wikipedii a odkaz na obrázok zobrazujúci danú vec (ak je k dispozícii). V prípade, že daná vec je opísaná vo viacerých jazykoch na Wikipedii, sú pridané krátke a dlhé abstrakty a odkazy na zodpovedajúce články. DBpedia ponúka tri rôzne klasifikačné schémy pre veci:

1. klasifikácia Wikipedie
2. YAGO klasifikácia
3. Word Net Synset Links

DBpedia je prístupná online cez verejný koncový bod SPARQL a ako Linked Data (viac o spôsobe prístupu na [32]). Linked Data je metóda ako publikovať dáta ne webe a navzájom ich prepájať. Tieto dáta sú potom prístupné použitím prehliadača sémantického webu rovnako, ako sú tradičné dokumenty prístupné pomocou tradičných webových prehliadačov. Rozdiel je, že prehliadač sémantického webu umožňuje nasledovať aj RDF linky, ktoré môžu použiť aj vyhľadávacie služby. Prepojenie DBpedie s ďalšími datasetmi ilustruje obrázok (neposkytuje úplné zobrazenie, prebratý z [32] 22. 3. 2011).



Ako jeden z vedúcich projektov aktivity Linking Open Data, kladie DBpedia základ pre tzv. „Web of Data“. Mnoho poskytovateľov dát začína publikovať a prepájať dáta podľa Linked Data princípov Tima Berners-Leea. Výsledkom je, že Web of Data pozostáva z obrovského počtu RDF tripletov, ktoré opisujú rôzne oblasti. Vzhľadom na to, že DBpedia je tiež robená ako Linked Data a definuje URI pre veľa konceptov, začali títo poskytovatelia dát nastavovať svoje RDF linky na DBpediu, čím sa z DBpedie stal centrálny prepájací bod pre Web of Data. Keďže DBpedia je odvodená z Wikipedie, je distribuovaná pod rovnakou licenciou ako Wikipedia, čo v súčasnosti znamená dvojité licencovanie (Creative Commons Attribution-ShareAlike 3.0 a GNU Free Documentation Licence).

Jednou z vlastností sémantického webu je schopnosť počítača pracovať s našimi dátami a uvažovať nad nimi (v zmysle, že im rozumie). Zároveň mnohí autori vyzdvihujú schopnosť prepájať dáta. Tu prichádza do pozornosti projekt FOAF, ktorý sa zaoberá prepojením osobných údajov v strojovo čitateľnej podobe. Zmyslom tohto projektu je vytvorenie stránok, ktoré opisujú ľudí, veci, miesta a pod. a vyjadrujú aj, aké majú medzi sebou vzťahy. Víziou projektu je web, v ktorom si môžeme stránky a nástroje, aké chceme bez toho, aby sme sa oddelili od priateľov, ktorí si vybrali niečo iné. FOAF umožňuje zdieľať a prepájať informácie z rôznych zdrojov a narábať s nimi novými spôsobmi. FOAF začal v roku 2000 ako experimentálny projekt a neskôr sa podstatnejšie rozšíril. Jeho základnou myšlienkou je vytváranie prepojení medzi objektmi. K tomu nám FOAF poskytuje základný aparát na asistenciu pri oznamovaní o prepojeniach s vecami, ktoré s nami súvisia. Použitím FOAF pomáhame strojovému porozumeniu našej stránky a tým získavame informácie o vzťahoch, ktoré spájajú ľudí, miesta a veci opísané na webe. Na spracovanie informácií z našej stránky

