

Univerzita Karlova v Praze

Filozofická fakulta

Ústav informačních studií a knihovnictví

Studijní obor: Studia nových médií

Bc. Eduard Piňos

Sémantický web – ontologický Internet budoucnosti

Diplomová práce

Praha 2009

Vedoucí diplomové práce: Mgr. Josef Šlerka

Oponent diplomové práce:

Datum obhajoby:

Hodnocení:

UNIVERZITA KARLOVA V PRAZE

Filozofická fakulta

Katedra / ústav: Ústav informačních studií a knihovnictví

Školní rok: 2008/2009

ZADÁNÍ DIPLOMOVÉ PRÁCE

Jméno a příjmení studenta: Bc. Eduard Piňos

Datum narození: 27.10.1982

Obor studia / kombinace: Studia nových médií

Název práce v češtině: Sémantický web – ontologický internet budoucnosti

Název práce v angličtině: Semantic web – The ontological future Internet

Zásady pro vypracování:

Cílem práce je teoretické pojednání a výzkum rozvíjejícího se konceptu sémantického webu a nových principů architektury internetu budoucnosti.

Předpokládaná osnova práce je rozčleněna do tří základních bodů:

- 1) Deskripce a rozbor pojmu sémantický web; analýza současného stavu vývoje sémantického webu dle monitoringu World Wide Web Consortium a výzkumů na předních světových univerzitách (založeno na pracích Albert-László Barabásiho, Félix Guattariho, Tima Bernerse-Leeho, Grigorise Antoniaua a Franka van Harmelena).
- 2) Cíle hnutí kolem sémantického webu. Technologie obousměrného linkování dat a nástup webu třetí generace. Praktické využití nových forem indexování, publikování a vyhledávání v internetových databázích. OWL (Web Ontology Language), deskripční logika jako nadstavba formální logiky pro sémantické sítě a objektově orientované programovací jazyky (založeno na pracích F. Baadera, D. Calvanese, D. L. McGuinness, D. Nardi, P. F. Patel-Schneidera, Iana Horrockse, John F. Sowa a Sinuhe Arroyo).

- 3) Dopady sémantického webu na nové sociální uskupení internetových komunit, na proměny mezi-lidské komunikace a přístupu k informačním zdrojům a mediální praxi. Umělá inteligence a kolektivní vědomí počítačů nové generace a jejich uživatelů v rámci světové sítě. (založeno na pracích M.McLuhana, Toma Grubera, J. Currana, W. Boddyho, M.Castellse, W. Gibsona, Raye Kurzweila, Vernora Vinge, Gillesse Deleuze).

Diplomová práce bude připravena v souladu s platnými vnitřními předpisy FF UK a dalšími metodickými pokyny a normativními dokumenty.

Seznam odborné literatury:

1. NEWMAN Mark, BARABASI Albert-Laszlo, WATTS Duncan J.: *The Structure and Dynamics of Networks*. [S.l.]: Princeton University Press, 2006. 624 s. ISBN-13: 978-0691113579
2. CASTELLS, Manuel: *The Internet Galaxy: Reflections on the Internet, Business, and Society*. [S.l.]: Oxford University Press, 2003. 304 s. ISBN-10: 0199255776.
3. DELEUZE, Gilles, GUATTARI, Felix,: *Thousand Plateaus*. [S.1.]: Continuum International Publishing Group, 2004. 704 s. ISBN-13: 978-0826476944.
4. BERNERS-LEE, Tim, FENSEL Dieter (ed), HENDLER James A. (ed), LIEBERMAN Henry (ed), WAHLSTER Wolfgang (ed): *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential* . [S.l.]: MIT Press, 2004. 503 s. ISBN-13: 978-0262562126

Vedoucí práce: Josef Šlerka

Datum zadání práce: 25.5. 2008

L.S.

.....
Vedoucí součásti – ředitel ÚISK FF UK

.....
Děkan FF UK

V Praze dne 25.5.2008

Prohlášení:

Prohlašuji, že jsem diplomovou práci zpracoval samostatně, a že jsem uvedl všechny použité informační zdroje.

V Praze, 20. srpna 2009

.....

.....

podpis diplomanta

PIÑOS, Eduard. *Sémantický web – ontologický Internet budoucnosti [Semantic web – The ontological future Internet]*. Praha, 2009. 72 s. Diplomová práce. Univerzita Karlova v Praze, Filozofická fakulta, Ústav informačních studií a knihovnictví 2009. Vedoucí diplomové práce: Mgr. Josef Šlerka

Abstrakt

Tématem práce je popis historie, stavu a budoucího vývoje sémantického webu a jeho technologií v rámci předpokladů World Wide Web Consortium – W3C. Jde především o základní seznámení s principy fungování tohoto konceptu ve spojitosti s nenaplněnými možnostmi World Wide Web v oblasti poskytování informací a vědomostí pomocí automatizovaného strojového zpracování dat. První část práce si klade za cíl poskytnout podrobnější informace o logickém rámci a technologických aspektech, na nichž je sémantický web postaven. Jedná se zejména o RDF – Resource Description Framework, jako základní datový model, z něhož vychází většina sémantických webových jazyků a aplikací a koncept URI – Uniform Resource Identifier, používaný pro identifikaci webových i newebových zdrojů.

Druhá část práce se věnuje především ekonomickým a dalším problémům, které brání sémantickému webu v masovém rozvoji. Kromě toho jsou zde podrobně popsány existující sémantické aplikace a jejich možnosti využití v čele s projektem znalostní báze několika německých univerzit - DBpedia.

V poslední části práce jsou zmíněny filosofické konotace spojené s proměnou uživatelské praxe v rámci sémantického webu a dopad těchto skutečností na webovou komunitu a globální societu jako takovou.

Klíčová slova

sémantický web, ontologie, RDF, OWL, XML, WEB 3.0, znalostní báze, sémantické odkazy, inteligentní agenti

Abstract

The theme of this thesis is a description of the history, current state and the future development of the semantic web and its technologies within the assumptions of World Wide Web Consortium – W3C. It mainly covers the principles of this concept in correlation with an unfulfilled potential of World Wide Web in the area of information retrieval using machine-readable data. The first part of the text provides deeper information about logical basis and technological aspects of the semantic web. Above all it deals with the basic data model for most of the semantic languages and applications – the RDF (Resource Description Framework) and the URI (Uniform Resource Identifier) used for the identification of web based or even non-web based resources.

The second part of the thesis deals with some basic facts of the problems of the semantic web to gain mass support for its development. Among this there is a broader description of the most important contemporary semantic applications, including the leading one – the DBpedia knowledge base created by several German universities.

In the last part the philosophical connotations linked to the transformation of user experience within the semantic web are mentioned, so as is the overall impact on the web community and the global society.

Keywords

Semantic web, an ontology, RDF, OWL, XML, WEB 3.0, knowledge base, semantic links, intelligent agents

OBSAH

OBSAH.....	8
1. ÚVOD.....	10
2. TECHNOLOGIE SÉMANTICKÉHO WEBU.....	14
2.1. Základní informační zdroj – URI	15
2.2. RDF a XML	18
2.2.1. XML – první krůčky k sémantickým technologiím.....	19
2.2.2. Jak RDF funguje?.....	21
2.3. Co je to v počítačové vědě ontologie a proč je důležitá pro sémantické webové jazyky	23
2.3.1. Z čeho se ontologie skládá?.....	24
2.4. OWL a DAML – ontologické programování v praxi.....	24
2.4.1. DAML + OIL.....	24
2.4.2. OWL	27
2.5. Principy práce se sémantickými technologiemi – doplnění	28
2.5.1. SPARQL – dotazovací jazyk pro datový model RDF.....	28
2.5.2. Deskripční logika – vzor pro sémantické dokazování.....	28
2.5.3. DRY - pravidlo „Don't repeat yourself“	28
2.5.4. Mikroformáty – sémantické prvky v HTML.....	29
3. SOUČASNÝ STAV SÉMANTICKÉHO WEBU V PRAXI A JEHO EKONOMICKÉ ASPEKTY.....	30
3.1. Proměna webových služeb s příchodem sémantických technologií	33
3.2. Co má společného WEB 2.0 a sémantický web a lze je propojit?	35
3.3. Co je to WEB 3.0 a co bude po něm?.....	38
3.4. Sémantika, BPM a B2B.....	40
3.5. Specializované sémantické aplikace pro běžné uživatele v praxi. .	43
3.5.1. Twine, Mywikibiz a DBpedia.....	43
3.5.2. Sociální polo-sémantické aplikace	50
4. Filosofické aspekty sémantického webu.....	55
4.1. Proměna síťového paradigmatu díky sémantickému webu.....	55
4.1.1 Ubiquitous WEB.....	57
4.1.2. Simulace a realita.....	58

<i>4.1.3. Filozofové proti sémantickému webu</i>	<i>58</i>
<u>5. Závěr.....</u>	<u>60</u>
<u>Seznam literatury.....</u>	<u>62</u>
Citovaná literatura.....	62
Použitá literatura.....	69
<u>Evidence výpůjček.....</u>	<u>71</u>
<u>Prohlášení:.....</u>	<u>71</u>

1. ÚVOD

Masové rozšíření síťové technologie World Wide Web můžeme datovat na konec 20. století s příchodem nových a rychlejších technologií pro přenos dat. Dnes již je World Wide Web (dále jen WWW nebo WEB) tak běžnou součástí našeho života, jako třeba denní tisk, televize a knihy a nikoho ani nenapadne se nad tím pozastavit. Zároveň si málokdo uvědomuje, že myšlenky, které vedly ke vzniku světové informační sítě, směřují ještě mnohem dále k lidem, kteří předběhli svými představami o několik desítek let svou dobu, ale většinou bohužel i její technické možnosti. Něco podobného by se s trochou nadsázky dalo říct i o sémantickém webu a jeho průkopnících.

Když v roce 1945 přišel americký vynálezce Vanevar Bush s nápadem na stroj, který by dokázal vytvořit odkazy mezi různými typy dokumentů a zpětně se pomocí nich orientovat, asi nikdo netušil, že právě položil základ síťové technologii, na níž je současný WEB postaven - hypertextu. Jeho hypotetický stroj nazval Memex a očekával, že vytvořené linky a vůbec veškeré dokumenty by se daly ukládat na mikrofilm (55). Stroj nebyl pochopitelně nikdy uveden do provozu, ale samotná myšlenka, že mezi dokumenty (potažmo daty) jsou přirozené a logické vazby a uživatelé by měli možnost je sledovat, stačila k tomu, aby se věci daly do pohybu.

Zhruba o 15 let později Američan Ted Nelson jako první použil pojem *hypertext* ve svém článku „Souborová struktura pro složité, měnící se a nedeterminované“ (21). Později následovaly vynálezy prvních systému na editaci *hyperlinků* a pokusy o vytvoření primitivních počítačových sítí. Teprve když se však do vývoje vložila americká armáda, respektive její výzkumné centrum DARPA, začalo vše nabírat reálný rozměr a první počítačová síť – ARPANET, propojující vládní instituce a univerzity byla na světě. Později byl pro ARPANET vynalezen protokol TCP/IP (Internet Protocol Suite), který překonal jeden z hlavních problémů, a sice nekompatibilitu mezi jednotlivými protokoly určenými pro přenos paketových dat. To vše se událo na začátku 80. let 20. století, doba Internetu však měla teprve přijít. Po několikaletém výzkumu, který probíhal na hranici devadesátých let 20. století, představili vědci pod vedením Tima Bernerse-Leeho a Bena Segala z mezinárodního

institutu CERN¹ hlavní součástí Internetu, jak ho známe dnes - WWW (41). Byl ustanoven značkovací jazyk HTML (Hypertext Markup Language) pro zobrazování dat na WWW, stejně jako protokol pro jeho přenos v rámci sítě – HTTP (který se postupně stal unifikovaným protokolem pro celý Internet). Od té doby už se WEB jen nadále rozrůstá a jsou k němu přiřazovány nové a nové technologie, které různým způsobem rozšiřují jeho funkčnost a možnosti využití pro uživatele. Ať už se však jedná o javascript, flash, CSS, XHTML a nepřehledné množství dalších technologií (s výjimkou XML či například systému JSON) – veškerý vývoj v této oblasti nikdy nic neměnil na základní funkci webu – tedy pouhém zobrazování informací a *linků* (zadaných lidskými uživateli) mezi nimi.

Společně s dalšími Internetovými službami (jako e-mail, sdílení dat, internetová telefonie a další) proměnil WEB lidskou schopnost čerpání a sdílení informací jednou pro vždy. I tak ale někteří vizionáři, jmenovitě Tim Berners-Lee, už ve chvíli kdy WEB vytvářeli, chápali, že jeho skutečný potenciál bude obrovským způsobem nenaplněn, pokud se nezmění jeho základní aspekt, potažmo funkcionalita. Výjimečný vědec si uvědomil, že HTML a všechny jeho přidružené technologie jsou jazyky stvořené lidmi pro počítače, ale ve své podstatě jsou pochopitelné pouze lidmi. Počítač nerozumí přirozenému lidskému jazyku, ale nerozumí ani HTML – jedná se pro něj pouze o soubor příkazů určujících, co a jak může zobrazit. Nechápe však význam zobrazovaných informací a nedokáže určit jejich relevanci. Právě tento zásadní nedostatek vedl Bernerse-Leeho k definici pojmu sémantický web, neboli také *web of knowledge*, mimo jiné popsáném v jeho slavném článku pro Scientific American z roku 2001 (62). V té době měli ovšem lidé kolem sémantického webu už několik let velmi ucelenou představu ohledně toho, jak by skutečná znalostní síť měla fungovat a několik skupin počítačových vědců již pracovalo na vývoji datového modelu, který by umožňoval počítačům rozpoznávat a abstrahovat smysl zobrazovaných informací a dat

Výsledkem je datový model RDF – *Resource Description Framework* postavený na systému trojic a logických grafů, umožňující počítači definovat z dané množiny smysl vloženého slova (nebo jiné formy dat). To samé platí i pro tzv. *sémantické linky*. V běžném HTML je totiž odkaz prostě jen směřováním uživatele na jinou URL, pomocí RDF však může uživatel (respektive už jeho webový prohlížeč – či inteligentní softwarový agent) rozpoznat, zda je pro něj tento odkaz vůbec relevantní. RDF samo o sobě však není webovým jazykem, jak se laická veřejnost občas domnívá, nýbrž pouze systémem (modelem) na němž

¹ Více informací o výzkumném středisku CERN lze nalézt na webové adrese:
<http://public.web.cern.ch/public/>

může být WEB postaven. Pro vlastní syntaxi používá RDF specifikací jako XML a dalších, o nichž bude řeč v pozdějších kapitolách.

Dalšími velice důležitými pojmy pro pochopení sémantického webu jsou URI a *ontologie*. URI - *Universal Resource Identifier* dává každému zdroji (a to i ne-webovému) jednu konkrétní unikátní adresu. Tento fakt je pro sémantický web velice důležitý, jelikož umožňuje počítačovým systémům vyhnout se potenciální dvojznačnosti, kterou nemají v binárním systému jakoukoliv možnost rozpoznat. *Ontologie* jsou pak jakýmsi instancemi určujícími klasifikaci, hierarchii a taxonomii souborů dat, pomocí nichž se systém orientuje například v rozpoznávání smyslu slov v konkrétním oboru. Ontologie tak vznikají většinou kolem nějakého tématu, vědního oboru (například farmacie) a podobně.

Jak je patrné z velice zjednodušeného úvodu do technologií sémantického webu, nejedná se o žádnou sci-fi představu vzdálené budoucnosti, či o pátrání po umělé inteligenci podobné člověku, jak by si možná někteří lidé představovali. Naopak, jde o soubor sofistikovaných vývojářských nástrojů a principů k dosažení relativně prostého cíle – *strojové zpracovatelnosti dat*. Často využívané pojmy, jako „smysl“ a „znalost“ jsou v tomto případě přeci jen trochu nadsazené, jelikož počítačové systémy (tedy tak, jak je známe dnes – pokusy například s neurálními počítači nejsou brány v potaz) v pravém slova smyslu samozřejmě nebudou chápat výraz typu „strom“ a nedokážou si abstrahovat jeho fyzickou podobu. Co však v případě sémantického webu poměrně snadno zvládnou, je určit, že na stromě se vyskytuje listí a ke svému životu potřebuje vodu, nebo že například takový smrk je také stromem (jinými slovy počítač určí pomocí logického systému RDF a příslušné ontologie smrk, jako podkategorii stromu a z toho usoudí: smrk je strom). Taková schopnost ovšem radikálně proměňuje možnosti WWW pro uživatele a mnohonásobuje svou informační sílu a funkcionalitu jako takovou. Nasnadě jsou pak donedávna tak nepředstavitelné vlastnosti jako například vyhledávání v přirozeném jazyce.

Jak již bylo řečeno (čistě principiálně), z hlediska podstaty nové technologie současného WWW uživatelům nic opravdu nového nepřináší – stále jde pouze o zobrazování „nějakých“ informací, potažmo o jejich přenos. Veškeré vyhledávání je založeno na statistických algoritmech a shodnosti klíčových slov (či znaků). Trochu jiné světlo do vývoje Internetu nebo spíše WWW přinesl boom v podobě WEB 2.0. Z hlediska technologií a jejich významu pro počítačové systémy se ani zde nic tak významného nezměnilo, ale z pohledu uživatele jde o revoluční proměnu v otázce distribuce informací.

Uživatelé vytvářejí sociální distribuční sítě a systém folksonomií, které nahrazují rigidní taxonomie za označení, jež mohou reální uživatele skutečně potřebovat. Právě zde je patrný jeden ze synergických momentů WEB 2.0 a sémantického webu – je to vytváření *metadat*, která umožňují ve svém důsledku mnohem efektivnější uživatelskou zkušenost a práci s WWW než doposud. Rozdíl je však v tom, že sémantický web má za cíl tento proces plně automatizovat a ne ho využívat jen jako doplněk komunikace mezi uživateli, tak jak je tomu v případě WEB 2.0.

Cílem této práce je podat čtenáři ucelený obraz o tom, co to sémantický web je, jaký je jeho účel, co má společného s představou webu třetí generace – WEB 3.0, v jaké fázi se nachází jeho vývoj v současnosti, co mu brání v masovém rozšíření a jaké (pokud vůbec nějaké) jsou dopady tohoto konceptu na lidskou společnost. Za tímto účelem je text rozdělen do tří tematických celků. První část je věnována především technologickému pozadí sémantického webu, jehož znalost je zásadní pro pochopení potenciálu ale i omezení, která sebou (v oblasti práce s daty a jejich smyslem) nese. Ve druhém celku je možné nalézt informace zejména o tom, jaký je reálný stav sémantického webu v současné době. Jsou zde představeny nejen skutečné a fungující on-line i offline *sémantické aplikace* a možnosti jejich využití, ale i technologická a ekonomická omezení, která vývoj sémantického webu už několik let poměrně významně brzdí. V závěrečném oddíle jsou stručně zmíněny filosofické aspekty celé problematiky v návaznosti na teorie některých významných světových myslitelů v oblasti médií a filosofie informatiky.