(aj zo všetkých ostatných stránok) FOAF používa technológiu RDF. FOAF opisuje veci s použitím jednoduchých ideí inšpirovaných webom. V týchto opisoch sú rôzne druhy vecí a odkazov, ktoré sa nazývajú vlastnosti (properties). Typy vecí, o ktorých sa vo FOAF hovorí sú triedy (classes). FOAF je teda definovaný ako slovník výrazov, ktoré sú buď triedou alebo vlastnosťou. Ostatné projekty majú vlastné sústavy tried a vlastností, z ktorých mnohé sú prepojené s tými, ktoré sú definované vo FOAF. Opisy vo FOAF sú publikované ako prepojené dokumenty na webe. Výsledkom toho je sieť dokumentov, ktoré opisujú sieť ľudí (a vecí). Hoci tieto dokumenty nie sú vždy v zhode alebo sú zavádzajúce, majú jednu dôležitú vlastnosť, a to, že ľahko môžu byť spájané, čím umožňujú čiastočným a decentralizovaným opisom byť kombinované zaujímavým spôsobom. FOAF používa množstvo výrazov, ktoré opisujú ľudí, skupiny alebo dokumenty. Rôzne druhy aplikácií môžu používať alebo ignorovať rôzne časti FOAF (napríklad budeme ignorovať archaické a historické časti a ostatné výrazy rozdelíme na výrazy, ktoré majú zmysel iba na webe a na výrazy univerzálne použiteľné pri prepájaní ľudí a informácií). FOAF výrazy sa delia do kategórií:

1. Core – tieto výrazy opisujú ľudí a skupiny a sú nezávislé na čase a technológii, čiže môžu byť použité na opis ľudí v súčasnosti, kultúrneho dedičstva alebo digitálnej knižnice
2. Social Web – výrazy opisujúce internetové účty, adresáre a iné aktivity založené na webe
3. Linked Data utilities – výrazy, ktoré boli použité na najmä vzdelávacie účely alebo technické termíny, ktoré podporujú širšiu snahu o prepájanie informácií

V súčasnosti je vo FOAF použitá slovná zásoba verzie 0.98.

Často adresovaný problém súčasného webu je nedokonalosť webového vyhľadávania. V mnohých článkoch týkajúcich sa sémantického webu je tento problém uvádzaný ako prvý príklad, kde by táto technológia mohla priniesť výrazné zlepšenie. Nedokonalosť vyhľadávania sa veľmi vypuklo prejavuje vo vyhľadávaní firiem a obchodov. Odpoveďou na tento problém by mal byť projekt GoodRelations. Cieľom projektu je vytvoriť jazyk, ktorý by umožňoval konkrétnu špecifikáciu ponuky tovarov a služieb firiem na webe. GoodRelations vytvára malé dátové balíky, ktoré opisujú daný produkt, jeho vlastnosti, cenu, prevádzky firmy, kde je dostupný, ich otváracie hodiny, možnosti platby a pod. Takýto balíček sa jednoducho pridá do existujúcej stránky firmy. GoodRelations je teda produkt, ktorý môže byť pridaný do existujúcich statických a dynamických stránok a môže byť spracovaný počítačmi. To zvyšuje viditeľnosť firmy pri vyhľadávaní na webe. Táto zvýšené viditeľnosť je

dosiahnutá precíznym opisom produktu a schopnosťou počítača priradiť ho k vyhľadávaniu so zodpovedajúcimi atribútmi (potrebami používateľa). GoodRelations sa vytvorí pomocou nástroja GoodRelations Snippet Generator, v ktorom firma opíše svoje produkty a tento nástroj vygeneruje kus HTML kódu, ktorý si firma pridá do svojej stránky. GoodRelations má nasledujúce vlastnosti:

- Jedna slovná zásoba pre veľa cieľov – GoodRelations bude použiteľná pre tradičné aj nové (sémantické) vyhľadávače
- Multi-syntax – GoodRelations dáta môžu byť publikované ako RDFa v HTML, Microdata, RDF/XML, Turtle, dataRSS a N3
- Minimálny dopad na veľkosť stránky a dobu načítania
- Informácie o spoločnosti a predajniach aj s otváracími hodinami
- Detailné informácie o cene zahŕňajúce aj zľavy a rozpätia cien
- Možnosti platby a dodania spolu s osobnými poplatkami
- Produktové modely, varianty, náhradné diely
- Premenlivá úroveň detailov – jednoduché prípady ostanú jednoduché, komplexné scenáre sú tiež podporované
- Vhodné pre B2B aj B2C trhy
- Vhodné pre priemysel, spotrebnú elektroniku, súčiastky aj služby
- Podporuje zoznamy prianí a vyhlasovanie súťaží
- Možná kombinácia s Open Graph protokolom Facebooku