V neposlední řadě je třeba upozornit na pojmosloví využitě v rámci této práce. Jde především o technické pojmy spojené se sémantickým webem. České termíny (vzhledem k existenci minimálního počtu odborných prací na toto téma v českém jazyce), ještě nejsou zcela ustáleny. Drtivá většina autorů, která o toto téma byť jen okrajově zavadila, se rekrutuje z vyučujících a studentů kateder informatiky na dvou až třech českých univerzitách a translace zavedené anglické terminologie si přizpůsobují každý svému projevu. Autor této práce si proto vyhrazuje právo na užívání vlastních překladů této terminologie, založených především na oficiálních definicích základních pojmů sémantického webu ze strany World Wide Web Consortium – W3C (69).

2. TECHNOLOGIE SÉMANTICKÉHO WEBU

První věcí, kterou je si třeba při zkoumání technologií sémantického webu uvědomit je, že tyto technologie vycházejí a navazují na principy WWW, tak jak jej známe dnes. Když v roce 1963 definoval Ted Nelson slovo *hypertext* (21), byly položeny nejen základy WWW ale zároveň i sémantického webu, který (alespoň ve své současné podobě) je „pouhou“ *možností*, vývojovým stupněm či extenzí webu současného. V praxi to znamená, že sémantický web nepotřebuje ke svému plnohodnotnému fungování například jiná hardwarová zařízení nebo operační systémy osobních počítačů než využívá standardní WWW. Stejně tak základní technologie WWW musejí zůstat pro chod sémantického webu zachovány. Sémantický web také funguje na bázi klient-server komunikace. Stejně tak může (lépe řečeno musí) využívat HTTP protokol při komunikaci mezi klientem (webovým prohlížečem) a serverem. Stejná situace je i u vlastního přenosu dat, fungujícího na základě komunikačního protokolu TCP/IP (Internet Protocol Suite). Hlavní změnou je tedy vlastní distribuce a především získávání dat, které je uzpůsobeno k proměně informačního zdroje (který může dešifrovat pouze člověk) na zdroj znalostní (který může zpracovat i stroj).

Jestliže je základním stavebním kamenem a standardem WWW již od jeho počátku (tedy někdy od začátku 90. let 20. století) jazyk HTML (Hypertext Markup Language), pro sémantický web je to (v současné době) RDF (Resource Description Framework) – metadatový model pro zdroje (objekty) a vztahy mezi nimi (66, s.12). Samotné HTML (respektive dnešní XHTML) nelze chápat v porovnání s RDF jako kódovací jazyky stojící proti sobě. RDF již z principu není určeno k reprezentaci zobrazovaných elementů a nelze ho žádným způsobem stylovat po vizuální stránce. Jeho účelem není na rozdíl od HTML stavba webu, ale popis zdrojů a forem linkování mezi nimi. V současnosti hojně vyžívaný jazyk XML (eXtensible Markup Language) – jehož elementární prvky jsou implementovány už i v samotném XHTML je jakýmsi předvojem sémantických jazyků a schémat. Také účelem XML je pouze definice pravidel pro nakládání s datovými zdroji a vlastní obsah jeho elementů a značek (stejně jako u RDF) není definován. Technologie sémantického webu současnosti tedy mohou (nebo spíše musí) existovat paralelně s technologiemi rozšířenými v „běžném“ WWW.

Sémantický web nemění způsob zobrazování a fungování WWW jako takového, umožňuje pouze jiný model práce s informacemi a po technické stránce je zcela závislý na HTML a stejně tak nemá žádný vliv na komunikaci mezi klientem a serverem pomocí PHP a na další aktuální standardy ustanovené W3C. Prvotní jednotkou sémantického webu je *ontologie* (v angličtině „an ontology“ – na rozdíl od filosofického pojmu „ontology“), která formálně popisuje sféru diskurzu – skládající se z konečného počtu výrazů (66, s. 10).

RDF popisuje situaci – určení *ontologie*. RDF samotné není pro strojové zpracování ideální – jde pouze o datový model a ne o webový jazyk. Za účelem vlastního kódování byl vyvinut jazyk RDF Schema (dále již jen RDFS), které využívá podobnou skladbu a formu pro tvorbu *ontologií* jako má například XML (66, s. 52). Dalším vývojovým stupněm na poli sémantických webových jazyků je OWL (Ontology Web Language), které vzniklo z důvodu omezených možností RDFS. Modelový základ RDF je při svém popisu vztahů redukováno de facto na standardní formu predikátu s binárním výsledkem ANO/NE - RDFS je pak omezeno rozsahem a hierarchií tříd, podtříd a vlastností (66, s. 111). OWL oproti tomu nabízí bohatší možnosti využití sémantických principů. Fungování, principy a stavbu RDF, RDFS a OWL si detailněji přiblížíme v následujících kapitolách, stejně jako logický rámeček, na jehož základě vznikly a příklady dalších technologií sémantického webu.

2.1. Základní informační zdroj – URI

Zcela klíčovou součástí sémantického webu jsou zdroje, bez nichž by neexistovala data, jež by mohla být reprezentována. Veškeré zdroje na WWW (i mimo něj) se dají univerzálně označit jako URI (Universal Resource Identifier). URI se používá už od počátku WWW a jeho hlavním předpokladem je myšlenka, že každý zdroj je unikátně identifikovatelný. Typicky se na WWW URI skládá z lokace zdroje (tedy URL – Uniform Resource Location) a z názvu zdroje (URN – Uniform Resource Name). I když už víme, kde se zdroj nachází a jak se jmenuje, je třeba se k němu nějak dostat. To lze učinit s pomocí nějakého protokolu pro přenos dat, jako je například HTTP – Hypertext Transfer Protocol.

Samotná myšlenka URI, odpoutáme-li jí od konkrétní aplikace ve stávající podobě WWW, je však mnohem zajímavější. Standard URI umožňuje použití odlišných identifikátorů ve stejném kontextu, i přesto že mechanismy přístupu ke zdrojům, jež identifikují, se mohou lišit. Takto URI vytváří prostor pro uniformní sémantickou interpretaci

společných syntaktických konvencí napříč různými typy identifikátorů zdrojů. URI tím tak umožňuje vznik nových forem identifikátorů a jejich využití v mnoha různých kontextech a nových aplikacích a protokolech, aniž by to nějak negativně ovlivnilo ty současné zaběhnuté (61).

URI je standardem pro určitou syntax využívaným k identifikaci abstraktního či fyzického zdroje na WWW (56, s. 59). Skládá se vždy ze *schématu* odděleného dvojtečkou

a z *hierarchické části*, součástí může být i *dotaz* a *fragment*. Schéma musí vždy začínat písmenem, následováno může být již libovolným počtem písmen, čísel a některých dalších znaků. Hierarchická část bývá od schématu zpravidla oddělena ještě dvojitým symbolem /. Dotaz musí být od hierarchické části oddělen otazníkem a fragment symbolem křížku #. Fragment a dotaz však nejsou povinnou součástí URI. Konkrétní URI pro unikátní soubor na FTP serveru může tedy vypadat například takto:

`ftp://ftp.skoteinos.smiley.cz/eda/text.doc`

nebo takto:

`http://s3.amazonaws.com/basec/1239071/2812563/navrh01a.jpg`

ale i takto:

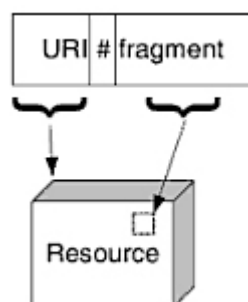
`mailto:eduard.pinos@gmail.com`.

Pro myšlenku sémantického webu je důležitý především fakt, že pomocí URI se lze za pomoci sémanticky (strojově zpracovatelné syntaxe) dostat vždy přesně k jednomu jedinému unikátnímu zdroji.

Specifičtější formou URI je URIref (URI reference), což je vlastně URI doplněná o fragment, způsobem, který byl popsán výše. Důležité je, že pro fragmenty jsou vždy použity znaky písmen a to vždy jen ve standardu *unicode*. URIref je důležité, jelikož díky němu RDF identifikuje zdroje, k čemuž by mu normální URI nestačilo (38, s. 866). Je to z toho důvodu, že způsob fungování RDF staví na stavbě sémantických trojic - *subjekt*, *predikát*, *objekt* (podrobněji v kapitole věnované přímo RDF) a pouhé URI bez fragmentového identifikátoru by RDF nemohlo do těchto kategorií, potřebných k získání sémantické *znalosti*, rozřadit. URIref je tedy vlastně absolutní reference, kterou každá URI může, ale nemusí mít. URIref může vypadat například takto:

www.skoteinos.smiley.cz/dokumenty-pracovni#ucetnictvi.

V rámci URL takový fragment většinou identifikuje specifické místo, kde se v rámci dokumentu identifikovaném danou URL nachází (32, s. 55). Pro lepší představu, proč je URIref pro přesnou identifikaci fragmentu zdroje důležitá, poslouží následující jednoduché schéma (32. S. 55):



Často se stává, že se URI či URIref zaměňuje s konceptem *namespaces* (dalo by se volně přeložit jako *jmenné prostory*), typického pro XML. Namespaces se zde využívají k vytváření unikátních jmen elementů a atributů značkovacího jazyka. To umožňuje propojení různých typů XML značkovacích jazyků (nebo lépe řečeno slovníků) a jejich společné fungování v rámci jedné aplikace. Každý XML slovník má svůj namespace, který se definuje pomocí URI (zpravidla URL) (29). Díky namespaces totiž nedochází k dvojznačnosti identických jmen. V praxi to vypadá tak, že se každý XML *tag* skládá ze dvou částí - *prefixu* a *lokální části*: `<xsd:local>`. Na začátek dokumentu je tedy třeba definovat namespace deklarací toho na co prefix odkazuje:

```
<xsd:schema xmlns:xsd=http://www.w3.org/2001/XMLSchema>
```

Značka `xsd` pak znamená, že se jedná o tag se specifikací XMLSchema (47, s. 42-44). Namespace se často zaměňuje s URI, protože jsou v podstatě totožné, co se týče syntaxe a lexikálních komponentů. V rámci specifikace URI se však počítá i s jeho předpokládaným využitím (61). To samé ovšem neplatí pro namespace – to se pouze „tváří“ jako URI, ale ve skutečnosti má pouze rozlišovací účel – ne identifikační.

2.2. RDF a XML

Sémantický web je postaven na principu *metadat*. Jak již bylo řečeno, jeho hlavním účelem je učinit identifikovatelné datové zdroje přístupné strojovému zpracování, tedy změnit informaci na znalost. Stroje (alespoň v současném stavu technologického vývoje) nejsou na rozdíl od člověka schopny jakékoliv abstrakce nad rámec svého programu. Běžné počítače pracují na základě binárních dat (pomineme-li pokusy s kvantovými a neutrálními počítači) a informaci, kterou získají, nejsou schopny interpretovat, ale pouze předat. Stejně tak fungují v podstatě všechny webové aplikace dneška včetně webových prohlížečů. Jediným způsobem, jak docílit toho, aby stroj „pochopil“ danou informaci a získal tak znalost, s kterou může dále (automatizovaně) pracovat je zavedení metadat. Proto byla v sekci W3C, která se zabývá vývojem a standardy sémantického webu, ustanovena zvláštní pracovní skupina za účelem vytvoření webového metajazyka vhodného pro tyto účely. A právě zde vzniklo RDF – jako základní datový standard pro procesování metadat.

Cílem RDF je vytvořit univerzální mechanismus pro veškeré aplikační oblasti, aniž by vytvářelo a priori jakýkoliv soud o této oblasti (31). RDF samo o sobě tedy ve skutečnosti není (jak už bylo zmíněno) kódovacím jazykem, nýbrž je to čistě datový model definující vztahy mezi datovými vstupy. K tomu, aby bylo RDF nějakým způsobem reprezentováno a přenášeno mezi aplikacemi, potřebuje nějakou konkrétní určitou syntax. V současnosti je v RDF pro použití na webových stránkách nejčastěji využíváno XML syntaxe, i když je možné ho reprezentovat i pomocí jiných syntaxí. Kromě samotného RDF existuje také (jak už bylo naznačeno výše) kódovací jazyk – slovník RDFS, které definuje zdroje pomocí podobné terminologie jako XML – určuje jejich vlastnosti pomocí metatagů. Tyto vlastnosti pak mohou být přiřazeny k určitým *třídám* za vzniku konkrétní taxonomie. RDFS staví svůj sémantický princip na tom, že vlastnosti jsou definovány odděleně od tříd. Kdokoliv programuje aplikaci na základě RDF datového modelu s využitím pravidel RDFS, může definovat, že libovolná vlastnost se dá asociovat s určitou třídou nebo více třídami (46, s. 52). Než bude popsáno, jak vlastní RDF vlastně vypadá, je dobré nastínit stavbu a vlastnosti XML, jehož syntax může RDF využívat.

2.2.1. XML – první krůčky k sémantickým technologiím

XML je anglická zkratka pro *Extensible Markup Language* – tedy ve volném překladu „rozšiřitelný značkovací jazyk“. Nejdůležitější částí zkratky XML je právě první písmeno X, naznačující, že se nejedná o jazyk uzavřený. Základem XML je totiž především poměrně jednoduchá o to však přesnější syntax, nikoliv definovaný soubor povolených výrazů. Výhodou XML (a důvodem proč se hojně rozšířilo především v on-line databázích či specifických webových aplikacích) je fakt, že na rozdíl od HTML umožňuje implementovat na WEB data, aniž by tato byla jakkoliv propojena se stylováním dokumentu (ze sémantického hlediska jsou stylovací tagy HTML zcela irelevantní). U XML si vlastní tagy (neboli značky) může každý vývojář vydefinovat v podstatě sám. Děje se tak většinou s ohledem na oblast, v níž je ten který aplikační jazyk (resp. XML slovník) využíván. Proto vznikly specifické jazyky jako například: SVG (Scalable Vector Graphics) pro datový zápis vektorové grafiky; MathML (Mathematical Markup Language) pro zápis matematických vzorců; CML (Chemical Markup Language) využívaný pro chemické prvky a další (14, s. 13). Výhodou XML pak je, že jeho otevřenost v podstatě umožňuje mezi jednotlivými jazyky - slovníky volně přecházet, případně si vytvářet zcela vlastní kombinace dle potřeby. Vždy záleží jen na tom, zda autor daného kódu chce, aby jeho data mohla číst jen jedna konkrétní aplikace, nebo raději využije nějakého obecnějšího XML slovníku s širším polem použití.

Správné formátování XML je v jednotlivých aplikacích kontrolováno pomocí tzv. *syntaktické analýzy* (neboli *parseru* – ten odděluje čistá data od meta-značek a znakových dat, u nichž pak analyzuje správnost syntaxe). Typicky bude takový parser obsahovat například jakýkoliv webový prohlížeč, který umožňuje čtení XML dokumentů. Všeobecně ho musí obsahovat každá aplikace schopná číst XML (14. s. 5). Aby mohlo být XML chápáno jako skutečný standard, vydává W3C aktualizace k pravidlům syntaktické analýzy a k lexikální gramatice – tedy pravidlům pro skladbu sekvencí znaků a jejich dělení. Na základě těchto pravidel je teprve možné s XML volně pracovat

Původně XML vzniklo jako odnož mnohem složitějšího jazyka SGML (Standard General Markup Language), jehož hlavní účel byl učinit data (především textová) snadno zpracovatelná strojem. XML se ustanovilo již s předpokladem, že bude vytvářeno se sémantickými omezeními. Původní myšlenka tedy zůstala zachována, ale přibyl k ní (byť

primitivní) sémantický rozměr. Samotné XML se skládá z *entit* neboli úložných jednotek, obsahujících syntakticky analyzovatelná nebo neanalyzovatelná data. Syntakticky analyzovatelná data (dále jen *parsovatelná* data) se dělí na značky (meta-tagy) nebo na znaková data. Značky umožňují hierarchizaci dokumentu a poskytují logickou strukturu jejímu úložišti. Pomocí této struktury pak k datům může přistupovat tzv. XML procesor, ovládaný softwarovou aplikací (57).

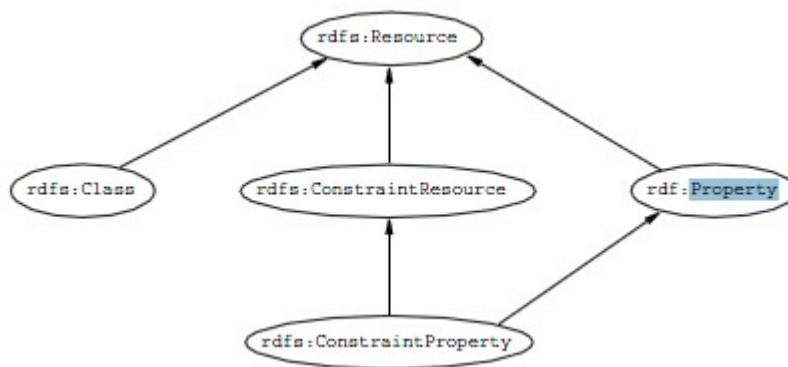
To, jak je XML dokument strukturován a jaké značky smí použít, definují jazyky (slovníky) řazené do skupiny *XML schema*. Jaké XML schema dokument využívá je definováno zpravidla v deklaraci dokumentu. Jednotlivé části dokumentu však mohou využívat různá schémata. Toho se dá snadno docílit pomocí prefixů, tak jak je popsáno výše v části věnované XML namespaces. Dokumenty napsané podle konkrétního schéma mohou pak být aplikacemi kontrolovány, zda odpovídají deklarovanému schématu a jsou tedy validní. Správně napsaný XML dokument, který projde parserem nemusí být vždy nutně validní – jelikož pro jeho potřeby nemusí být podstatné, aby odpovídal ustanoveným standardům například pro využití na WWW (může jít například o specializovanou intranetovou aplikaci pro B2B prostředí, kde se XML hojně využívá). Nejznámější XML schema jazyk je DTD (Document Type Definition). Problém DTD je, že používá poměrně velmi omezenou syntax, která XML dokumenty napsané podle tohoto schématu v podstatě zbavuje možnosti poskytovat nějaká skutečně hodnotná sémantická data. To na rozdíl od DTD umožňuje W3C XML Schema Language (který se často milně slučuje s XML schema). Díky němu je možné vložit do dokumentu i tak konkrétní strojově zpracovatelné sémantické prvky jako například informaci typu: „tento dokument je datován mezi lety 1999 a 2021“ (36, s. 280). Právě proto RDF aplikace často využívají XML Schema Language za vzniku RDF/XML syntaxe, o níž bude řeč později.

Celkově vzato ovšem XML není určeno k popisu smyslu dat. Poskytuje jednotný rámec a nástroje pro výměnu dat mezi aplikacemi, ale sémantické možnosti jsou spíše jeho vedlejší produkt. Struktura a vnořování tagů XML je postavená tak, že záleží plně na aplikaci, jak si jí interpretuje a nemá tudíž pevně daná omezení. Z těchto důvodů je XML pro potřeby sémantického webu nedostačujícím jazykem, i když je pro vývoj dalších sémantických webových jazyků klíčové.