GoodRelations používa vlastnú ontológiu s veľkým počtom definovaných termínov z oblasti obchodu (ich definície na [34] – publications). Táto ontológia obsahuje konceptuálne entity aj pre n-árne vzťahy (OWL podporuje iba binárne vzťahy). Pre vytvorenie tejto ontológie autori navrhujú použiť syntax OWL DL, aby mohol RDFS-style reasoner spracovať všetky relevantné odvodenia a ontológia by bola použiteľná s OWL DL ontológiami a bázami poznatkov bez toho, aby bol výsledný model OWL Full. Takto navrhnuté riešenie má nasledovné výhody:

- Všetky dáta anotované použitím GoodRelations môžu byť správne interpretované s použitím RDFS-style reasonera
- Ontológia a dáta v kombinácii s OWL produktmi a ontológiami služieb sa nestanú OWL Full (môžu byť použité s DL alebo DLP reasonerom)
- Je umožnené definovať dôležité koncepty pre ontológie produktov a služieb, keďže ich treba importovať budúcimi verziami eClassOWL a iných ontológií

Bližšia špecifikácia GoodRelations ontológie je na [34], v sekcii publikácií. Pri používaní GoodRelations je potrebné mať unikátne identifikátory nie len pre webové zdroje (stránky), ale aj pre všetky konceptuálna prvky z používanej oblasti. Tu sa neodporúča brať URI existujúcich webových zdrojov, lebo tieto môžu zmiešavať viacero konceptuálnych entít. Prístupy k tvorbe identifikátorov pre konceptuálne entity sú opísané na [34], v sekcii publikácií. Projekt beží od roku 2005 a počas svojho vývoja sa prispôsoboval viacerým požiadavkám partnerov. Ontológia GoodRelations je dostupná pod licenciou Creative Commons Attribution 3.0.

3.3 Porovnanie vybraných projektov

Hoci každý z vyššie uvedených projektov je samostatný, dajú sa medzi nimi nájsť isté podobnosti a odlišnosti. Zamierame sa na tri projekty, ktoré práve boli obsiahlejšie opísané. Projekty DBpedia aj FOAF sa zameriavajú na prepájanie dát, ale každý z nich pracuje v tejto oblasti po svojom a používa iný princíp. Podobnosť sa dá objaviť aj medzi FOAF a GoodRelations, keďže oba projekty pracujú so strojovým spracovaním dát. Hlavným rozdielom medzi DBpediou a FOAF je, že DBpedia prepája rôzne dáta (rôzne datasety) do jednej veľkej siete, zatiaľ čo FOAF prepája dáta z jedného datasetu (zo svojho datasetu) medzi sebou. Výsledným efektom DBpedie má byť schopnosť používateľov pýtať sa komplikované dotazy a schopnosť počítača tieto dotazy správne vyhodnotiť a poskytnúť zmysluplné odpovede. Oproti tomu, FOAF má umožniť ľuďom vyjadriť vzťahy medzi sebou navzájom alebo medzi sebou a nejakými vecami a potom pomocou týchto vzťahov ďalej operovať (dobrým príkladom môže byť Web of Trust, kde je dôvera k neznámym dokumentom odvodená od dôvery našich priateľov a od dôvery k našim priateľom). Vyjadrovanie vzťahov by sme mohli prirovnať k pridávaniu si priateľov v súčasných sociálnych sieťach (v zmysle, že obsahu od priateľov v súčasných sociálnych sieťach dôverujeme a zároveň sprístupňujeme svoj obsah pre nich, pričom užívatelia, ktorí nie sú medzi našimi priateľmi nemajú prístup k nášmu obsahu a ani my k ich obsahu – paralela s dôverou, resp. nedôverou). Zároveň je rozdiel aj medzi strojovým spracovaním dát v projektoch FOAF a GoodRelations. Spracovanie dát v FOAF prebieha ako bolo opísané vyššie, čiže je to istá forma odvodzovania. GoodRelations sa zameriava na podporu internetového vyhľadávania a preto zmyslom tohto spracovania dát je nájdenie zhody medzi zadanou požiadavkou a ponúkanými produktmi. V tomto zmysle je zase GoodRelations mierne podobné DBpedii (oba projekty hľadajú odpovede na zadané požiadavky, pričom je dôležité si uvedomiť, že odpovede aj požiadavky, resp. otázky sa v oboch projektoch zásadne líšia). Všetky tri projekty pracujú so svojou vlastnou technológiou (hlavne s vlastnou

ontológiou), ale aj použitie tejto technológie je odlišné. DBpedia vytvára svoju vlastnú ontológiu (a vzhľadom na rozsiahlosť informácií, ktoré musí spracovať ešte nie je kompletná) a umiestňuje ju u seba a používateľom iba dáva možnosť opýtať sa na niečo z nej. FOAF aj GoodRelations umiestňujú svoje technológie (GoodRelations ontológiu, resp. FOAF slovník) priamo u používateľov. Výrazy FOAF si pridá do svojej stránky každý sám a tak vyjadrí svoje vzťahy k okoliu. Opisy pomocou GoodRelations si do svojich stránok pridajú všetci poskytovatelia výrobkov a služieb a tieto opisy potom zužitkujú webové vyhľadávače.

Praktické použitie projektu FOAF sa naskytuje priam samo – je to nie len vyjadrenie vzťahov v štýle sociálnych sietí, ale a čo by mohlo mať aj väčší prínos je už známe vyjadrenie vzťahov medzi firmami [47]. Tu by bolo jednoznačne prospešné vyjadriť kto sú partneri danej firmy, kto sú dodávatelia a odberatelia, aké sú konkurenčné firmy. To by pomohlo zorientovať sa na trhu. Zároveň by sa dalo pekne vyjadriť „toto je náš výrobok (služba)“. Pri praktickom použití DBpedie momentálne nie je veľký priestor na variácie vzhľadom na jej poslanie a cieľ. Jedna stránka s množstvom informácií, nad ktorými sa budeme vedieť dotazovať. Zaijímavé (najmä z obchodného hľadiska) je použitie technológií v GoodRelations. Myšlienka zostáva rovnaká ako v cieľoch projektu, ale význam získava najmä v podmienkach malých ekonomík ako je Slovensko. Tu by podrobný opis ponúk firiem zvýšil ich šancu na presadenie sa nie len na domácom trhu (po zohľadnení kvality slovenských výrobkov), ale aj v zahraničí a výrazne zvýšil ich konkurencieschopnosť pomocou zlepšeného vyhľadávania na webe.

Záver

Objem prác vykonávaných online neustále rastie a preto je zefektívňovanie pracovného prostredia, v našom prípade prostredia webu, veľmi dôležitou otázkou. Dlhodobá vízia sémantického webu poskytuje značné uľahčenie vykonávaných činností (s pomocou inteligentných agentov) a v dôsledku toho aj zníženie nákladov spojených s týmito činnosťami. Na presadenie a rozšírenie technológií sémantického webu bolo zrealizovaných veľa projektov (zväčša vedeckého charakteru) a mnohé projekty stále prebiehajú alebo vznikajú nové.