2.2.2. Jak RDF funguje?

Jak už bylo napsáno, RDF je jako datový model složen z logických trojic. Ty mohou reprezentovat tyto veličiny: objekt, atribut a hodnota (46, s. 51). Vlastní trojice lze v RDF vytvářet ve složení: *subjekt, sloveso (predikát), objekt*. RDF trojice je možné abstrahovat a chápat jako jednoduchý graf, kde subjekt a objekt jsou uzly a sloveso je hrana orientované určitým směrem (48). Taková trojice se pak nazývá *výrokem*. Tímto způsobem složený výrok – graf může snadno reprezentovat URI pomocí vztahu mezi uzly. Podobnost s *formální logikou* zde není náhodná - RDF z ní totiž vychází. RDF tedy staví na modelové sémantice, která funguje tak, že definuje pravidla, pomocí kterých program identifikuje, kdy je výrok platný a kdy ne a jaký závěr je ze série výroků možno vytvořit (26 s. 465). RDF používá ke své reprezentaci zpravidla XML syntax, což ovšem neznamená (jak bylo již uvedeno), že by jiné syntaxe nebyly možné. RDF nespécifikuje žádnou oblast ani typ svého využití, což je podstatná vlastnost tohoto datového modelu. Pro specifické využití RDF je zapotřebí RDFS. Volba názvu pro RDFS nebyla nejšťastnější, jelikož se nabízí analogie s XML Schema. Ve skutečnosti se však jedná o dvě zcela rozdílné věci. Zatímco XML Schema uvádí omezení a pravidla pro strukturu stavby XML dokumentů, RDFS je systém pro určování slovníku vhodného pro využití RDF v konkrétní aplikační sféře (66, s.62). Pomocí RDFS tedy můžeme například určit, ke kterým objektům se mohou vztahovat jaké vlastnosti, a jaké mohou pojmout hodnoty. Je také možné identifikovat vztahy mezi objekty. Ve skutečnosti je to tedy tak, že až vlastní aplikace RDFS na RDF datový model umožňuje danou reprezentaci informace zpracovat softwarově a zároveň sémanticky zcela korektně.

Rozebereme-li detailně části RDF trojice, zjistíme, že se jedná zpravidla o různé typy zdrojů reprezentované pomocí URI. Představit si objekt jako zdroj není komplikované, u vlastností je to ovšem už složitější. Pro pochopení vlastnosti jako zdroje nám pomůže následující graf (66, s. 87):



Jedná se o zvláštní typ zdroje, jehož esenciální vlastností je to, že znázorňuje nějaký vztah. V praxi (máme-li na mysli tu webovou) jde zpravidla o URL. Hodnoty pak mohou být také zdrojem, ale může se jednat i o atomickou strukturu, která se nijak dále neurčuje (66, s. 64). Trojice - tedy výroky mohou být propojeny do sémantické sítě s dalšími výroky a tím tvořit další sady trojic vyplývající z těchto spojení. Zároveň je možné použít výroky jako zdroje a vytvářet výroky o těchto výrociích. V RDF se jedná o běžnou praxi.

Aplikace určené ke zpracování zdrojů popsaných pomocí RDF (resp. RDFS) musí disponovat plnohodnotným dotazovacím jazykem, podobně jako je tomu třeba u databází. Proto vznikl SPARQL – dotazovací jazyk na bázi SQL určený k dotazování v RDF dokumentech. Protože RDF je postaveno na znakové reprezentaci grafů i SPARQL využívá k dotazování systému grafů. SPARQL pak porovnává dotazovací „graf“ s vloženým vzorem. Výsledné odpovědi vrací v tabulkové formě, případně jako šablonu (za předpokladu, že nenalezne odpovídající odpověď) určenou ke tvorbě nových RDF dat. Samotná aplikace RDF v rámci RDFS jazyka/slovníku vypadá například následovně:

rdfs:Class (nebo rdf:type) – třída věcí. Z užití nic neplyne: používá se pro snadnost orientace.

rdfs:Property – sloveso. Řekneme o něčem, že je to sloveso.

rdfs:subClassOf – řekneme, že něco je subtřída.

rdfs:subPropertyOf – řekne o něčem, že toto má nadřazené sloveso.

RDFS nám tak umožňuje mluvit o třídách a vlastnostech věcí a jejich hierarchii (48).

2.3. Co je to v počítačové vědě ontologie a proč je důležitá pro sémantické webové jazyky

Původní řecký význam slova ontologie znamená doslova „řeč o tom jsoucím“, v přeneseném významu se pak jedná o filosofickou disciplínu zabývající se povahou bytí. V informatice však asi nejznámější definice ontologie zní: „Ontologie je specifická reprezentace konceptualizace,“ (40, s. 907-928) (příčemž konceptualizací se míní abstraktní model světa). Počítačová věda si tento termín tedy vypůjčila z filosofie, ale přidala k němu předložku „an“ (v anglickém jazyce) aby náležitě odlišila oba dva výrazy. Vzniklý termín „an ontology“ se používá pro označení formální reprezentace jakékoliv množiny konceptů (48). Zároveň ontologie určuje terminologii, taxonomii, klasifikaci, vztahy mezi koncepty a axiomy instance. Jinými slovy ontologie v počítačové vědě „vysvětluje“ – tedy popisuje veškeré koncepty ve své množině. Koncepty jsou pak třídami objektů (66, s. 10). Tyto mohou reprezentovat doslova cokoli, od hmotných živých bytostí, přes tabulky až po čistě abstraktní entity jako například čísla.

Pro vývoj sémantických webových jazyků je ustanovení ontologie zásadním momentem, bez něhož nemohou fungovat. Esenciální vlastností ontologie je její strojová čitelnost. Později upravená základní definice ontologie říká: „Ontologie je formální specifikací sdílené konceptualizace (58, s. 12).“ Významný posun zde nastává právě ve slově „formální“ jelikož to určuje, že ontologie musí být strojově čitelná – jinak by pro počítačovou vědu jako koncept neměla žádný smysl. Samotná ontologie ve smyslu sémantického webu jde však ještě dále, protože její funkčnost pro potřeby strojového zpracování musí být postavena na určitém logickém systému. Proto vznikly i další definice ontologie, které mnohem více vyhovují těmto potřebám: „Ontologie je logická teorie poskytující explicitní parciální vysvětlení konceptualizace (39).“ Ontologie využívané pro vývoj sémantických jazyků jsou z pravidla postaveny na deskripční logice, a proto je jejich definice jako logických teorií poměrně přesná.²

I v samotné počítačové vědě (a pochopitelně i mimo ní) občas pojem ontologie milně zaměňuje s pojmem taxonomie. Je to de facto degradace konceptu ontologie na pouhý systém kategorizace a hierarchizace tříd, podtříd a jejich definicí. Ontologie se však na rozdíl

² Definicí pojmu ontologie existuje v počítačové vědě ještě celá řada a nemělo by smysl je zde všechny vyjmenovávat.

od tradiční logické taxonomie neomezuje na pouhý popis, ale na základě axiomatických principů umožňuje sdílet určitou znalost o světě. Za určitých okolností se mohou za ontologie považovat taxonomické systémy například v oblasti e-commerce nebo webové vyhledávací systémy, ty však téměř rezignují na jakákoliv omezení pro sémantiku své oblasti a neposkytují sami o sobě proto v podstatě žádnou znalost. Proto se ontologie někdy dělí na tzv. *odlehčené* a *plnohodnotné ontologie* (49, s. 3). *Odlehčené ontologie* se užívají v dnešních

ne-sémantických webových technologiích poměrně běžně a mají blíže k již zmíněným taxonomiím. Plnohodnotné ontologie jsou takové, kterými se zabývá programování znalostních systémů, a jsou sémanticky relevantní.

2.3.1. Z čeho se ontologie skládá?

Jak bylo popsáno v minulé kapitole, aby bylo vůbec možné vytvořit nějakou znalostní bázi, je nutné nejdříve připravit ontologii, na níž bude postavena. Stavba ontologie se skládá z dvou částí – 1) ze znalostního modelování (definuje třídy, role, koncepty, vztahy mezi nimi atd.) a 2) z paradigmatu znalostní reprezentace, v němž se pohybujeme (definuje formální rámec, užití deskripční logiky nebo logiky prvního řádu atp.) (50, s. 9). Vlastní stavební komponenty ontologie jsou následující: *třídy* (reprezentují určité koncepty – např. automobily, lidé); *vztahy* (ty reprezentují propojení, souvislosti a návaznosti mezi třídami definovanými koncepty); *funkce* (to jsou speciální typy relací mezi koncepty, u nichž je element n unikátní pro všechny elementy předcházející elementu $n1$); *formální axiomy* (modelové – vždy pravdivé – výroky); a *instance* (reprezentující elementy nebo individuality v ontologii) (7, s. 87). Komponenty, které odlišují plnohodnotnou ontologii od odlehčené a od běžné taxonomie jsou pak zpravidla právě funkce, instance a zejména axiomy.

2.4. OWL a DAML – ontologické programování v praxi

2.4.1. DAML + OIL

Sémantický web se rozvíjí velmi dynamicky a záhy po ustálení RDF a RDFS na konci 90. let minulého století přišly výzkumné týmy pracující především pod hlavičkou W3C na to,

že tyto systémy jsou samy o sobě nedostačující. V řadě případů se relativně omezená expresivita samotného RDF a jeho nadstavby RDFS (v případě RDF na binární vztahy predikátů a u RDFS na hierarchii podtříd) ukázala jako nevhodná pro aplikace, umožňující plné využití možností sémantického webu. Proto vzápětí začaly vznikat nové sémantické jazyky, které už byly stavěny od počátku na principu ontologií (od toho ontologické jazyky) (66, s. 109-110). Tak přišly na svět specifikace DAML+OIL (DARPA Agent Markup Language + Ontology Interchange Language) a OWL (Ontology Web Language). Jestliže jsme si řekli, že ontologie se skládají ze znalostního modelování – definice tříd, rolí atd. a z paradigmatu znalostní reprezentace – definic formálního rámce, pak z toho vyplývá, že ontologický jazyk musí mít tyto náležitosti: dobře definovanou syntax, formální sémantiku, ustálené vyjadřovací konvence, efektivní logický základ a dostačující expresivní potenciál (66, s. 110).

DAML se původně vyvíjel nezávisle na W3C ve výzkumném experimentálním centru americké armády DARPA (Defence Advanced Research Projects Agency). Jeho vývoj začal v roce 1999 a záměrem bylo vytvořit bohatší slovník, než jaký nabízelo doposud využívané RDF/XML a RDFS. Vyšlo totiž najevo, že slovníky využívané pro RDF do té doby, založené na RDF grafovém modelu, postrádaly dostatečné množství datových typů a ustálené výrazivo pro výpočty. Prvním slovníkem vhodným pro rozšíření (jinak značně omezených) možností RDF byl DAML-ONT, který ihned následoval OIL (Ontology Interference Layer – vyvinutý výzkumnými týmy v EU) za vzniku DAML-OIL využívajícím konstrukty z tzv. rámcových systémů (Frame-based AI systems) (18). Ty jsou vlastně jednoduchými znalostními reprezentacemi, v nichž je každý rámec doplněn balíkem definičních a deskriptivních informací (22). DAML-OIL proto umožňuje například využívat *instituci unikátní vlastnosti*, což u RDFS nebylo možné a docházelo ke zbytečným informačním ztrátám. DAML-OIL má tak jako tak samozřejmě (stejně jako všechny sémantické jazyky/slovníky) svůj základ v XML syntaxi a XML Schema strukturování, zároveň využívá pochopitelně i RDF datového modelu, který ovšem výrazně rozšiřuje. Umožňuje například mnohem přesnější specifikaci skládaných datových formátů. Jestliže se u běžného RDFS určil datový typ hodnoty nějakého subjektu pomocí nějaké vlastnosti třeba takového typu:

```
<rdfs:Property rdf:ID="productNumber">  
<rdfs:label>Product Number</rdfs:label>  
<rdfs:domain rdf:resource="#Product"/>
```

```
<rdfs:range rdf:resource="http://www.w3.org/2000/01/rdf-schema#Literal"/>
</rdfs:Property>
```

pak počítač získal znalost o tom, že tento subjekt bude řadou znaků (18). Jestliže se na výše uvedenou informaci podívá člověk, pravděpodobně usoudí, že reprezentací tohoto subjektu označeného jako „product number“ (tedy číslo produktu) bude číslo. Takovou informaci však z daného namespace, které mu RDFS určuje, počítač nevyčte. Je tedy postaven před nejednoznačný výrok. Pomocí DAML-OIL by se však stejná informace zobrazila takto:

```
<daml:DatatypeProperty rdf:ID="productNumber">
<rdfs:label>Product Number</rdfs:label>
<rdfs:domain rdf:resource="#Product"/>
<rdfs:range rdf:resource=
"http://www.w3.org/2000/10/XMLSchema#nonNegativeInteger"/>
</daml:DatatypeProperty>
```

Díky RDF hodnotové specifikaci **DatatypeProperty** tak DAML-OIL omezí datový formát na jednoznačnou množinu možných vstupů. Pro vložení této informace bylo, jak je možné vidět výše, použito namespace pro XML Schema. Stejně tak lze však využít vlastního DAML-OIL namespace, nebo vytvářet uživatelské hodnoty dle potřeby.

Podobně jako v případě datových formátů rozšiřuje DAML-OIL možnosti RDF/XML a RDFS v případě *unikátních vlastností subjektů* (subject unique properties) a přichází s novým pohledem na systém tříd. Právě ten je hlavním přínosem pro RDF ze strany DAML-OIL. **Daml:Class** je totiž definována jako podtřída **rdfs:Class**. V RDF systému často dochází na identifikaci třídy pomocí výčtu jejích členů. Na tom by ještě nebylo nic komplikovaného. Problém je, že v RDFS není konstruováno na to, že takovýto výčet by měl být uzavřenou množinou. DAML-OIL nabízí mnohem přesnější identifikaci pomocí elementů: **daml:oneOf** a **daml:collection**, kdy určuje, že daná třída je omezena přesnou skupinou uzavřených instancí (8).

Všeobecně byl DAML prvním opravdu významným posunem k mnohem komplexnějšímu jazyku vhodnému pro potřeby sémantického webu. Jeho vývoj už však v současné době prakticky nepokračuje, jelikož byl de-facto převzat do OWL a nadále běží již zcela pod patronací světového webového konsorcia W3C.

2.4.2. OWL

Na tomto místě je vhodné si znovu připomenout, co dělá ontologický jazyk ontologickým jazykem: Je to především formální sémantika. Její hlavní vlastností je to, že popisuje smysl vědomosti naprosto přesně – tedy tak, aby jej žádní dva lidé (a především stroje) nemohli interpretovat nestejným způsobem. Formální sémantiku si ontologické programování vypůjčilo z oboru matematické logiky, kde se využívá především pro potřeby usuzování (reasoning). V znalostní ontologii lze takového usuzování využít například k určení náležitosti do určité třídy, rovnosti mezi třídami, klasifikaci či konzistentnosti syntaxe (66, s. 110).

Tento typ sémantiky je předpokladem k procesu ověřování a zdůvodňování, pomocí něhož lze zjistit inkonsistence v ontologiích, zkontrolovat nechtěné vztahy mezi třídami a automaticky klasifikovat podmnožiny v třídách. Vytvořit silný ověřovací nástroj společně s přesnou vyjadřovací schopností sémantického jazyka kombinovaného s plnohodnotnou logikou byla hlavní motivace pracovní skupiny W3C pro práci na OWL.

Jazyková specifikace OWL se dělí na tři kategorie: je to OWL Full, OWL DL a OWL Lite. OWL Full umožňuje využívat jazykových primitiv a lze z ní určovat omezení pro celé instance a především skupiny instancí. Kromě toho je také OWL Full při dodržení některých formálních předpokladů kompatibilní s RDF a RDFS. Problém této verze OWL však je jeho komplexita. Jde o tak složitý systém, že de facto zabraňuje využití nějakého efektivního ověřovacího procesu (37).

Další dvě kategorie OWL jsou ve skutečnosti jen podkategoriemi OWL Full. V OWL DL (DL je zde zkratkou pro description logic) je možné použít stavební prvky OWL a RDF, systém například zabraňuje kombinování různých stavebních tagů a tím zajišťuje kompatibilitu celého systému odpovídající fungující deskripční logice. Nevýhodou OWL DL je, že se stává zcela nekompatibilní s RDFS. Poslední variantou OWL je OWL Lite, což je jazyk zcela vycházející z OWL s tím, že je založen na ještě větším množství omezení. Například zcela eliminuje výčty tříd, *nespojitelné výroky* a *arbitrární kardinalitu*. Tato omezení zajišťují, že pro vývojáře OWL aplikací je práce s OWL Lite mnohem snadnější a efektivnější. Samozřejmě je to na úkor vyjadřovací (expresivní) síly jazyka.

OWL je doposavad nejrozvinutějším stupněm sémantického webového jazyka. V praxi se však na webu zatím příliš neuplatnil, jelikož jeho nasazení brání jeho formální rozsáhlost a také nekompatibilita příslušných softwarových aplikací.

2.5. Principy práce se sémantickými technologiemi – doplnění

2.5.1. SPARQL – dotazovací jazyk pro datový model RDF

Stejně jako klasické webové programovací jazyky potřebují pro komunikaci se serverem dotazovací jazyk (SQL), i sémantický systém ontologií vyžaduje speciální způsob dotazování. Ten zajišťuje zvláštní jazyk SPARQL vytvořený přímo za tímto účelem. Jde o poměrně prostý princip, kdy SPARQL porovnává vzory RDF grafů cílových dat s grafem, který vytvoří z příslušného dotazu. Tyto vzory jsou v podstatě stejné jako klasické RDF grafy s tím rozdílem, že mohou obsahovat proměnné místo některých zdrojů či predikátů (46, s. 56).

Pomocí SPARQL lze tedy především získávat informace z RDF grafů v podobě URI, zdrojových uzlů nebo znaků. Kromě toho je možné rozklíčovat podgrafy RDF a následně konstruovat nové RDF grafy na základě těch, které vznikly ze samotného dotazu (46, s. 56).

2.5.2. Deskripční logika – vzor pro sémantické dokazování

Základní motivací, jež stála za zrodem systému deskripční logiky, byla možnost formální reprezentace znalosti (o světě) na co nejvyšší úrovni. To by samo o sobě ještě nebylo nic nového, protože stejný úkol mají v podstatě všechny logické systémy. Deskripční logika však již od svého vzniku přímo navazuje na ne-logické způsoby zobrazování sémantických vlastností znalosti, pro něž je zpravidla využíváno grafických prvků a rámců. Její následné využití v počítačové vědě, proto například umožňuje výstavbu inteligentních systému, které jsou schopné z formální reprezentace externích dat (znalosti) extrahovat implicitní konsekvence (23).

2.5.3. DRY - pravidlo „Don't repeat yourself“

DRY je poměrně prostý princip, který je jedním z hlavních předpokladů sémantického webu a sémantických aplikací. Jde o poučku, jak již z anglické „don't repeat yourself“ vyplývá, kdy je pro vývoj jakékoliv znalostní báze (či ontologie) zcela esenciální, aby měla každá obsažená informace – znalost pouze jednu jedinou jednoznačnou reprezentaci. Dodržování tohoto pravidla z podstaty zajišťuje funkčnost jakéhokoliv sémantického jazyka postaveného na funkčních logických pravidlech (9).

2.5.4. Mikroformáty – sémantické prvky v HTML

Nutnost sémantického zpracování dat pomocí metatagů nepodceňují ani zastánci současných webových technologií. Právě z toho důvodu vznikla i jakási třetí cesta, kterou by se vývoj současného WWW mohl ubírat. Na rozdíl od čistě sémantických datových modelů a jazyků, které otázku zobrazování a stylování dat nijak neřeší, jsou tzv. *mikroformáty* zakomponovány přímo do běžných HTML šablon a dodávají tak, jinak zcela omezenému stylovacímu jazyku, sémantický rozměr.