Cieľom tejto práce bolo poskytnúť stručný prehľad projektov z oblasti šírenia myšlienok sémantického webu a pokúsiť sa navrhnúť konkrétne uplatnenie niektorých z nich. Na projekty sme sa pozerali z hľadiska cieľov projektu (čo sa snažili alebo snažia presadiť) a z hľadiska použitých technológií. S ohľadom na skutočnosť, že myšlienka sémantického webu je známa už pomerne dlho (od roku 1999) sme sa v úvode práce zamerali na projekty, ktoré boli realizované a ukončené v predchádzajúcom období. V tejto časti uvádzame aj projekty, na ktorých sa podieľali aj zástupcovia so Slovenska.

V poslednej kapitole práce sa venujeme projektom, ktoré sú stále aktívne. Tu podávame výpočet niekoľkých projektov spolu s ich cieľmi a technológiami použitými na dosiahnutie uvedených cieľov. Vzhľadom na rozsiahlosť témy neposkytujeme úplný zoznam v súčasnosti prebiehajúcich projektov a zároveň vyberáme tri projekty, ktorým sa venujeme podrobnejšie v druhej časti poslednej kapitoly. Tu najprv uvádzame aké ciele projekt má a rozoberáme technológie použité v projekte do väčšej hĺbky. Na záver porovnávame vybrané tri projekty ako zástupcov väčších skupín projektov s podobnými technológiami. Po porovnaní nasleduje náš návrh praktického uplatnenia projektov, ktoré nachádzame najmä v ekonomickej oblasti (projekty FOAF a GoodRelations) so zvýraznením pozitívneho efektu, ktorý by ich uplatnenie malo v slovenskej ekonomike. Medzi hlavné konkurenčné výhody slovenskej ekonomiky patrí kvalita. Slovenská ekonomika je malá a preto sa naši výrobcovia ťažko uplatňujú na medzinárodnom trhu. Aplikovanie projektov FOAF a GoodRelations v oblasti podnikania by malo za následok presnejšie vyhľadávanie podľa konkrétnejších kritérií a v konečnom dôsledku zlepšenie konkurencieschopnosti najmä malých a stredných firiem, nie len na Slovensku, ale celosvetovo.

[1] O'REILLY, T. *What Is Web 2.0* [online]. [cit. 4. 2. 2011]. Dostupné na internete: <http://oreilly.com/web2/archive/what-is-web-20.html>
[2] Kolektív autorov. *Web 2.0* [online]. [cit. 4. 2. 2011]. Dostupné na internete: <http://www.paulgraham.com/web20.html>
[3] Kolektív autorov. *Web 2.0* [online]. [cit. 4. 2. 2011]. Dostupné na internete: <http://en.wikipedia.org/wiki/Web_2.0>
[4] Kolektív autorov. *Web 2.0: Nové príležitosti, nové riziká* [online]. [cit. 4. 2. 2011]. Dostupné na internete: <http://www.euractiv.sk/informacna-spolocnost/zoznam_liniek/web-20-nove-prilezitosti-nove-rizika>
[5] Kolektív autorov. *Pros and Cons to Web 2.0* [online]. [cit. 10. 2. 2011]. Dostupné na internete: <http://www.ajaxwith.com/Pros-and-Cons-to-Web-20.html>
[6] Exforsys Inc. *Advantages and Disadvantages of Web 2.0* [online]. 25. 4. 2007. 26. 4. 2007. [cit. 10. 2. 2011]. Dostupné na internete: <http://www.exforsys.com/tutorials/web-2.0/advantages-and-disadvantages-of-web-2.0.html>
[7] Kolektív autorov. *Understand Web 2.0 Security Issues - As Easy as 2, 1, 3* [online]. [cit. 10. 2. 2011]. Dostupné na internete: <http://1raindrop.typepad.com/1_raindrop/2007/03/understand_web_.html>
[8] Kolektív autorov. *Web 2.0- The disadvantages* [online]. [cit. 10. 2. 2011]. Dostupné na internete: <http://knol.google.com/k/web-2-0-the-disadvantages#>
[9] KIDMAN, A. *Digg revolt highlights shortcomings of Web 2.0 view* [online]. 2. 5. 2007. [cit. 10. 2. 2011]. Dostupné na internete: <http://www.itwire.com/your-it-news/home-it/11776-digg-revolt-highlights-shortcomings-of-web-20-view>
[10] Kolektív autorov. *Knowledge Web FP6-507482* [online]. [cit. 15. 2. 2011]. Dostupné na internete: <http://knowledgeweb.semanticweb.org/semanticportal/sewView/frames.html>
[11] Kolektív autorov. *Semantically-Enabled Knowledge Technologies* [online]. [cit. 15. 2. 2011]. Dostupné na internete: <http://www.sekt-project.com/>
[12] Kolektív autorov. *Data, Information, and Process Integration with Semantic Web Services - Project DIP* – [online]. [cit. 15. 2. 2011]. Dostupné na internete: <http://dip.semanticweb.org/index.html>
[13] Kolektív autorov. *aceMedia* [online]. [cit. 15. 2. 2011]. Dostupné na internete: <http://www.acemedia.org/aceMedia/index.html>
[14] Kolektív autorov. *About AIM@SHAPE* [online]. 14. 10. 2010. [cit. 15. 2. 2011]. Dostupné na internete: <http://www.aim-at-shape.net/>
[15] Kolektív autorov. *Knowledge and Content Management with semantically enriched objects* [online]. [cit. 15. 2. 2011]. Dostupné na internete: <http://metokis.salzburgresearch.at/>
[16] Kolektív autorov. *Welcome to the NeOn Project* [online]. 24. 2. 2010. [cit. 15. 2. 2011]. Dostupné na internete: <http://www.neon-project.org/nw/Welcome_to_the_NeOn_Project>
[17] Kolektív autorov. *Reasoning on the Web with Rules and Semantics* [online]. [cit. 15. 2. 2011]. Dostupné na internete: <http://rewerse.net/>
[18] Kolektív autorov. *Integrated Project SUPER* [online]. [cit. 15. 2. 2011]. Dostupné na internete: <http://www.ip-super.org/component/option,com_frontpage/Itemid,1/>
[19] Kolektív autorov. *SWAD-Europe* [online]. [cit. 15. 2. 2011]. Dostupné na internete: <http://www.w3.org/2001/sw/Europe/>