Běžný HTML kód doplněný o sémantický tag určující například něčí jméno (zpravidla provozovatele webové prezentace) může vypadat následovně:

```
<div class="vcard">
<div class="fn">Eduard Piňos</div>
<div class="nickname">Euripides</div>
<div class="org">mediaform</div>
<div class="tel">+ 420 000 000 000</div>
<aclass="url"href="http://www.skoteinos.smiley.cz/">http://www.skoteinos.smiley.cz/
</a>
</div>
```

Hlavní účel mikroformátů od začátku byl umožnit strojové zpracování některých speciálních dat, ukrytých v kódu. Jedná se hlavně o události, data či geografické lokace a odkazy na jiné weby. Obzvláště takto označené odkazy se ukázaly jako vhodný nástroj pro

zvýšení možnosti zaznamenání těchto propojení vyhledávači a i společnosti jako Google oficiálně oznámily přijetí mikroformátů.

Skvělým příkladem využití mikroformátů jsou tzv. *sémantické wikipédie*, které umožňují ideální přechod z pouhé prezentace k základní formální znalosti přidáním například sémantického odkazu. Zároveň je z takové formální „vědomosti“ možné získat abstraktní reprezentaci informace za využití dotazování a dynamických stránek (59).

Problémem tohoto kompromisního řešení původní myšlenky sémantického webu (tedy vložit na WWW takový typ dat, který bude snadno strojově zpracovatelný a bude vytvářet znalostní báze) je to, že právě na svůj původní účel tato technologie tak trochu rezignuje. Mikroformáty sice umožňují vložit do HTML jistou sémantickou znalost, avšak nevytváří žádné formální ontologie či implicitní dynamické znalostní báze založené na předem dané konkrétní logické struktuře (34).

Dalším výrazným omezením mikroformátů je fakt, že vznikají především v open source komunitě vývojářů a to zcela nezávisle na ostatních systémech sémantického webu. Vyžadují tak například jiný – unikátní parsovací (ověřovací) software než XML či RDF. Tím vzniká další nezávislý systém, jenž sice přispívá myšlence sémantického webu, zároveň je však jeho slepou uličkou. Paradoxní situací je, že právě mikroformáty jsou cestou, kterou si pro aplikaci sémantických prvků zvolila nová specifikace standardu HTML - HTML5, za níž stojí organizace WHATWG (Web Hypertext Application Technology Working Group), přímo napojená na W3C. Společně s ní vidí v této syntéze HTML a sémantických metaprvků budoucnost i velcí hráči na poli internetového vyhledávání v čele s už zmiňovaným Google a Yahoo (nutno však dodat, že Google již dlouhou dobu podporuje také rozvoj XML) a provozovateli WEB 2.0 serverů a aplikací, což staví původní myšlenku sémantického webu Tima Bernerse- Leeho do obtížné pozice.

3. SOUČASNÝ STAV SÉMANTICKÉHO WEBU V PRAXI A JEHO EKONOMICKÉ ASPEKTY

Sémantický web je pouhým pokračovatelem starých myšlenek o logice informací a jejich zobrazování. Přitom jde o zásadní posun v celkovém konceptu WWW jako takového.

Vše lze shrnout do prostého sdělení, že sémantický web znamená skutečnou síť dat a nejen síť neorganizovaných informací. Zde již tolikrát omílaný princip počítačem zpracovatelných informací je totiž zcela zásadním faktem, který mění celý koncept fungování WWW. Počítače převyšují lidské logické schopnosti především v rychlosti. Umožníme-li jim tedy, aby vykonávaly jisté logické operace za nás – jako například rozpoznání smyslu dotazu a automatické třídění vyhovujících odpovědí – výrazně tak urychlíme získávání relevantních informací a výrazně rozšíříme svou znalostní kapacitu. To však není vše. Sémantický web, tak jak ho vidí někteří síťoví vizionáři, může přejít až v něco, co by se dalo nazvat *adaptivním hypermédiem* (13). Jak se již z názvu dá usoudit, jde o nový typ média, který nabízí mnohem větší personalizaci, než současný WEB. Tato personalizace však neprobíhá na základě vizuálního stylu zobrazování (jak to již dnes známe například z celé řady WEB 2.0 aplikací), nýbrž na poli třídění dat a práce s informacemi. Každý uživatel by tak mohl získat relevantní soubor dat upravený na základě jeho preferencí přesně pro jeho potřebu. Přičemž by k tomu využil stejnou webovou aplikaci, jako milióny jiných uživatelů po celém světě.

V současném WWW toto není možné, jelikož zobrazovaná výsledná informace je (na jakkoliv komplexně definovaný dotaz) právě jedna.

Dalším zásadním faktem, který lze už nyní pozorovat v praxi, je proměna fungování některých sociálních aplikací a WEB 2.0 v závislosti na pronikání sémantických prvků a principů do praxe. Tzv. *folksonomie* je ideálním příkladem interoperability sémantického webu a WEB 2.0. Folksonomie je forma kategorizace webových zdrojů, kterou vytvářejí sami uživatelé webových aplikací. Jde o soubor tagů (značek) s jedním nebo více klíčovými slovy, jež se zdají tomu či onomu uživateli relevantní ve spojitosti s daným zdrojem (20, s. 729). Na základě většinové shody jsou pak zdroje kategorizovány dle jednotlivých klíčových slov. V kombinaci se sémantickým zobrazováním dat na webu (například pomocí mikroformátů) zde vzniká zcela nový způsob nakládání se zdrojem informací. Uživateli vytvořený obsah webu je také uživateli kategorizován ovšem s jednou zásadní výhodou, a sice možností strojového zpracování takového systému. Lidská schopnost náhledu na věc (informaci) v kontextu se pojí s obří kapacitou logických operací stroje za vzniku zcela nového typu znalostní báze.

Otázkou je, proč je sémantický web, skýtající tak rozsáhlé možnosti nejen v oblasti WEB 2.0 ale i pro běžné uživatelské aplikace (například v oblasti time-managementu)

a především *enterprise* a B2B sféru, neustále na jakési druhé koleji, jde-li o jeho uvedení do běžného použití. Navíc když RDF standard a navazující sémantické webové jazyky a specifikace jsou vyvíjeny už řadu let a de facto jsou připraveny pro masové požití. Jak je tedy možné, že vývojáři ale i velké korporace a klíčový hráči na poli WWW se k této možnosti staví stále neutrálně až odmítavě? Proč již nenastala éra WEB 3.0 – sémantických aplikací, vyhledávání v přirozeném jazyce a automatizovaných softwárových agentů? Existuje zhruba šest hlavních důvodů k tomuto stavu:

1. Současný WEB už tak zpřístupňuje velké množství osobních dat, se sémantickým webem by tato redukce anonymity mohla (za určitých okolností) nabrat zcela nový (pro většinu lidí děsivý) rozměr.

2. S předešlým souvisí hrozba zneužití takovýchto informací například k cílenému kontextovému spamování a dalším možným negativním jevům.

3. Jeden z hlavních technických problémů je (prozatím) neexistence tzv. softwárových agentů, co by uměly efektivně číst přirozený jazyk a poskládat odpověď na dotaz například v jiném pořadí slov než ve vyhledaném dokumentu (či dokonce více dokumentech). Tento problém částečně řeší existující sémantické technologie v čele s RDF a OWL, ale i pro čtení informací strukturovaných pomocí těchto metadat v podstatě žádné 100% funkční nástroje ještě neexistují (byť se na jejich vývoji intenzivně pracuje).

4. Změna paradigmatu „hodnoty“ na internetu - v současnosti probíhá monetizace většiny komerčních webových služeb (nejedná-li se přímo o služby placené uživatelem) pomocí reklamních sdělení. Sémantický web se svou možností zcela personalizovaného brouzdání po WWW (respektive získávání informací, znalostí) přímo ohrožuje možné příjmy z reklamy. Jedná se o podobný problém, na nějž naráží provozovatele například informačních portálů již dnes při využívání různých *content delivery feeds* – tedy kanálů pro šíření nového obsahu webu pomocí speciálního čtecího softwaru (například RSS).

5. Nedořešenou otázkou je také nekompatibilita slovníků používaných různými uživateli a různými softwárovými systémy. I v rámci jednoho jazyka může existovat celá řada slovníků, které jeden výraz interpretují každý jiným způsobem, může se tak stát, že implicitní informace, které data sémantického webu obsahují, mohou být počítačem interpretovány nesprávným způsobem (11).

6. Posledním a možná nejpodstatnějším důvodem k tak pomalému nástupu sémantického webu do praxe je konzervativní postoj a „vyčkávací taktika“ velkých

internetových hráčů v čele s nejrozšířenějšími světovými vyhledávači a především společností Google.

V následující části práce si tyto skutečnosti pokusíme podrobněji objasnit a zároveň si poukážeme na již existující sémantické aplikace a jejich možný dopad na ekonomiku či fungování internetu jako takového.

3.1. Proměna webových služeb s příchodem sémantických technologií

Využití technologií sémantického webu by mohlo mít v praxi mimo jiné velký dopad na způsob fungování tzv. webových služeb (*web services*). Tato, dnes již samozřejmá, součást WWW byla v minulosti hlavním motorem masového rozšíření webových technologií a především zdrojem ekonomické síly, kterou současné počítačové sítě skýtají. Webovou službou může z principu být v podstatě jakákoliv webová stránka poskytující specifickou odpověď na nějaký dotaz. Tak můžeme za webovou službu považovat například jakýkoliv vyhledávač, jehož výsledná URL nám dává odpověď na náš vyhledávací dotaz (17, s. 172). Přesnější definice webové služby však uvádí, že jde o softwarové aplikace, které mohou být nalezeny, popsány a zpřístupněny pomocí XML rámce a standardních webových protokolů (míněno HTTP). Tyto služby pak lze distribuovat pomocí extranetů, intranetů nebo samozřejmě Internetu (47, s. 58). Jiná definice by nám pak mohla například říct, že webová služba je něco, co umožňuje nějakou formu komerčního využití komunikace mezi obchodníkem (poskytovatelem služeb) a klientem pomocí klient- server technologií. Tato definice však pro problematiku sémantického webu není až tak důležitá.

Vraťme se nyní k XML. Naprostá většina aktivních softwarových aplikací fungujících on-line, ať už se jedná o B2B systémy, rezervační servery, e-shopy či například CRM (Customer Relationship Management) systémy, je postavena právě na XML syntaxi. Pro komunikaci se servery pak používají webové služby, jak už bylo naznačeno, protokol HTTP. Pro komunikaci mezi serverem a webovou službou je užíváno zpráv zasílaných v protokolu SOAP (Simple Object Application Protocol) (47, s. 58). Všechny tyto standardy byly přijaty proto, aby se vyřešil hlavní problém internetového obchodu – integrace mezi jednotlivými aplikacemi a systémy. Drtivá většina společností a provozovatelů webových služeb nyní považuje právě XML a SOAP za jediné využitelné technologie, jelikož jakákoliv shoda

v záplavě technických řešení na tak širokém poli jako je WWW, je shoda velmi obtížně dosažitelná. V minulých deseti letech se s obrovským množstvím nových webových technologií ukázalo, že klíčem k úspěchu není (bohužel) jejich technická vyspělost, ale jejich adaptabilita a masové rozšíření. Na využití SOAP jako standardu v rámci aplikací webových služeb se shodli světoví vývojářští giganti v čele se společnostmi Microsoft, Sun či IBM dokonce i společně s open source komunitou – což je znamením toho, že tato technologie má mimořádný potenciál na dlouhodobý úspěch (47, s. 63).

Jenomže, co když RDF a jeho slovníky jako OWL nabízejí mnohem komplexnější a flexibilnější využití webových služeb a jejich interoperabilitu s inteligentními agenty a jiným softwarem? Proč tedy není možné ihned přejít na sémantickou technologii, nabízející webovým službám nový rozměr automatizace celé řady procesů, při kterých je v současnosti stále zapotřebí nadbytečné lidské intervence? Odpověď je snadná a vyplývá jak z výše uvedeného, tak z bodů vysvětlující hlavní problémy uvedení sémantického webu v praxi popsaných v předešlých odstavcích.

Předním důvodem, proč webové služby zůstanou ještě dlouho při XML syntaxi a SOAP komunikačním protokolu, je pohodlný (avšak z obchodního hlediska zcela logický) postoj vývojářských firem. Ty se snaží co nejdéle těžit zisk z úspěšně rozšířené technologie, místo toho, aby se pouštěly do riskantního podniku s novými neozkoušenými technologiemi sémantického webu. Navíc ani v komunitě lidí kolem těchto technologií nepanuje naprostá shoda a ony samy mohou podléhat ještě zásadním vývojovým změnám. Taková technologie však není pro firmy pod tlakem akcionářů, vyžadujících neustálý růst, žádoucí – tedy alespoň do doby, než se její využívání stane, v rámci potřeb nových aplikací, nutností a tím i možností pro generování potenciálním ziskem. Druhým důvodem je pak způsob vyvolávání webových služeb a nakládání se získanými daty. Odpůrci sémantického webu totiž tvrdí, že představa automatické komunikace pomocí inteligentních agentů jde přímo proti zásadám WWW a principům, které umožnily jeho tak rapidní a masové rozšíření (17, s. 173). Jde zde především o obavu ze zneužití: Je rozumné nechat obchodování v obrovských částkách a nakládání s privátními daty získanými v rámci webových služeb, plně na komunikaci počítač-počítač místo na řetězci člověk-počítač-člověk-počítač?

Na příkladu webových služeb můžeme snadno pozorovat některé z uvedených důvodů, které brání masovému rozšíření sémantických technologií. Máme zde sice už řadu let

fungující předvoj sémantického webu – v podobě XML, ten však paradoxně (vzhledem ke svému masovému nasazení především v B2B sektoru) zabraňuje adaptaci sofistikovanějších forem sémantických technologií především ze strany enterprise subjektů.

3.2. Co má společného WEB 2.0 a sémantický web a lze je propojit?

Web 2.0, tak jak jej známe dnes, dozajista revolučně proměnil nahlížení na WWW jako informační médium. Zásadně také změnil chování uživatelů na WWW. Díky sociálním sítím a uživateli vytvářenému obsahu se stal WEB mnohem rychlejším a komplexnějším zdrojem informací, než tomu bylo dříve, kdy informační tok drželi v rukou především designovaní jednotlivci či organizace. Problém WEB 2.0 ve spojitosti se sémantickými technologiemi je však poměrně zřetelný. Existuje zde sice komunikační triumvirát člověk-počítač-informace, avšak tento je zcela nerovnoměrný (53). Jak již bylo zmíněno, WWW (zastoupeno v tomto případě jeho vyústěním směrem k uživateli – tedy počítačem) je zde v pozici „pouhého“ média – tedy nosiče informací, přitom, jak už víme, potenciál WWW je mnohem větší. Může se snadno stát téměř nevyčerpatelnou znalostní bází, která informace nejen přenáší ve formě dat, ale tyto data zároveň interpretuje. V současném WEB 2.0 přitom je jakákoliv znalost (potenciálně získaná ze zobrazených – přenášených dat) zcela iluzorní do chvíle, než do informačního řetězce zasáhne lidský faktor (53). S pomocí sémantických technologií a sociálních aplikací, které mají přístup k mnohem bohatším uživatelským profilům, by bylo možné propojit uživatele sdílející podobné zájmy na mnohem vyšší úrovni než je tomu možné dnes (25, s. 24).

Zmínka o interpretaci dat je v případě WEB 2.0 obzvláště podstatná, jelikož se v rámci jeho (samovolného komunitního) vývoje ukázala jako jedna z nejzajímavějších technologií folksonomie, o níž už byla řeč. Tento nový typ klasifikace dat stojící v juxtaopozici ke klasické taxonomii založené na hierarchické bázi, přibližuje data kolující po síti mnohem více přirozenému jazyku - činí tedy to, co je jedním z hlavních cílů hnutí kolem sémantického webu: sjednotit přirozený (lidský) jazyk s (počítačům srozumitelným) webovým. K tomuto využívá folksonomie systém tagování – tedy jedné z hlavních (v současnosti dostupných) forem vkládání sémantických metadat do kódu. Překážka však zůstává stejná – máme zde lidmi vytvořenou kategorizaci dat pomocí přirozeného jazyka, kterou jsou počítače schopny hierarchizovat, přiřazovat k zobrazovaným informačním

zdrojům a zahrnovat je do výsledků vyhledávání; stále se však jedná pouze o mediátor mezi člověkem vloženými daty a člověkem – uživatelem, který tyto data musí intencionálně hledat. Tato skutečnost nastává z toho důvodu, že ač aplikace WEB 2.0 ve velkém měřítku zakomponovávají emergentní webové technologie do svého vývoje (jako například AJAX, RSS, Atom, SOAP, XML, FOAF). RDF a dalším sémantickým technologiím se vývojáři naopak zatím spíše vyhýbají. Přílišnou výjimkou nejsou ani mikroformáty, které se sem tam objeví v nějakém projektu příbuzném WEB 2.0. Zpravidla především proto, že za touto technologií stojí silná open source komunita kolem sdružení Mozilla a pro její vlajkovou loď – třetí generaci prohlížeče Firefox už vytvořila celou řadu plug-inů, pro čtení těchto elementárních sémantických dat naroubovaných na HTML.

Odpůrci sémantického webu často argumentují, že datový model RDF a obzvláště jazyk OWL jsou příliš komplexní na to, aby se jej byli schopni naučit běžní uživatelé, kteří jsou ochotní pojmout tak maximálně základy HTML tagů – o XML nemluvě (25, s. 24). A i když je z mnoha příkladů sociálních sítí, blogů a aplikací (Wikipedia, Flickr...atd.) dnes už jasné, že uživatelé jsou ochotni poskytovat metadata, není dořešen způsob, jakým by to mohli dělat. Jednoduchou alternativou (nebo spíše jakousi berličkou pro aktuální potřeby WWW) mohou být právě zmíněné mikroformáty případně primitivní sémantické HTML tagy. Tyto zvládnou používat i zcela běžní uživatelé WWW a umožní tak šířit alespoň povědomí o tom, jak lze výrazně automatizovat a obohatit možnosti WEB 2.0.

Jak je možné, že se však tedy toto zdánlivě logické spojení – sociálních sítí a znalostního webu stále nedaří masově rozšířit? Zde přicházejí na řadu dva z výše uváděných bodů – a sice neochota (či minimálně zdrženlivost) především vyhledávacích gigantů podpořit tyto technologie a změna paradigmatu „hodnoty“ a s ním spojená problematika distribuce reklamních sdělení v prostředí sémantického webu. Společnost Google a její zásadní proměna přístupu k reklamě (všeobecně) je toho krásným příkladem. V klasickém reklamním i internetovém reklamním prostředí platí jakýsi axiom, který by se dal popsat asi takto: *Relevantní je smysl*. Google, který čerpá drtivou většinu svých (nutno dodat astronomických) příjmů z kontextové PPC (Pay Per Click) reklamy, tento reklamní axiom ovšem zkrátka ignoroval a přišel s vlastním řešením. Místo toho, aby analyzoval jemné nuance ve smyslu webových dat a významy datových linků – tedy místo toho, aby de facto z komerční nutnosti uspíšil a připravil nástup sémantického webu (obzvláště v sociálních sítích a uživatelsky orientovaných intuitivních aplikacích), zvolil Google cestu

komplikovaného systému algoritmů a obří analýzy webových dat. Vznikl tak nový reklamní axiom v následující podobě: *Relevantní je statistická podobnost* (25, s.203). Smysl už není podstatný. Dalo by se snadno namítat, že Google je sice klíčovým hráčem na poli WWW, který v komerční úspěšnosti v podstatě nemá žádnou konkurenci, ale i tak nemůže sám o sobě mít takový vliv, aby upozadil technologii s tak obrovským potenciálem. Je třeba si ovšem uvědomit, že valná část sociálních sítí (v čele například s jednou z těch zcela prvních – systémem MySpace) zobrazování Google reklam umožňuje a čerpá z nich značný příjem. Navíc podíl Google na trhu veškeré internetové reklamy je (po akvizici reklamního giganta Double Click z roku 2008) téměř 70%. Z toho vyplývá, že jakákoliv cesta, kterou se tato společnost vydá, zcela zásadně ovlivní veškerý vývoj na WWW. To samé platí i pro samotný WEB 2.0. Úspěch WEB 2.0 aplikací a sítí, je totiž stejnou měrou závislý právě na tom, že se Google rozhodl jít tímto směrem a veškeré své produkty (AJAXovým emailovým klientem Gmail počínaje a vlastní sociální sítí Orkut konče) včetně toho hlavního – vyhledávače (viz možnost vkládat uživatelské komentáře k výsledkům vyhledávání v rámci Google účtu viditelného ostatními uživateli a tím – statisticky - ovlivňovat jejich relevanci) přizpůsobil tomuto trendu.