[20] BOURNEZ, C. *Web Services and Semantics (WS2) Project* [online]. 3. 4. 2007. [cit. 15. 2. 2011]. Dostupné na internete: <http://www.w3.org/2004/WS2/>
[21] Stanford-Smith, B. 2001. *E-work and E-commerce: Novel Solutions and Practices for a Global Networked Economy.* Lisse: IOS Press, 2001. 1400s. ISBN-13: 978-1586032050
[22] Kolektív autorov. *EÚ R&D projekty v oblasti eGovernmentu* [online]. [cit. 10. 3. 2011]. Dostupné na internete: <http://www.egov.sk/EUprojekty>
[23] Kolektív autorov. *Access-eGov Project* [online]. [cit. 10. 3. 2011]. Dostupné na internete: <http://www.accessegov.org/acegov/web/uk/index.jsp>
[24] Kolektív autorov. *Semantic-enabled Agile Knowledge-based E-Government* [online]. [cit. 10. 3. 2011]. Dostupné na internete: <http://www.sake-project.org/>
[25] Kolektív autorov. *SAKE – Agilný eGovernment* [online]. [cit. 10. 3. 2011]. Dostupné na internete: <http://www.sake-project.org/index.php?id=49>
[26] RISTVEJ, J. *Vedecké metódy* [online]. 31. 5. 2010. [cit. 11. 3. 2011]. Dostupné na internete: <http://www.trilobit.fai.utb.cz/vedecke-metody>
[27] HANÁČEK, J. *Vedecké a nevedecké metódy, druhy, charakteristika, príklady* [online]. [cit. 11. 3. 2011]. Dostupné na internete: <http://www.docstoc.com/docs/1663957/Vedecke_a_nevedecke_metody_-_VP_2007-08-JH>
[28] ČALKOVSKÁ, A. *Metóda vedy* [online]. [cit. 11. 3. 2011]. Dostupné na internete: <http://www.lefa.sk/internet/HandoutyJH/slov/03.pdf>
[29] SWARTZ, A. *The Semantic Web In Breadth* [online]. 1. 5. 2002. [cit. 23. 12. 2010]. Dostupné na internete: <http://logicerror.com/semanticWeb-long >
[30] PALMER, S. B. *The Semantic Web: An Introduction* [online]. 9. 2001. [cit. 27. 12. 2010]. Dostupné na internete: <http://infomesh.net/2001/swintro/>
[31] Kolektív autorov. *Firemná stránka* [online]. [cit. 18. 3. 2011]. Dostupné na internete: <http://www.w3.org/>
[32] Kolektív autorov. *DBpedia* [online]. 21. 4. 2011. [cit. 18. 3. 2011]. Dostupné na internete: <http://dbpedia.org/About>
[33] Kolektív autorov. *The Friend of a Friend (FOAF) project* [online]. [cit. 18. 3. 2011]. Dostupné na internete: <http://www.foaf-project.org/>
[34] HEPP, M. *GoodRelations: The Web ontology for E-Commerce* [online]. [cit. 18. 3. 2011]. Dostupné na internete: <http://www.heppnetz.de/projects/goodrelations/>
[35] Kolektív autorov. *SIOC – project* [online]. [cit. 18. 3. 2011]. Dostupné na internete: <http://sioc-project.org/>
[36] Kolektív autorov. *Semantic Interoperability of Metadata and Information in unLike Environments* [online]. 27. 2. 2007. [cit. 18. 3. 2011]. Dostupné na internete: <http://simile.mit.edu/>
[37] Kolektív autorov. *NextBio* [online]. [cit. 18. 3. 2011]. Dostupné na internete: <http://www.nextbio.com/b/nextbio.nb>
[38] Kolektív autorov. *Linking Open Data* [online]. 8. 4. 2011. [cit. 18. 3. 2011]. Dostupné na internete: <http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>
[39] Kolektív autorov. *OpenPSI* [online]. [cit. 18. 3. 2011]. Dostupné na internete: <http://www.openpsi.org/>
[40] Kolektív autorov. *METEOR-S: Semantic Web Services and Processes* [online]. [cit. 18. 3. 2011]. Dostupné na internete: <http://lsdis.cs.uga.edu/projects/meteor-s/>

[41] Kolektív autorov. *SKOS Simple Knowledge Organization System* [online]. 23. 5. 2010. [cit. 18. 3. 2011]. Dostupné na internete: <http://www.w3.org/2004/02/skos/>
[42], [43], [44], [45] Kolektív autorov. *Semantic Web Projects* [online]. [cit. 18. 3. 2011]. Dostupné na internete: <http://domino.research.ibm.com/comm/research_projects.nsf/pages/semanticweb.Semantic%20Web%20Projects.html>
[46] Kolektív autorov. *Semantic Web Portal Project* [online]. [cit. 18. 3. 2011]. Dostupné na internete: <http://sw-portal.deri.at/>
[47] Kolektív autorov. *Sociálna sieť firiem v Slovenskej republike* [online]. [cit. 18. 3. 2011]. Dostupné na internete: <http://foaf.sk/>
[48] Kolektív autorov. *Redakčný systém (CMS)* [online]. [cit. 18. 3. 2011]. Dostupné na internete: <http://www.creativesites.sk/redakcny-system-cms-joomla/>
[49] Kolektív autorov. *Dspace* [online]. [cit. 18. 3. 2011]. Dostupné na internete: <http://www.dspace.org/>
[50] NATIONS, D. *What is a Mashup?* [online]. [cit. 18. 3. 2011]. Dostupné na internete: <http://webtrends.about.com/od/webmashups/a/what-is-mashup.htm>