Kromě Google je zde ještě jedna zásadní síť (společnost), okolo níž se točí vývoj WEB 2.0, a která by mohla propojení sémantických technologií uspišit – Facebook. Překážkou zde ovšem je, stejně jako v případě Google, ekonomická návratnost. Ne snad, že by takové sociálně - sémantické síť s počítačem zpracovatelnými – uživateli generovanými daty neskýtala obrovský komerční potenciál, ale existují zde zásadní komplikace. Facebook také investoval obrovské množství finančních prostředků do analýzy (uživatelských) dat, aby mohl inzerentům nabídnout co nejpřesnější zacílení jejich reklamních sdělení. A tak v závislosti na proinvestované prostředky potřebuje, aby se mu tyto vrátily v podobě zisků ze svých tří zvolených reklamních modelů: *Sociálních reklam* – zobrazující cílenou reklamu v návaznosti na informace získané z uživatelského profilu; tzv. *návnad* - forma reklamy, kdy se mohou fiktivní či skuteční uživatelé (nebo slavné osobnosti) stát fanoušky nějaké značky, či jí pomocí svého profilu jinak propagovat a z *insight marketingu* založeném na hloubkové socio-demografické analýze uživatelů a uživatelských skupin (12).

Z výše uvedeného vyplývá, že všechny tyto reklamní modely, stejně jako systém vlastní analýzy dat, vznikly ještě dříve, než se zdálo, že by sémantické technologie mohly být v takovém stavu, aby je mohli používat doslova všichni uživatelé. I když by právě automatické strojové rozpoznání smyslu obsahu dat například u uživatelských profilů mohlo

několikanásobně zvýšit efektivitu a potenciální přesnost zacílení zobrazovaných reklam, není možné, aby se na tuto technologii přešlo dříve, než se vrátí prostředky vložené do té minulé. Druhým problémem zde zůstává již naznačená komplikace s nedořešeným způsobem distribuce reklamních sdělení v sémantických sítích. Předně by to vyžadovalo adaptaci těchto technologií hlavně také ze strany inzerentů, Pak by zde ale vyvstala mnohem zásadnější otázka: Kdo je cílem reklamního sdělení? Uživatel, jehož kontakty se zajímají nebo by se mohly zajímat (na základě sémantické analýzy) o daný produkt, nebo uživatel sám (65)? I na tuto otázku by mohl počítač automaticky odpovědět právě na základě sémantické analýzy linků mezi uživateli. Jenže co, když si uživatel personalizuje (a zautomatizuje – jelikož to je jeden z hlavních cílů sémantického webu) sociální webovou službu natolik (například tak, že si nechá zobrazovat jen určitá pro něj obsahově relevantní data od svých kontaktů), že se jakémukoliv reklamnímu sdělení vyhne? Tato zásadní otázka zůstává prozatím nezodpovězena a je proto logické, že provozovatele služeb WEB 2.0 jsou stejně jako vyhledávací společnosti k sémantickým technologiím spíše zdrženliví, jelikož prostředí současného WWW je především business dominiem a ztrátu příjmů z reklamy nelze riskovat.

3.3. Co je to WEB 3.0 a co bude po něm?

Příchod WEB 2.0. byl pro další vývoj WWW naprosto zásadní a zcela změnil pohled na směřování a možnosti využívání této světové sítě. I když je tento fenomén vlastně starý jen pár let a proces jeho vývoje ještě zdaleka není ukončen, řada lidí už se poohlíží po tom, kam bude WWW směřovat dál – k WEB 3.0. Mohlo by se zdát, že sémantický web jako takový by se (za předpokladu, že se skutečně rozšíří natolik, jak by si jeho propagátoři přáli) dal označit právě jako web třetí generace. Ovšem zacházet s tímto termínem takto povrchně by bylo poněkud lehkovážné. Právě proto je většina expertů s tímto krokem opatrná. Dalo by se očekávat, že druhou možnou variantou pro vznik WEB 3.0 je sestavit jednoduchou rovnici: WEB 2.0 + sémantický web = WEB 3.0. Koneckonců obrovské možnosti extenze (ale i s tím spojené komplikace) WEB 2.0 aplikací o sémantické technologie už byly zmíněny výše. Ovšem ani v tomto případě nelze takovou úvahu tak snadno přijmout. Existuje zde totiž zásadní rozdíl ve významu terminologie. Zatímco WEB 2.0 označuje de facto jakýsi sociální aspekt WWW technologií, sémantický web je označení především pro nové technologie samotné. Dalo by se namítat, že pod WEB 2.0 se také zahrnují různé webové technologie (například AJAX, RSS či systém WIKI). To je sice pravda, ale vznik těchto nástrojů byl

pouze komplementární součástí aplikací vzniknuvších za nějakým konkrétním účelem sociální aktivity na WWW. Jinými slovy WEB 2.0 je především o interakci uživatelů s WWW. Jenomže v případě sémantického webu se jedná hlavně o novou sadu technologických nástrojů umožňující samotným vývojářům kombinovat data a služby z různých typů zdrojů (25, s. 23). Máme zde tedy v podstatě dva rozdílné přístupy k webu, které lze jen obtížně spojit pod jeden souhrnný název.

Jak tedy ovšem správně určit, co je to WEB 3.0 a jak bude vypadat? Ta správná cesta je začít se dívat na tuto problematiku z jiného úhlu pohledu. Jak již bylo naznačeno, nelze prostě označit za WEB 3.0 něco, co by bylo pokračovatelem WEB 2.0, nejde tuto nálepku přiřadit ani sémantickému webu a dokonce ani kombinaci obojího. Je to proto, že správné nazírání na WEB 3.0 je z perspektivy času. Nejde totiž o technologii jako takovou – ale spíše o její vývojový stupeň v čase (83). Ani to však neznamená, že by vývoj probíhal formou skoku a nástup WEB 3.0 by byl ze dne na den. Není to ani tak, že by současné technologie webu byly opuštěny a zcela odstaveny. Pokusme se spíše představit si WWW jako lego. Veškeré dostupné technologie se využívají pro stavbu jednoho otevřeného systému s obrovským potenciálem. Nová technologická řešení jsou jen dalšími stavebními kameny umožňujícími webu nabízet uživatelům nové způsoby využití. WEB 3.0 by měl všeobecně splňovat tyto cíle: rozšířené možnosti vyhledávání, on-line kolaborace (především mezi systémy) a pokročilá reklamní komunikace, zavedení inteligentních aplikací a transformace webu ze souborového serveru na dynamickou znalostní databázi (83). Takový stav vývoje WWW se dá očekávat někdy mezi lety 2010 až 2020 a prerekvizity k němu můžeme pozorovat vlastně všude kolem sebe. Jako příklad lze uvést třeba rozšíření *mash-up* aplikací, *widgetů*, *social taggingu*, aplikace mikroformátů, RSS a Atom, *RDF úložiště* a využívání OWL pro vládní a korporátní účely, sémantické databáze (DBpedia a další), ale i nová specifikace HTML 5 a další (83). Výrazným krokem k WEB 3.0 je také oznámení masového nasazení webového operačního systému Chrome společností Google, jelikož právě WEB OS je možné považovat za jeden ze znaků pro příchod fáze WEB 3.0.

WEB se bude samozřejmě vyvíjet i dál a určitě se nezastaví ve své třetí generaci, budou následovat další, které se vzhledem k obrovské rychlosti technologického vývoje dají jen těžko odhadovat. Ob jednu fázi tedy k WEB 4.0 se přeci jen vyjádřit lze, jelikož představy o ní přímo navazují na současný stav sémantického webu. Čtvrtá generace WWW by totiž měla přejít z „webu znalosti“ do stavu „inteligentního webu“ (83). Při laických

diskuzích o sémantickém webu se často licituje s problematikou přirozeného jazyka a s umělou inteligencí. Je sice pravda, že právě dotazování (vyhledávání) na WWW v přirozeném jazyce je jedním z cílů komunity kolem sémantického webu, ale současná realita je této možnosti velice vzdálená. Technologie sémantického webu nyní řeší především strojovou zpracovatelnost dat a extrakci imanentní znalosti, která je v nich zahrnuta. Existují sice projekty jako WORDNET (70), snažící se „umělou inteligencí“ obejít a dostat na WEB sémantiku takříkajíc „na tvrdo“ - tedy pomocí databáze vzorů co největšího možného množství sémantických vztahů mezi slovy. I to je samozřejmě určitá cesta k ideálu komunikace s počítačem (webem) v přirozeném jazyce, háček se však skrývá v tom, že se jedná spíše o vývojový krok zpět (či minimálně úkrok stranou), který rezignuje na inteligentní systémy a navíc je výrazně omezen typem jazyka. WORDNET skýtá opravdu široké spektrum sémantických relací od synonym, antonym, homonym, meronym, troponym, holonym až po logické vztahy mezi slovesy (24). Je však pochopitelné, že tato databáze prozatím vzniká pouze v jednom jazyce – angličtině a představa o jejím vytvoření pro všechny jazyky, tak aby taková technologie mohla být otevřená skutečně všem uživatelům WWW, je absurdní. Dalším problémem sémantického webu (jak již bylo naznačeno ve výše uvedeném výčtu) je prozatím neexistence osobních softwarových agentů, kteří by na straně klienta mohli komunikovat s webovými aplikacemi a servery, analyzovat sémantická data a porovnávat je s požadovaným dotazem nebo typem akce. Právě inteligentní agenti a reálná možnost jejich využití nebo běžné používání přirozeného jazyka pro získávání dat na WWW se považují za hlavní znaky WEB 4.0 – čili „inteligentního webu“ (83). I zde je třeba připomenout, že jakákoliv další generace WWW stále navazuje na ty předešlé a z logiky věci jejich technologie a uživatelské aspekty absorbuje a nelikviduje.

3.4. Sémantika, BPM a B2B

Business sféra bývá zpravidla hlavním tahounem, pokud jde o prosazování nových technologií a postupů. Právě v rámci tohoto odvětví už řadu let řeší IT specialisté problematiku automatizace procesů pro řízení chodu (obzvláště velkých korporací – tzv. enterprise) firem. Tyto procesy zahrnují uchovávání, transfer, analýzu a kontrolu veškerých dat spojených například s účetnictvím, řízením lidských zdrojů, sledováním skladových zásob a mnoho dalších. Jde tedy o velice komplexní systém požadavků, vyžadující vývoj neustále nových specializovaných aplikací pro potřeby vzniku nových produktů i služeb. Když tedy

manager (či jiný řídicí pracovník) potřebuje vidět třeba výpis veškerých finančních operací za určité období – musí mu IT oddělení připravit nějaký nástroj, který z uložených zdrojů v databázi požadovaná data vyselektuje (za předpokladu, že takový nástroj zatím ve firmě neexistuje). Manager tedy získá potřebná data, jež může analyzovat. Co když chce ale manager zároveň uvést do prodeje nový produkt, pro nějž (jak usoudil na základě analýzy dat získaných od IT oddělení), je třeba vytvořit novou aplikaci na sledování finančních operací. Musí tedy opět s požadavkem na IT oddělení a to musí aplikaci připravit. Problém tohoto postupu je zde jasný – i když je řada procesů v dnešních firmách automatizována, může se velice snadno stát, že nastane situace, kdy je třeba se vrátit v práci s daty k lidskému faktoru, který je pomalý, neefektivní a omylný (44).

Tuto situaci se snaží organizace řešit pomocí BPM (Business Process Management), jehož hlavní přínos je v důrazu na jednoduchost a přístupnost GUI (Graphic User Interface) business aplikací, tak aby s nimi mohli pracovat především pracovníci bez hlubšího IT vzdělání a nebyli tak při každé operaci odkázáni na IT oddělení. Přehledné GUI je většinou augmentováno o procesní workflow a možnost integrace business aplikací různého typu (44). Toto je možné zejména proto, že enterprise (business) sféra už dávno pochopila výhody XML jako standardu pro syntaxi při vývoji takovýchto aplikací a je tak možné navzájem převádět jednotlivá data aplikací. I přes využívání XML, je ovšem prostor business aplikací v organizaci nepřístupný na sémantické úrovni a jako instance BPM diskurzu je nedostačující (44). Řešení je následující: je třeba sémanticky reprezentovat a popsat každý atomární proces v organizaci, vytvořit kompletní ontologii IT prostředí organizace (hardwarem počínaje, operačními systémy konče), shromáždit znalost o instanci pravidel organizace a zakomponovat je do axiomů vybraného sémantického jazyka, specifikovat dotazovací jazyk a k řešení konkrétních cílů vyžít prostředí sémantických webových služeb (44). Z uvedeného by se mohlo zdát, že organizace by mohly už dávno pracovat na masovém rozšíření takovýchto systému, které by výrazně zefektivnili jejich procesy, stejně jako tomu bylo v případě nasazení XML.

Mimo některých nedořešených otázek ohledně vývoje sémantických jazyků a jejich vhodnosti pro daná komerční odvětví zde ale přichází na řadu již zmiňovaný problém nekompatibility slovníků. V business sféře je tento problém obzvláště palčivý, jelikož pro obchod a dané odvětví je v dnešní globalizované ekonomice nutno specifikovat přesný standard využitelný kdekoli na světě. Vytvořit takto komplexní slovník, sloužící počítači jako dostatečná báze k tomu, aby mohl řídit procesy v organizaci (jako například

mnohamilionové přesuny finančních prostředků) zcela automaticky bez nutnosti neustálé lidské kontroly, je velice komplikovaná vývojářská, lingvistická ale také diplomatická (bereme-li v potaz nutnost jednání na široké platformě světových korporací) záležitost.

Stejná situace jako v případě BPM je i v oblasti B2B aplikací. Obzvláště zde se zasloužilo za masový rozvoj těchto technologií XML, ale zakomponování hlubších sémantických prvků tu podléhá stejným komplikacím jako v případě BPM, navíc problém slovníků (vzhledem k tomu, že jde o systémy umožňující business komunikaci především na nadnárodní úrovni) je tu ještě více nasnadě. Když dojde na komplexní B2B operace vyvstávají zde ve spojitosti s potenciálním nasazením sémantických technologií další dvě důležité otázky: 1) Jak strukturovat složení jednotlivých služeb? 2) Jaká jsou vhodná sémantická omezení pro účast různých služeb v těchto rozdílných složeních (54, s. 118)? Pro B2B (a všeobecně e-commerce služby) by tuto otázku mohli vyřešit softwaroví agenti, kteří by na obou stranách (jak na straně aplikace, tak na straně klienta) automaticky komunikovaly na základě specifikované ontologie. Pro tyto účely vznikl i zvláštní jazyk ASDL (Agent Service Description Language), pomocí něhož jsou definovány procesy a reakce daných webových služeb.

Pokusy reálně zakomponovat RDF do automatických obchodních procesů probíhají už delší dobu, avšak funkčních a úspěšných aplikací je doslova jen pár. Jednou z nich je například RDF systém Sesame (76), který je open source projektem společnosti Aduna. Sesame je postaveno na specifikaci RDF Schema k reprezentaci a dotazování dat a je možné jej zakomponovat do jakékoliv webové aplikace. Pomocí Sesame lze také propojit například B2B nástroje pro integraci a správu business dat s RDF repositářem. Bohužel však i tady brání reálnému využití zřetelná komplikace – tentokrát se jedná o nedostatečnou expresivní sílu dotazovacího jazyka RQL, který Sesame (místo pro sémantický web vhodnějšího SPARQL) využívá (19).

Jak je vidět na příkladu Sesame, zakomponování sémantických principů do každodenního použití v business sféře naráží především na technické problémy a přílišné množství jazyků a specifikací, které jsou často použity v ne zcela ideální kombinaci. V neposlední řadě je třeba zmínit, že adaptace sémantického webu ze strany business komunity vyžaduje ještě jeden důležitý krok – a sice interpretaci odborných pojmů počítačové vědy pro potřeby obchodní sféry. Pojmy jako ontologie, RDF, sémantické úložiště a další jsou pro tzv. „decision makery“ na straně organizací zcela nepodstatné. Je tedy třeba přijít s konkrétní terminologií a příklady, pomocí nichž bude přínos sémantických technologií

pro zefektivnění a urychlení obchodních procesů viditelný i lidem bez hlubšího technologického pozadí (54, s. 9).

3.5. Specializované sémantické aplikace pro běžné uživatele v praxi

I přes komplikace, které byly naznačeny v předešlých odstavcích, se už některé sémantické aplikace začínají objevovat v běžném použití. Stále však existuje silná hranice mezi specializovanými aplikacemi určenými pro zvláštní účely a webovými službami, které mají blízko sociálním aplikacím WEB 2.0 určenými skutečně pro koncové uživatele WWW. Všechny aplikace, o nichž zde bude řeč, jsou však stále poměrně značně vzdáleny představě sémantického webu a zdaleka nenaplňují jeho opravdový potenciál. Velice často také jde o pouhé alfa či beta verze s omezenou funkcí a sémantických technologií se někdy dotýkají jen okrajově. Lze v podstatě tvrdit, že plnohodnotnou aplikaci či pouhou webovou stránku, jež by využívala například jazyka OWL pro kategorizování dat na současném (běžně přístupném) WWW (až na opravdové výjimky) nenajdeme. Na druhou stranu lze narazit na aplikace, které se jistým rysem sémantického webu blíží, a to hlavně v aplikaci metadat umožňující určitou úroveň rozpoznávání smyslu dotazu počítačem. Tyto projekty většinou vycházejí nějakou formou z úspěšného konceptu informační (ne znalostní) báze ve stylu Wikipedia. Zajímavostí je, že parserů, programů na tvorbu ontologií (dokonce i takových, které jsou vhodné pro běžné uživatele založené na přehledném GUI), různých úložišť RDF trojic a dalších nástrojů potřebných pro stavbu sémantického webu je k dispozici poměrně dostatek – zpravidla navíc dostupných pod otevřenými licencemi (nejčastěji LGPL či GPL) - na rozdíl od vlastních front-end aplikací (73).

3.5.1. Twine, Mywikibiz a DBpedia

Asi nejlepšími příklady tzv. *Real-world* sémantických aplikací jsou počiny typu TWINE,³ Mywikibiz⁴ a DBpedia.⁵ Druhé dvě aplikace jsou postaveny částečně na základech původního systému Wikipedia, každá z nich však používá poměrně odlišný přístup, jak při implementaci zdrojů dat, tak i při financování a samotné práce s RDF. TWINE je oproti tomu

³ Lze najít na webové adrese: www.twine.com

⁴ Lze najít na webové adrese: www.mywikibiz.com

⁵ Lze najít na webové adrese: www.dbpedia.org

aplikací poněkud jiného typu – jeho hlavním cílem je sice sdílení znalostí a naplnění informačních možností sémantického webu – to vše je ale založeno na principech uživatelského sdílení a provázanosti, tak jak je to běžné u jiných sociálních sítí (jako Facebook, Twitter, LinkedIn...atd).

První aplikací, která zde bude podrobněji popsána je projekt **TWINE** společnosti Radar Networks. Tato americká softwarová firma je, mimo jiné také díky svému řediteli a zakladateli (vizionářský podnikatel Nova Spivack), jedním z hlavních průkopníků vývoje sémantických aplikací pro běžnou uživatelskou základnu. TWINE je z uvedených aplikací sémantickému webu bezkonkurenčně nejbližší, navíc je dokonalým příkladem propojení sociálních sítí a možností RDF. Lidé stojící za tímto webovým nástrojem rádi prohlašují, že jde o první opravdovou mainstreamovou sémantickou aplikaci.

TWINE je, jak už bylo řečeno, především znalostní bází – ovšem svou informační sílu nezakládá na seskupování ohromného množství dat a jeho postupné sémantické analýze, nýbrž na intencionálním propojování uživatelů se zájmem o určitou oblast (či oblasti). Cílem TWINE není rozšiřování sociálních kontaktů, ale něco co by se dalo nejlépe popsat jako „knowledge networking“ (neboli budování své vlastní znalostní sítě – uživatelů) (30). Tato myšlenka by se dala sledovat až někam do hluboké historie vzniku WWW k prapůvodnímu ARPANETU, který používali vědci pro komunikaci a sdílení poznatků z výzkumu pomocí krátkých textových zpráv (předchůdců dnešních e-mailů). Není proto divu, že se tento, na webu se stále opakující nápad, znovu objevil i v konvergenci sémantických a sociálních sítí.

TWINE je opravdovou výjimkou v rámci sémantických aplikací - a reálně využívá pro práci s ontologiemi jazyk OWL. Kompletní aplikace je pak postavena na datovém modelu RDF s XML syntaxí. Práce se systémem TWINE je jednoduchá a je soustředěná kolem vyhledávaných témat. Tato mohou uživatele (podobně jako například ve Wikipedia) zakládat, doplňovat a všelijak upravovat. Základní hledání je založeno na klasickém full-text, ale sémantická síla TWINE leží teprve v samotných výsledcích tohoto vyhledávání. Uživatelé mohou brouzdat témata a hledat v nich souvislosti s dalšími informacemi a to ať už pomocí uživatelsky generovaných tagů nebo doporučeními samotného systému, vzniklých právě na základě RDF trojic a detekce URI.

Hlavní síla TWINE spočívá, jak už bylo zmíněno, v propojování lidí sdílejících podobné zájmy, aby si tak uživatelé mohli navzájem doplňovat znalosti. Toto je možné díky funkci ukládání témat. To by ještě nebylo nic tak zvláštního – vždyť to samé umožňuje

například síť Facebook, když se uživatel přihlásí na jakoukoliv „fanouškovskou“ stránku. Rozdíl je zde však v práci s daty uvnitř systému a z toho vyplývající relevancí nabízených spojení. Facebook udělá zkrátka pouze to, že ukáže všechny uživatele, kteří se na tuto stránku přihlásili. Oproti tomu TWINE (díky tomu, že „chápe“ smysl uživatelských dat) nabídne propojení s jedinci, kteří jsou nejen přihlášení ke stejnému tématu, ale mají podobné zájmy na více úrovních. Systém tak může porovnat a seřadit uživatele velice přesně pomocí jejich profilu a plného spektra sémantického grafu, který vznikne, když se spojí veškeré URI, na něž uživatel sémanticky odkazuje. Vizualizace tohoto sémantického grafu (na jehož představě je celý datový model RDF založen) je dokonce možné generovat pomocí externí aplikace vyvinuté speciálně pro TWINE. Uživatelé tak mohou vidět, jak vypadá jejich znalostní síť v grafické podobě. TWINE je jistě zajímavým počinem a dobrým příslibem do budoucnosti, jde o jisté potvrzení ve stylu: „sémantický web je zde a funguje“. Na druhou stranu jde o aplikaci, která stále vyžaduje velkou dávku lidského úsilí, aby se uživatel dostal k požadované informaci. Také využití přirozeného jazyka je zde spíše minimální. Opravdový přínos TWINE tkví tak především v oboustranných sémantických odkazech a z nich vzniklé vědomostní síti.

Mywikibiz je projektem skupiny lidí, kteří se podíleli na zakládání původní Wikipedie. Co se týče vzhledu, vkládání dat i funkčnosti je tento server vystavěn zcela právě na enginu originální Wikipedia. Je tu však jedná podstatná výjimka. Mywikibiz využívá upravenou verzi softwaru *mediawiki*, nazvanou trefně *semantic mediawiki*. Ta umožňuje uživatelům označit jakýkoliv příspěvek do této báze libovolným počtem klíčových slov ve formě sémantických tagů.⁶ To by samo o sobě nebylo ještě nic tak ohromujícího, ale to co činí Mywikibiz zcela výjimečným projektem je jeho synergie současného WWW a technologie RDF, při zachování naprosté kompatibility a využitelnosti pro koncové uživatele. Uživatelé totiž mohou vkládat sémantická data pomocí jednoduchého kódu bez jakékoliv znalosti RDF. Tyto data mohou sami dělit na *kategorie*, *vztahy* a *atributy*. Kód, který se pak do vlastního HTML stránky, vypadá například takto (77):

```
<br>
<div style="overflow:auto;height:1px;">
[[Keyword:=Heating oil]]
[[Keyword:=Fuel]]
```

⁶ Více viz www.mediawiki.com

```

[[Keyword:=Home heating]]
[[County_Name:=Somerset]]
[[County_Name:=Warren]]
[[County_Name:=Morris]]
[[City:=Readington]]
[[City:=Clinton]]
[[City:=Flemington]]
[[State_Code:=NJ]]
[[Has_Susidiary::Acme Excavation]]
[[Region_Located_In::North Jersey]]
</div>
<br>

```

Mywikibiz tedy pracuje na systému založeném na RDF a pro vkládání dat využívá XML syntax. Systém pak sám vytváří RDF trojice a ukládá je do sémantického repozitáře. Výsledná data jsou potom jak sémanticky validní a strojově čitelná, tak i navíc dodávají příspěvkům v této znalostní bázi velkou výhodu v rámci SEO (Search Engine Optimalization). Je tomu tak důvodu, že sémantické XML tagy jako například [[Keyword]] či [[City]] jsou již současnými vyhledávači plnohodnotně indexovány a pomocí nich získá crawler (indexovací robot) mnohem přesnější data o obsahu stránky než indexací běžného HTML.

Aby toho nebylo málo, vytvořili vývojáři Mywikibiz i zvláštní dotazovací jazyk označený ASK. Ten umožňuje registrovaným uživatelům klást i sémanticky konkrétní dotazy na bázi trojic. Dotaz v jazyku ASK může vypadat například takto (77):

```

<ask>
[[City:=*]]
[[State Code:=*|State]]
[[Zip:=*]]
[[Latitude:=>36.0]][[Latitude:=<40.0]][[Latitude:=*|Latitude]]
</ask>

```

Výsledkem takového dotazu je seznam všech měst, ve všech (amerických) státech (nezávisle na poštovním směrovacím čísle), které se nacházejí mezi zeměpisnou šířkou 36 až 40 stupňů včetně.

Webová služba Mywikibiz je dobrým příkladem, jak lze i za současné (pro sémantický web komplikované situace) využít datového modelu RDF k budování informačně bohatších a navíc komerčně (jak již napovídá název stránky a důraz na výsledky SEO) zajímavých WEB projektů.

Aplikací, jež je v současnosti ve spojitosti s reálným využitím sémantického webu asi nejvíce zmiňovaná, je bezesporu **DBpedia**. Je však potřeba upřesnit, že termín DBpedia pro označení aplikace není úplně vhodný – jedná se spíše o skutečnou znalostní bázi, zdroj dat. Jde o společný projekt dvou německých univerzit - Freie Universität Berlin a Universität Leipzig a sdružení OpenLink Software (72). Na rozdíl od Mywikibiz, jde o projekt zcela nekomerční a jeho hlavním účelem je veřejný výzkum a rozvoj v oblasti sémantických technologií. Zároveň výzkumný tým, stojící za Dbpedia, považuje svou znalostní bázi za první zárodek vznikajícího *Web of Data* – skutečného sémantického webu, tak jak jej chápe do budoucna především W3C. Stavební rámec DBpedia se stal zároveň předpokladem k vzniku celé řady open source aplikací s ní spojených (většinou se jedná o různé méně či více funkční dotazovací klienty pro možnost hledání v tomto systému). Informační síla DBpedia spočívá v extrakci dat z neustále rozvíjející se databáze klasické Wikipedia.org a jejich přeukládání ve formě RDF trojic. To skýtá celou škálu potenciálních využití včetně možnosti implementace výsledků dotazů na DBpedia pomocí jazyka SPARQL (například ve formě infoboxu) do běžné webové stránky pomocí javascriptu na straně klienta nebo PHP příkazu na serveru (67, s. 730). Takový dotaz je pak na stránce neustále (automaticky) doplňován, tak jak jsou doplňována data uživateli originální Wikipedia.

DBpedia se aktuálně skládá ze 103 milionů RDF trojic a v kombinaci s interlinky s dalšími otevřenými datovými bázemi dává dohromady neuvěřitelné 2 miliardy RDF trojic a je tak naprosto bezkonkurenčně nejrozsáhlejší znalostní bázi na světě (67, s. 723). Systém interlinků spojuje DBpedia se zdroji dat, které se zaměřují většinou na budování ontologií v konkrétní oblasti, jako například zeměpisná data, hudba, databáze knih či informace o světových státech (mezi tyto datové soubory patří projekty jako: Geonames (4), MusicBrainz (78), Project Gutenberg (68), Freebase (79), a mnoho dalších). Samotná

DBpedia tak nyní poskytuje informace o 1,95 miliónů „věcí“ včetně 80000 osob, 70000 míst, 35000 hudebních alb, 12000 filmů, 657000 odkazů na obrazové materiály, 160000 odkazů na relevantní webové stránky, to vše v sémanticky obohaceném datovém hávu. Navíc lze tento výčet doplnit ještě o 180000 externích odkazů na jiné RDF datové soubory a 207000 Wikipedia kategorií (67, s. 725).

Velmi zajímavý je způsob, jakým DBpedia extrahuje data z Wikipedia. Pomocí databázových úložišť jsou využívány dvě základní metody k extrahování těchto dat: 1) mapování vztahů mezi datovými vstupy, jež jsou již uloženy v relační databázi Wikipedie a jejich koverze na RDF formát, 2) získávání dodatečných informací přímo z textů článků a infoboxů. Stylování Wikipedia příspěvků a infoboxů využívá jednotnou šablonu (tzv. template), extrahovací algoritmus DBpedia vyhledává tyto šablony a pomocí porovnávání vzorů vybírá podstatná data a za pomoci parseru je převádí do RDF trojic. Pokud jde o odkazy uložené v mediawiki šablonách – ty jsou transformovány do příslušných URI (67, s. 725).

Vyhledávání v DBpedia je samostatnou kapitolou. Jak již bylo napsáno výše, existuje celá řada interních i externích klientů určených k tomuto účelu. Zároveň je však nutno dodat, že řada z nich teprve probíhá vývojem a nemusí vždy 100% fungovat. Je tu také možnost celou bázi DBpedia (nebo její části) stáhnout a pracovat s ní pomocí některého z offline klientů (což je ovšem pro běžného uživatele zcela nepraktická varianta). Aktuálně asi nejpoužitelnější dotazovací aplikace, poskytující uživateli alespoň základní grafické rozhraní pro lepší orientaci je SNORQL – postavená jak jinak na dotazovacím jazyku SPARQL (71). Pochopitelně je třeba znát alespoň základy SPARQL syntaxe, aby bylo možné v DBpedia efektivně vyhledávat. Na rozdíl od jazyka ASK na Mywikibiz tady totiž autoři nezaobalili SPARQL do nějaké uživatelsky jednodušší formy a nechávají ho v podstatě v syrové podobě. Dotaz v SNORQL pak tedy vypadá typicky takto:

```
PREFIX dbo: <http://dbpedia.org/ontology/>  
SELECT ?name ?birth ?death ?person  
WHERE {?person dbpedia2:birthPlace <http://dbpedia.org/resource/Prague> .  
?person dbo:birthdate ?birth . ?person foaf:name ?name . ?person  
dbo:deathdate ?death  
FILTER (?birth < "1900-01-01"^^xsd:date) . }
```


Tento dotaz má za následek genezi seznamu všech osob narozených v Praze před 1.1. 1900 nacházejících se v záznamech DBpedia respektive Wikipedia.

Jiným způsobem jak získávat data z DBpedia je brouzdání v URI na základě diferenciací *oboustranně linkovaných dat*. Toto sémantické „serfování“ mezi URI lze provádět pomocí jedno z dostupných sémantických prohlížečů, ty z vložené URI vygenerují veškeré linky na jiné URI, které v sobě toto obsahuje. Ale i v případě těchto specializovaných prohlížečů existuje problém nefunkčnosti sémantických aplikací třetích stran – jedná se často o experimentální projekty, které jsou buďto ve stádiu dlouhodobého vývoje, případně jsou už naopak takřikajíc ponechány svému osudu bez nutné údržby.

Přes všechny nedostatky sémantických aplikací (včetně těch spojených s DBpedia) zde existuje jedna, která naskýtá skutečný pohled do budoucího paradigmatu používání WWW. Jde o extenzi **DBpedia Mobile** primárně určenou pro zařízení iPhone a přístroje s operačním systémem Windows Mobile (na variantách pro další mobilní zařízení a operační systémy se pracuje) (15, s. 344). Tato speciální aplikace totiž přichází z tzv. mash-up (spojením dvou různorodých aplikací) systému Google Maps (postaveného na XML) a DBpedia. Aplikace je napojena na zařízení pro zachycování družicových zeměpisných GPS dat a na základě jejich identifikace nabízí uživateli informace o URI – reprezentaci „věcí“ v jeho přímém okolí (60). Je to umožněno zejména díky základnímu principu sémantického webu – strojové zpracovatelnosti dat. Zeměpisná data jsou zanesena v ontologii DBpedia v RDF formátu, a proto je může aplikace vybavená parserem správně analyzovat a snadno vygenerovat relevantní výsledky na základě GPS lokace udávané mobilním zařízením. Kromě toho je DBpedia Mobile otevřena i dalším datovým úložištím a v současné době už je možné například získat informace o hudbě a hudebních skupinách spojených s polohou, v níž se uživatel nachází pomocí báze MusicBrainz. Totéž platí o knihách a Projektu Gutenberg (60). Tato utilita předznamenává cestu, kterou se bude WEB ubírat – směrem k vysoce relevantním, sémantickým, individualizovaným a automaticky generovaným datům a k mobilitě. Navíc DBpedia mobile naznačuje ještě jednu zatím tak trochu skrytou vlastnost sémantického webu – a sice aspekt tzv. *Ubiquitous Web* (představa sítě, která bude reálně všudypřítomná a bude spojovat *datasféru* s fyzickým světem), podpořené myšlenkou, že URI mohou (logicky) reprezentovat i ne-virtuální objekty reálného světa.

3.5.2. Sociální polo-sémantické aplikace

Jedním z hodně viditelných projektů, stojících někde na hraně sémantického webu a WEB 2.0 je **Freebase.org** softwarového sdružení Metaweb. I v tomto případě se jedná o svého druhu znalostní bázi (jako v podstatě u všech reálně fungujících současných sémantických aplikací), jejíž hlavním cílem je přiblížit informační potenciál webu jeho skutečným možnostem. Jak již název napovídá, Freebase se orientuje především na širokou veřejnost a jejím přispěvatelem se může stát doslova kdokoliv. Veškeré využívání dat z Freebase je zcela zdarma a je dokonce možné data přímo vkládat do vlastní webové prezentace pomocí speciálního kódu. Freebase obsahuje velké množství informací především z populárních oblastí, jako je hudba, film a umění, ale zároveň je zde možné najít informace například z politiky a dalších odvětví. Obsah Freebase závisí hlavně na rozsahu informovanosti WEB komunity a široké veřejnosti o její existenci a na jejím zpopularizování. Hlavní otázkou zůstává, k čemu je taková databáze vůbec dobrá, existuje-li zde už tolik rozšířená Wikipedia. Odpověď tkví v způsobu, jakým se na Freebase s daty pracuje a jak se uživatelům zobrazují. Zatímco Wikipedia strukturuje informace do běžných článků a infoboxů, Freebase zobrazuje data už předem strukturovaná v informačních blocích (80). Poskytuje tedy mnohem rychlejší přístup k hledaným datům. To je možné zejména proto, že Freebase je z části postavená právě na sémantických technologiích.

Z hlediska sémantického webu jako takového je ovšem Freebase pouze hybridním počinem, který využívá jistých podobných principů, ale rezignuje na některé z těch základních – jako například na možnost dotazování báze pomocí SPARQL. Místo toho používá svého dotazovacího jazyka MQL (Metaweb Query Language). Jakousi náplastí na tento nedostatek je podpora vyhledávání oboustranně linkovaných dat – podobně, jako to lze dělat v některých aplikacích pro DBpedia. Podstatným počinem společnosti Metaweb, která Freebase stvořila, bylo oproštění se od klasického databázového systému postaveného na tabulkách a zavedení RDF úložiště místo něj. API (Application Programming Interface) pro Freebase je několik a všechny jsou open source původu. Ne všechny jsou však sémantického původu, ale RDF je vždy samozřejmou možností, kterou si lze při editování obsahu vybrat. Uživatelé navíc mohou editovat nejen sémantické odkazy na data navzájem, ale i samotné ontologie (25, s. 25). Ty se v rámci Freebase nazývají *types* a vzhledem k tomu, že jsou volně otevřené, skýtají obrovský sémantický potenciál (15, s. 348). Hlavní výhoda

podobné znalostní báze je evidentní zejména při vyhledávání informací z většího množství zdrojů. Jestliže chci získat přesná data z několika článků či tabulek týkající se jednoho tématu, znamená to v případě běžného webu poměrně dost manuální práce, jelikož data zde nejsou přesně strukturována a využít tak běžně dostupné analytické nástroje není možné. Pokud jsou však tato data uložena i ve formě RDF trojic a sémantických linků (za předpokladu, že této možnosti uživatelé při jejich vkládání využili) je možné použít zvláštní data-mining crawly - ty poskytnou ze dat zdrojových, plnohodnotná data sémantická (10, s. 904). Aplikace typu Freebase jsou jistě pro vývoj sémantického webu dobrou předzvěstí do budoucna a ukazují, že jeho principy veřejně dostupných znalostí s předsevzetími sociálního hnutí (kolem WEB 2.0) mají mnoho společného.

Ve výčtu aplikací spojených se sémantickým webem rozhodně nesmí chybět ani projekt **Powerset** stejnojmenné softwarové společnosti. Jde o jeden z prvních (téměř) fungujících vyhledávačů nové generace – vyhledávačů umožňující klást dotazy v přirozeném jazyce. Výsledky dotazování pak čerpá Powerset z informačních zdrojů Wikipedia a Freebase. Přístup Powerset je však klasickou ukázkou hlavní překážky, která brání v rozvoji sémantického webu – chybějících standardů.

Powerset totiž pro vyhledávání a především indexování zdrojů využívá zcela odlišný princip, než bychom našli u všech výše zmíněných aplikací. Při indexování roboti hledají opakující se slova, kontext, smysl slov a různé kategorie – ne snad, že by se jednalo o roboty disponující skutečnou umělou inteligencí, nýbrž k tomu využívají zvláštní lingvistickou ontologii naplněnou biliony podobných dat. Jde tedy de facto o podobný „úrok stranou“ jako třeba v případě WORDNET. Získaná data se pak převádějí do podoby metadat, která jsou porovnávána s dotazy uživatelů. Co se týče vlastního dotazu, k jeho analýze se využívá metod NLP (Natural Language Processing) a porovnání vyhledávacího (neustále se učícího) algoritmu a parseru s kontextem zaindexovaných stránek (45). Powerset dokonce v jednu chvíli platil uživatelům dva (americké) centy za evaluaci sémantické relevance každého výsledku vyhledávání. Tím se podařilo získat obrovské množství lingvistických dat využitelných zpětně pro NLP.⁷ Z popsaného je tedy vidět, že Powerset je jakýmsi sémantickým vyhledávačem pro nesémantický web. Tato technologie poměrně kvalitně funguje a je skutečně schopna generovat velice relevantní výsledky (někdy by se dalo říct přímo skutečné odpovědi) na nekomplikované otázky typu: „Kdo je nejslavnější fotbalový

⁷ K tomu více viz NEWMAN, Aaron, THOMAS, Jeremy. *Enterprise 2.0. Implementation*, McGraw-Hill Professional, 2008. 406 s. ISBN-13: 9780071591607

hráč historie?“ (dotazy musí být pochopitelně kladeny v angličtině, jelikož pro jiné jazyky není Powerset uzpůsoben).

Kromě běžných výsledků vyhledávání nabízí Powerset uživatelům také tzv. *Factz* tedy lingvistické trojice složených ze subjektu, relace (slovesa) a objektu, zobrazované na základě statické četnosti těchto spojení v indexovaných datech. Platforma Powerset (i s celou firmou) byla v roce 2008 odkoupena jako jedna z mnoha akvizic gigantem Microsoft, který některé z jejich technologií využil pro své vyhledávače Bing a Live Search (i přesto tyto dvě aplikace stojí i nadále především na klasickém vyhledávání klíčových slov). Powerset je velice ambiciózním projektem, který však bohužel pouze ukazuje, že takto zvolená cesta má se skutečným sémantickým webem - a hlavně cílem reálného porozumění smyslu webových dat počítačem - jen pramálo společného.

Další aplikace, mající alespoň něco společného se sémantickým webem má smysl zmínit již jen okrajově. Tak například skupina Adaptive Blue stojí za utilitou **SmartLinks**. Ta umožňuje provozovatelům webových stránek vkládat ke svým odkazům dodatečné sémantické aplikace (tyto se mohou zobrazovat v podobě pop-up oken). Dalším poměrně zajímavým počinem je vyhledávač **Hakia**. I ten se pokouší o zprostředkování vyhledávání v přirozeném jazyce a soustředí se přitom na celé věty. K vlastnímu vyhledávání využívá dva principy – speciální infrastrukturu QDEX, která analyzuje lingvistické významy na úrovni vět a SemanticRank Algorithm, ten porovnává dotazy s relevantními ontologiemi a vytváří odkazy mezi jednotlivými daty.⁸ Přímá zkušenost s vyhledáváním na Hakia (která je stále v beta fázi testování) je však poněkud zklamáním, když uživatel zjistí, že většina výsledků je stále založena na porovnávání klíčových slov.

Mezi další sémantické projekty se řadí například aplikace **Spock** nebo plugin pro prohlížeč Firefox pojmenovaný **Gnosis** společnosti ClearForest. Druhá jmenovaná existuje i jako zvláštní program ke stažení, který se dá využívat nezávisle na prohlížeči. V principu jde „pouze“ o další NLP systém, který čerpá z obří databáze společností, technologií, organizací, produktů a osob. Zajímavostí je, že ClearForest byla odkoupena společností Thomson-Reuters (největší tiskovou agenturou na světě), která chce její produkty na poli sémantického webu využít především pro nové zpravodajské on-line produkty pro profesionály v oblasti médií. To aplikace **Spock** se specializuje na vyhledávání (především) osob, které kategorizuje na základě poměrně velkého počtu atributů. Specialitou Spock je

⁸

Více o Hakia také na: www.readwriteweb.com/archives/10_semantic_apps_to_watch.php

učící se systém, který přetváří časté fráze v dotazech do podob tagů - ty mohou pak uživatele upravovat a přiřazovat jim relevanci pomocí systému zpětné vazby.⁹

Ve výčtu aplikací, které by se daly označit jako sémantické nebo se tomuto označení alespoň blíží, by se dalo pokračovat ještě poměrně dlouhou dobu, ale těch opravdu významných a fungujících je ve skutečnosti velmi málo. Odlišným případem, který je třeba zmínit, je znalostí vyhledávač **Wolfram Alpha**. Ten byl s velkou mediální pozorností spuštěn 15.5.2009 a poměrně mylně byl okamžitě spojen se sémantickým webem. Tato medializace sice zčásti napomohla rozšíření povědomí o existenci sémantických technologií pro širokou (i odbornou) veřejnost, ale zároveň ukázala fatální nepochopení principů fungování sémantického webu, ze strany žurnalistické obce. To je možné pozorovat i na výstupech v českém tisku.¹⁰

Jedním z hlavních aspektů Wolfram Alpha je totiž fakt, že i když je jeho využití principům sémantického webu podobné, není postaveno na sémantických technologiích a nemá s nimi téměř nic společného. Wolfram Alpha je především vyhledávačem znalostním, jehož hlavním účelem je asociace vyhledávacích dotazů s matematickými daty. Nejedná se tedy o internetový vyhledávač v pravém slova smyslu, ale spíše neustále se rozšiřující databázi doplňovanou postupně o data dostupná na webu. Pochopitelně, že i k zobrazování relevantních matematických dat k textovým dotazům je nutné jisté sémantické podhoubí, pravda je ovšem taková, že Stephen Wolfram (vynálezce a tvůrce Wolfram Alpha) svou technologii přísně chrání a příliš se o ní neví. Jediné, co se o ní dá s jistotou tvrdit je její postavení na teoretických základech NKS (New Kind of Science) (2, s. 3) – souborem předpokladů k novému přístupu k vědě (jejichž je Wolfram, jak jinak, autorem) založeném na metodologii zkoumání tzv. „jednoduchých programů“ (simple programs), které jsou schopny překvapivě rozsáhlého množství variací složitého (a nepředvídatelného) chování (1).

Celkově by se o tzv. „real-world“ sémantických aplikacích (tedy takových, které už se začínají objevovat ve verzích ke skutečnému použití) dá říct, že, jak je naznačeno výše, jich existuje (na obrovský potenciál sémantického webu) opravdu jen velmi malé množství. Zároveň je ale třeba dodat, že toto odvětví softwaroví giganti ani open source komunita vůbec neopomíjejí. Ba naopak sémantické aplikace jsou pod drobnohledem celé řady programátorů a světové vývojářské firmy v nich vidí obrovský potenciál. Příklady můžeme

⁹ Více o Spock také na: www.readwriteweb.com/archives/10_semantic_apps_to_watch.php

¹⁰ Viz scestný článek Ondřeje Pohla - Glosa: Chiméra jménem sémantický web uveřejněný na serveru www.lidovky.cz

vidět i ve snaze velkých světových firem aplikovat (v různých formách, zatím většinou především pomocí XML u enterprise sféry, zatímco open source komunita staví spíše na rámci JSON – JavaScript Object Notion) sémantické principy a automatizaci do svých komerčních aplikací, jako například: e-speak od Hewlett-Packard – systému k popisu, registraci a dynamickému rozpoznávání e-sluzeb, programovací rámec .NET a soubor nástrojů BizTalk společnosti Microsoft, nebo systém Open Network vyvíjený korporací Sun. Opravdu vážně berou datový model RDF progresivní vývojářské nadnárodní společnosti Oracle a IBM. Ty na něm vystavěly své vlastní vývojářské prostředí pro programování webových služeb - Dynamic Services Framework v případě Oracle a Application Framework u IBM (27).

Obzvláště firma Oracle se snaží komunitu svých vývojářů podporovat k další práci a vzdělávání se v této oblasti. Mimo jiné i provozuje speciální on-line informační server - Semantic Technologies Center (75). Zde si může kdokoliv volně stáhnout Oracle Database Semantic Store (pro operační systém Linux), kterou společnost využívá pro tvorbu specializovaných aplikací – v současné době většinou intranetových RDF datových znalostních úložišť. V poslední době Oracle realizoval podobné projekty například pro různé farmaceutické společnosti a několik amerických univerzit (74). O tom, že hlavní hráči na poli softwaru berou sémantický web vážně, svědčí rozhodně i akvizice firmy Powerset Microsoftem, o níž už byla řeč. I tak by se však od těchto velkých společností ve vývoji „web of data“ dalo očekávat více – bohužel tomu brání především ekonomické důvody a problémy s neochotou společné integrace technologií a vývojářských prostředí – což už bylo podrobně popsáno v předešlých kapitolách. Přesto je ovšem možné vidět vývojářské společnosti (většinou menšího rozsahu), co už se do komerční produkce utilit, služeb a jednoduchých sémantických aplikací s vervou pustily. Příkladem takového podniku může být německá firma Semsol, mezi jejíž produkty patří například RDF adresář, sémantický kalendář, mikroblogovací systém s možností sémantického tagování, SPARQL grafický web-dotazovací rámec či vývojové prostředí pro nastavbu RDF a SPARQL kódu v PHP prostředí (82). I když se jedná většinou o jednoduché beta mini-aplikace, jde o příslib do budoucna a důkaz toho, že vývojářská komunita kolem sémantického webu žije a pomalu začíná reálně pronikat do komerční sféry. Dá se očekávat, že tento trend bude jistě pokračovat, ale rozhodně to, z již opakovaně zmíněných důvodů, nebude tak rychle, jak třeba W3C ještě před pár lety očekávalo.

4. Filosofické aspekty sémantického webu

Koncept sémantického webu zásadně mění pohled na světovou síť a možnosti získávání informací vůbec. Současné výsledky tohoto nového přístupu k práci s daty a možnostmi jejich ukládání a strukturaci sebou nesou konotace nejen technické, ekonomické a informační, ale také filosofické a sociologické, jelikož jejich aplikace a další vývoj mohou významně ovlivnit další vývoj naší globalizované společnosti.

4.1. Proměna síťového paradigmatu díky sémantickému webu

V oboru fyziky – tedy nauce o povaze a příčinách přírodních jevů se popisují tři velké změny paradigmatu. Nejprve se od naivních kosmogonií a kosmologií opírající se o nadpřirozené jevy a antropomorfní bohy přešlo na Aristotelovskou fyziku, která byla významně překonána až fyzikou Newtonskou o 2000 let později, avšak již o necelých 200 let později přišla další změna paradigmatu v podobě Einsteinovy speciální relativity a pozdějšímu vzniku kvantové mechaniky. Paradigma se vyznačuje určitým úhlem pohledu na danou problematiku, v jehož módu se pozorovatel, vědec či teoretik, v podstatě chtě nechtě, pohybuje (33, s. 24).

Stejně tak moderní informační věk, rozvoj výpočetních technologií a následný vznik počítačových sítí, především podle sociologa Manuela Castellse, vyžaduje zcela nový úhel pohledu sociálních věd (stejně tak jako vědy počítačové) na současný svět – změnu paradigmatu – vznik *paradigmatu síťového*. Castells při svých úvahách nad WWW sice

sémantický web v potaz nebral, ale ve skutečnosti je to právě tento soubor technologií, který nové paradigma posouvá ještě o kousek dál.

Základní definicí současné společnosti je (dle Castellse) její rozdělení na dvě bipolární sféry: „Sít“ a „Já“ (v podstatě metafyzické - „Já o sobě“) (52, s. 494). „Sít“ působí jako ustanovující sféra ekonomických globalistických tendencí a mód sociální komunikace mizícího času a prostoru – měnícího se na binární formu reprezentace – vše buďto „je tady a teď“ nebo to není – 1/0. „Já“ je místem, v němž redefinujeme a znovu-nabýváme identity, které reagují na měnící se strukturální uskupení sociálních a ekonomických aktivit v nových dynamicky se rozvíjejících sítích (28). Do této zásadní polarity ovšem vstupuje nový prvek - „smysl“. Jde o prvek, s nímž původní síť, jako „pouhé“ komunikační médium (i se svými převratnými možnostmi) nepočítala. Jedná se zde o snížení epistemologických nepřesností, informačního šumu a informačních ztrát, ke kterým nutně dochází při každé mediaci. Naše informační společnost se může v případě masového rozvoje sémantického webu dostat do situace, kdy už lidský uživatel nebude muset provádět nejen žádnou sémantickou, ale ani hermeneutickou analýzu (42, s. 88). Místo obyčejné komunikační sítě zde budeme mít síť znalostí (a vědomostí). V ideálním případě bude síť plně automatizována a personalizována, softwarové nástroje budou třídit potřebná data a poskytovat nám přesné a především relevantní informační celky. V takovém případě je tedy dost dobře možné, že nastane jakási syntéza obou protipólu výše uvedené dialektické dvojice – Sít' a „Já“.

Zmíněný vývoj síťového paradigmatu dává sémantickému webu zcela nový smysl a společně s proměnou sociálního sebeurčení se uživatelů v rámci sítě přináší odlišný pohled na to, co člověk – v prostředí webu třetí generace a jeho následovníku – je, nebo spíš může být. Ve své synergii „Sít“ a „Já“ proměňují uživatele – člověka na lidský terminál (5, s. 134). Kromě hardwarových uzlů by tak nový WEB disponoval systémem datových uzlů tvořených uživateli – navzájem doplňujícími existující ontologie, či vytvářejícími své vlastní a sdílejícími vědomosti a data s dalšími uživateli navzájem na základě preferencí, které už jsou ovšem automaticky určovány sémantickým webem. Dalo by se tedy říct, že díky svým novým sémantickým možnostem WEB zapříčiní, že *infosféra* pohltní biosféru (což je kupříkladu jedna z nejhorších obav francouzského myslitele Paula Virilia) (5, s. 9-15). Tuto zdánlivě vědecko-fantastickou představu (byť Virilio měl na mysli především proměnu lidských hodnot) podporují i některé další fakty přímo spojené se sémantickým webem. Prvně URI, jako základní označení cesty ke zdroji dat může (což si málokdo uvědomuje)

odkazovat i na informační zdroje mimo WEB – tedy mimo vlastní infosféru definovanou počítačovými sítěmi a hardwarem. Kód s využitím této „fyzické“ URI vypadá zhruba takto:

```
<p xmlns:foaf="http://xmlns.com/foaf/1.0/"
  xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#"
  about="_:WashingtonMonument"
  property="geo:lat_long" content="38.8895563,-77.0352546">
  During our trip we visited the
  <a rel="foaf:homepage" href="http://www.nps.gov/wamo/">
  Washington Monument</a>
</p>
```

Fyzický objekt ve světě je tedy určen například jeho přesnou geografickou polohou (16).

4.1.1 Ubiquitous WEB

Se sémantickým webem se pojí také představa takzvaného „všudypřítomného webu“ (mezinárodně používaný výraz zní: *ubiquitous web* nebo také *ubiquitous computing*). Tento termín už také definovalo W3C a extenzi webu popisuje jako interoperabilitu mezi webovými prohlížeči, službami, které v nich běží a zařízeními běžného užití v našem přirozeném prostředí. Technologicky by se jednalo v podstatě o zjednodušení, kdy by byl pro konfiguraci jednotlivých zařízení a obsahu sítě použit stejný programovací rámec založený na značkovacích jazycích. Technologie sémantického webu by v tomto případě byly mimo jiné využity pro popis toho, co které zařízení umí (51). V blízké budoucnosti tak budou uživatelé moci například nechat na počítači (respektive síti počítačů a zařízení) veškerou agendu běžných povinností. Softwarový agent automaticky zajistí obchodní schůzku na vhodné datum při synchronizaci s kalendářem všech zúčastněných, mezitím zadá do navigace auta GPS polohu místa, kam se má uživatel dostavit, v součinnosti s palubním počítačem posléze vypočítá průměrnou spotřebu a upozorní řidiče na případnou potřebu doplnit pohonné hmoty (samozřejmě mu i doporučí nejvýhodnější čerpací stanici na plánované trase). Tato velice fádňá ukázka možností ubiquitous web jen dokládá reálné propojení fyzického prostoru s infosférou.

4.1.2. Simulace a realita

Výše zmíněná - na první pohled jednoduchá - zjištění, v nichž je možné pozorovat vývojové předpoklady WWW, však kladou nové světlo na myšlenky, kterými si někteří postmoderní myslitelé zabývali už řadu let nazpět. Francouzský sociolog Jean Baudrillard například už dávno varoval před lavinou znaků – simulace, která zcela zakrývá naše přirozené životní prostředí – realitu. Dle jeho názoru právě simulace a hyperrealita (tedy akcelerované prostředí moderních masových a síťových médií) mažou veškeré reference ke všem lidským cílům a destruuji tak základní rozdělení hodnotového systému jakou jsou juxtapozice „pravdy“ a „lži“ či „dobra“ a „zla“ (64, s. 43) A skutečně – primitivní binární systém rozdělující svět na variantu A (1) a variantu B (0) sice pravdivou a nepravdivou hodnotu rozezná velice snadno ovšem pouze ve strojovém – nikoli psychologickém slova smyslu (a nic na tom nemění ani rozvíjející se výzkum v oblasti kvantových počítačů, který by „pouze“ binární svět rozšířil o další vrstvy proměnných). To, že současné stroje nemají žádný způsob, jak by rozpoznaly lidský hodnotový systém, pomocí něhož by odhadovaly intence lidského chování, je celkem jasné – a nic na tom nemění ani představa softwarových agentů a inteligentního webu budoucnosti. Zároveň je ale pravda, že pronikání síťového paradigmatu do reálného světa kolem nás se děje přímo před našima očima a má fatální důsledky na změny lidského chování – což šlo (a stále) lze pozorovat u tak relativně – primitivního posunu v technologiích jako je emergence WEB 2.0.

4.1.3. Filosofové proti sémantickému webu

I přes uvedené někteří filosofové zabývající se počítačovou vědou a jejím dopadem na člověka mají o reálných možnostech sémantického webu výrazné pochybnosti. Tak například Luciano Floridi - předseda International Association of Science and Computing (a člen Oxford a Heartfordshire akademické obce) kritizuje principy sémantického webu už od jeho základů včetně toho, že obviňuje autora této myšlenky – Tima Bernerse Leeho z matení veřejností. Tvrdí, že slova jako „sémantika“, „mysl“, „znalost“ a „inteligentní web“ jsou v tomto případě zcela metaforická nebo nesprávně použita – měla by být údajně raději nahrazena pojmy jako „integrace“, „agregace“, či „sdílení heterogenních dat, která jsou

ovšem velice konkrétně shlukována kolem předem daných témat“. Celkově je sémantický web jako koncept dle Floridiho odsouzen k zániku, jelikož abstrahovat masové rozšíření sémantických technologií napříč WWW pouze z příkladu několika specializovaných užití v konkrétních oborech je prý nelogické. Navíc, dle něj, nedochází k žádné zásadní změně, protože ontologie i samotný kód je prvotně produkován člověkem a koncovým uživatelem je opět člověk, čili „strojová zpracovatelnost dat“ je jen jakýmsi prázdným pojmem (43).

Floridi je ve svých kritikách na adresu sémantického webu až překvapivě tvrdý a v podstatě předvádí nepochopení cílů celé této myšlenky. Lidé kolem sémantického webu určitě neměli nikdy za cíl vytvořit na WWW jakousi umělou inteligenci, tak jak by si jí představovali například představitelé transhumanistů v čele s Vernorem Vingem. Ten předpokládá, že v budoucnu může dojít k bodu známému v počítačové vědě jako *singularita* – kdy si inteligentní systémy uvědomí sami sebe a začnou produkovat další a další mnohem inteligentnější umělé inteligence (6). Ve spojitosti se sémantickým webem a síťovým paradigmatem (či síťovým schizmatem - v případě prapodivného dvojvládní mezi „Sít“ a „Já“), tak jak zde bylo popsáno, je to zajímavé především proto, že za jednu z variant pro příchod singularity považuje Vinge „probuzení se do vědomí počítačových sítí – i s jejich uživateli“ (6).

Jak je vidět, i tak čistě technologická záležitost jako je vývoj sémantického webu a jeho technologií v sobě nese celou řadu filosofických a sociologických otázek. Mnoho z nich zapadá pouze do módu, v němž se postmoderní i současní myslitelé pohybují již od vzniku prvních masových a síťových médií. Na většinu lze, vzhledem k účelovosti (a prozaičnosti) celé problematiky sémantického webu stojící na poměrně jednoduchých principech, relativně snadno odpovědět. Jak bylo ovšem vidět na uvedených příkladech, rozhodně tyto otázky nelze podceňovat, stejně jako potenciální dopady sémantického webu na proměnu a fungování lidské society a forem lidského chování. Vždyť podobných pravidel, na nichž stojí sémantický web, se dožadoval už například Leibniz se svou představou stroje na rozpoznávání pravdivosti – Calculus Rationator (později těchto základů využil i Norbert Wiener - v počátcích kybernetiky (3, s. 197-219). Automatizace procesů a získávání implicitních - relevantních a pravdivých informací a jejich zakomponování do uživatelských interakcí (ať už v rámci virtuálního či reálného prostoru) tedy rozhodně nejsou nové myšlenky a již v sobě nesou určitou filosofickou tradici.

5. Závěr

V dynamickém prostředí dnešního WWW (potažmo celého Internetu) můžeme pozorovat nepřehledné množství trendů, hnutí, aplikací, emergentních technologií a téměř nekonečné množství dat. Nové světlo vnesl na síť WEB 2.0 a všechny jeho konotace a ať už je hodnocen jakkoliv, je evidentní, že WEB nejen zpopularizoval a přiblížil běžné veřejnosti, ale (byť mimoděk) položil do počítačových systému základy přirozené sémantiky v podobě tagování a folksonomií. Výhody sémantického webu a jeho široké možnosti využití byly v této práci popsány poměrně podrobně. Nezbývá se tedy než ptát: Je sémantický web opravdu dalším vývojovým stupněm v elektronické evoluci, nebo se jedná jen o slepou uličku? Jak už bylo naznačeno v kapitole o WEB 3.0 – nelze říct, že právě sémantický web sám o sobě a ani ve spojitosti se sociálními sítěmi toto označení snese. Zároveň by byla zcela mylná představa, že sémantický web je něčím, co má stávající web nahradit. Jde spíše o extenzi – nebo ještě lépe naplnění skutečného potenciálu, který WWW má.

Některé technologie sémantického webu jsou stále ve stádiu prvotního vývoje (například inteligentní agenti) a také jeden z jeho základních cílů – vyhledávání ve skutečně přirozeném jazyce je na míle vzdáleno od praxe, i tak je však jeho technologické podhoubí postaveno na velmi silných vývojářských a teoretických základech, jejichž demonstraci v podobě celé řady funkčních aplikací a sémantických úložišť již můžeme běžně pozorovat. Pro další rozvoj sémantického webu a naplnění cíle Tima Bernerse-Leeho a W3C – tedy vytvoření web of knowledge je klíčové sjednocení využívaných technologií – především v případě slovníků a syntaxí a dotazovacích jazyků a větší součinnost open source komunity a velkých korporací. Samozřejmostí je i hledání lepších modelů ekonomické návratnosti sémantických technologií, které by motivovaly především velké společnosti k vyčlenění většího rozpočtu na jejich vývoj.

Na nedostatek politické vůle k podpoře projektu sémantického webu, reprezentovaného kupříkladu začleněním této podpory do cílů již šestého Operačního programu EU (81), si W3C nemůže stěžovat. V čem však tato organizace zcela zaostává, je větší propagace výhod a celkové myšlenky kolem proměny WWW na dynamickou znalostní bázi nejen mezi širokou, ale dokonce i odbornou veřejností! Neinformovaná obec uživatelů ať už na laické, mediální, akademické ale i vývojářské úrovni se tak často k celé problematice staví velmi skepticky a pokud tento fakt W3C nezmění, je reálná naděje na rozšíření

sémantického webu jen mizivá. S podivem proto zůstává i nedávné spojení W3C s organizací WHATWG a jejich společná příprava standardu HTML5, jehož, jak se již ví, RDF není součástí a je na místo toho nahrazeno (co do funkčnosti a nasazení) velmi omezenými nástroji v podobě mikroformátů.

I přes veškeré problémy s nasazením sémantického webu je jasné, že budoucnost WWW jako takového do sebe počítačovou čitelnost dat, jejich kategorizaci a „znalost“ rozhodně zahrnuje. WEB 3.0 a další geneze WWW rozhodně budou nějakým způsobem ovlivněny sémantickými technologiemi, které se stanou součástí dalších specifikací HTML a jazyků, co přijdou po něm. I běžní uživatelé se (jak můžeme vidět na oznámení webového OS Google Chrome) v nějaké formě v budoucnosti dočkají webu postaveném na widgetech, inteligentních softwarových agentech, znalostních bázích, vyhledávání v přirozeném jazyce a logického „myšlení“ počítačových sítí. Zda to bude za pomoci RDF a současných syntaxí a slovníků podporovaných W3C zůstává otázkou. I přes horečnou aktivitu některých skalních zastánců sémantického webu a komunity vývojářů například kolem projektu DBpedia, zůstává totiž stále celý tento ambiciózní projekt tak trochu přešlapovat na místě.

Seznam literatury

Citovaná literatura

1. Wolfram Blog Team. *Stephen Wolfram on the quest for computable knowledge* [online]. June 29, 2009 [cit. 2009-08-06]. Dostupný z WWW: <<http://blog.wolfram.com/2009/06/29/stephen-wolfram-on-the-quest-for-computable-knowledge/>>.
2. WOLFRAM, Stephen. *A new kind of science*. Champaign (IL) : Wolfram Media, 2002. 1197 s. Dostupný též z WWW: <<http://www.wolframscience.com/nksonline/toc.html>>. ISBN 1-57955-008-8.
3. WIENER, Norbert. Time, communication, and the nervous system. *Annals of the New York academy of science. Teleological mechanisms*. 1948, vol. 50, issue 4. s. 197-220.
4. WICK, Mark (ed.). *GeoNames ontology* [online]. April 02, 2007 [cit. 2009-08-05]. Dostupný z WWW: <<http://www.geonames.org/ontology/>>.
5. VIRILIO, Paul. *Open Sky*. Translated by Julie Rose. London : Verso, 1997. 152 s. ISBN 185984880X.
6. VINGE, Vernor. *Technological singularity* [online]. March 31, 1993 [cit. 2009-08-08]. Dostupný z WWW: <<http://www-rohan.sdsu.edu/faculty/vinge/misc/WER2.html>>.
7. VENTER, Hein et al. *New approaches for security, privacy and trust in complex environments : proceedings of the IFIP TC-11 22nd International Information Security Conference (SEC 2007), 14-16 May 2007, Sandton, South Africa*. New York : Springer, c2007. xv, 498 s. ISBN 978-0387723662.
8. Van HARMELEN, Frank; HORROCKS, Ian; KLEIN, Michel. *Instance-OIL* [online]. November 28, 2000 [cit. 2009-07-03]. Dostupný z WWW: <<http://www.ontoknowledge.org/oil/syntax/Instance-OIL/>>.
9. VENERIS, Bill. *Orthogonality and the DRY principle. A conversation with Andy Hunt and Dave Thomas, Part II* [online]. October 10, 2003 [cit. 2009-07-03]. Dostupný z WWW: <<http://www.artima.com/intv/dry.html>>.

10. SIPOS, Ruben et al. Demo: HistoryViz -Visualizing Events and Relations Extraced from Wikipedia. In *The Semantic Web: Research and Applications, 6th European Semantic Web Conference, ESWC 2009, Heraklion, Crete, Greece, May 31-June 4, 2009*. London : Springer 2009. ISBN 978-3-642-02120-6.
11. SIMMONS, James. *5 problems of the semantic web* [online]. February 27, 2007 [cit. 2009-07-28]. Dostupný z WWW: <<http://www.semanticfocus.com/blog/entry/title/5-problems-of-the-semantic-web/>>.
12. SCHOFIELD, Jack. *Facebook's social advertising system and the rise of the fansumer* [online]. November 07, 2007 [cit. 2009-07-29]. Dostupný z WWW: <<http://www.guardian.co.uk/technology/blog/2007/nov/07/facebookssocialadvertisings>>.
13. SHADBOLT, Nigel; HALL, Wendy; BERNERS-LEE, Tim. *The semantic web revisited* [online]. 2006 [cit. 2009-07-05]. Dostupný z WWW: <<http://www.aktors.org/publications/selected-papers/2006-2007/1-6.pdf>>.
14. POWELL, Gavin. *Beginning XML databases*. Indianapolis (In) : Wiley Publishing, c2007. xxiii, 470 s. ISBN 978-0471791201.
15. POLLOCK, Jeffrey T. *Semantic web for dummies*. Indianapolis (In) : Willey Publishing, 2009. ISBN 978-0470396797.
16. PEMBERTON, Steven. *Use new URI scheme to address non-web resources* [online]. Last modified 2008-08-03 [cit. 2009-08-07]. Dostupný z WWW: <[http://rdfa.info/wiki/Uri-vs-resource-ambiguity-problem#Use New URI Scheme to Address Non-Web Resources](http://rdfa.info/wiki/Uri-vs-resource-ambiguity-problem#Use_New_URI_Scheme_to_Address_Non-Web_Resources)>.
17. PASSIN, Thomas B. *Explorer's guide to the semantic web*. Greenwich : Manning, c2004. xxii, 281 s. ISBN 978-1932394207.
18. OUELLET, Roxane; OGBUJI, Uche. *Introduction to DAML: Part I* [online]. January 30, 2002 [cit. 2009-07-02]. Dostupný z WWW: <<http://www.xml.com/pub/a/2002/01/30/daml1.html>>.
19. OMELAYENKO, Borys. Preliminary ontology modeling for B2B content integration. In *Proceedings of the first international workshop on electronic business hubs at the twelfth international conference on database and expert systems applications (DEXA-2001), Munich, Germany, September 3, 2001*. Odkaz na web autora: <www.cs.vu.nl/~borys>.

20. NGUYEN, Ngoc Thanh; GRZECK, Adam; HOWLETT, Robert J. *Agent and multi-agent systems: technologies and applications*. New York : Springer, 2007. 729 s. ISBN 978-3540728290.
21. NELSON, Theodor H. *Session 4: Complex information processing. 4.2: A file structure for the complex, the changing and the indeterminate* [online]. 1965 [cit. 2009-07-15]. Dostupný z WWW: <http://staff.cch.kcl.ac.uk/~wmccarty/DHO/readings/foreground/1965/Nelson,%20A%20file%20structure%20for%20the%20complex.pdf>.
22. NEBEL, Bernhard. *Frame-Based systems* [online]. MIT, 1999 [cit. 2009-07-02]. Dostupný z WWW: <http://www.cs.umbc.edu/771/current/papers/nebel.html>.
23. NARDI, Daniele; BRACHMAN, Ronald J. *An introduction to description logics* [online]. [cit. 2009-07-04]. Dostupný z WWW: <http://www.inf.unibz.it/~franconi/dl/course/dlhb/dlhb-01.pdf>.
24. MILLER, George, A. WordNet: a lexical database for english. *Communications of the ACN*. November 1995, vol. 38, no. 11. Dostupný také na WWW: http://reference.kfupm.edu.sa/content/w/o/wordnet_a_lexical_database_for_english_19806.pdf.
25. MIKA, Peter. *Social networks and the semantic web*. New York : Springer, 2007. 234 s. ISBN 978-0387710006.
26. MEERSMAN, Robert; TARI, Zahir (eds.). *On the move to meaningful internet systems 2004: OTM 2004 workshops : OTM confederated international workshops and posters, GADA, JTRES, MIOS, WORM, WOSE, PhDS, and INTEROP 2004, Agia Napa, Cyprus, October 25-29, 2004 : proceedings*. Berlin : Springer, 2004. xxv, 1648 s. ISBN 978-3540236641.
27. McILRAITH, Sheila A.; CAO SON, Tran; ZENG, Honglei. *Semantic Web Services* [online]. 2001 [cit. 2009-08-06]. Dostupný z WWW: <http://www.is.pku.edu.cn/~lz/teaching/dcn/references/semantic%20web%20service.pdf>.
28. MAREŠ, Michael. *Informační věk jako síťové paradigma* [online]. April 15, 2009 [cit. 2009-08-07]. Dostupný z WWW: <http://medialniproroci.blogspot.com/2009/04/informacni-vek-jako-sitove-paradigma.html>.
29. MALO, Roman. *XML namespaces* [prezentace Power Point]. Brno : Pef MZLU, 2006.

30. MacMANUS, Richard. *Twine: the first mainstream semantic web app?* [online]. October 18, 2007 [cit. 2009-08-03]. Dostupný z WWW: <http://www.readwriteweb.com/archives/twine_first_mainstream_semantic_web_app.php>.
31. LASSILA, Ora; SWICK, Ralph R. (eds.). *Resource Description Framework (RDF) model and syntax specification* [online]. January 05, 1999 [cit. 2009-06-25]. Dostupný z WWW: <<http://www.w3.org/TR/PR-rdf-syntax/>>.
32. LACY, Lee E. *Owl: representing information using the web ontology language*. Victoria : Trafford Publishing, 2005. 302 s. ISBN 978-1412034487.
33. KUHN, Thomas Samuel. *Struktura vědeckých revolucí*. 1. vyd. Praha : OIKOYMENH, 1997. 206 s. ISBN 80-86005-54-2.
34. KHARE, Rohit; ÇELIK, Tantek. *Microformats: a pragmatic path to the semantic web* [online]. May 2006 [cit. 2009-07-05]. Dostupný z WWW: <<http://www2006.org/programme/files/pdf/p116.pdf>>.
35. HUNTER, David et al. *Beginning XML*. 4th ed. Indianapolis (In) : Wiley Publishing, c2007. xxxvi, 1039 s. ISBN 978-0470114872.
36. HAROLD, Eliotte Rusty; MEANS, W. Scott. *XML in a nutshell*. 3rd ed. Beijing; Sebastopol : O'Reilly Media, c2004. xix, 689 s. ISBN 978-0596007645.
37. ANTONIOU, Grigoris; Van HARMELEN, Frank. *Web Ontology Language: OWL* [online]. [cit. 2009-07-03]. Dostupný z WWW: <<http://www.math.vu.nl/~frankh/postscript/OntoHandbook03OWL.pdf>>.
38. HALPIN, Terry; MORGAN, Tony. *Information modeling and relational databases*. 2nd ed. Burlington (Ma) : Elsevier/Morgan Kaufmann Publishers, c2008. xxvi, 943 s. ISBN 978-0123735683.
39. GUARINO, Nicola; GIARETTA, Pierdaniele. *Ontologies and knowledge bases* [online]. [cit. 2009-07-02]. Dostupný z WWW: <<http://www.loa-cnr.it/Papers/KBKS95.pdf>>.
40. GRUBER, Thomas R. *Toward principles for the design of ontologies* [online]. [cit. 2009-06-29]. Dostupný z WWW: <<http://www.itee.uq.edu.au/~infs3101/Readings/OntoEng.pdf>>.
41. GROMOV, Gregory R. *The roads and crossroads of internet history* [online]. [cit. 2009-06-25]. Dostupný z WWW: <<http://www.netvalley.com/intval2.html>>.

42. GADAMER, Hans-Georg. *Philosophical hermeneutics*. Translated and edited by David E. Linge. 30th anniversary edition. Berkeley : University of California Press, 2008. 304 s. ISBN 978-0520256408.
43. FLORIDI, Luciano. *Web 2.0 vs. the semantic web: a philosophical assessment* [online]. 2007 [cit. 2009-08-08]. Dostupný z WWW: <<http://www.philosophyofinformation.net/publications/pdf/w2vsw.pdf>>.
44. HEPP, Martin et al. *Semantic business process management: using semantic web services for a business process management* [online]. 2005-12-12 [cit. 2009-07-30]. Dostupný z WWW: <<http://people.kmi.open.ac.uk/domingue/papers/sbpm-sws-for-business-processes.pdf>>.
45. DODGE, Don. *Why powerset is important and different* [online]. Microsoft, Startup Zone. 2008-07-02 [cit. 2009-08-06]. Dostupný z WWW: <http://www.microsoftstartupzone.com/Blogs/the_next_big_thing/Lists/Posts/Post.aspx?List=6bab7b08-81ca-4602-bd97-4b7b2c893e88&ID=615>.
46. DEVEDŽIČ, Vladan. *Semantic web and education*. New York : Springer, c2006. xi, 353 s. ISBN 978-0387-35417-0.
47. DACONTA, Michael C.; ORBST, Leo Joseph; SMITH, Kevin T. *The semantic web: a guide to the future of XML, web services, and knowledge management*. Indianapolis : John Wiley and Sons, 2003. 312 s. ISBN 978-0471432579.
48. ČERNOŠ, Ondřej. *Sémantický web* [prezentace Power Point]. Praha : Univerzita Karlova, Ústav informčních studií a knihovnictví, 2009.
49. CORCHO, Óscar. *A layered declarative approach to ontology translation with knowledge preservation*. Volume 116 of Frontiers in artificial intelligence and applications. Amsterdam : IOS Press, 2005. 149 s. ISBN 978-1586034771.
50. CORCHO, Oscar, GÓMEZ-PÉREZ, Asunción; FERNANDEZ-LOPEZ, Mariano. *Ontological engineering: with examples from the areas of knowledge management, e-commerce and the semantic web*. London : Springer, 2004. 415 s. ISBN 978-1852335519.
51. World Wide Web Consorciium. *The ubiquitous web* [online]. Biot : W3C, 2005 [cit. 2009-08-08]. s. 5. Dostupný z WWW: <[http://www.w3.org/2005/Talks/0621-dsr-ubiweb/#\(5\)](http://www.w3.org/2005/Talks/0621-dsr-ubiweb/#(5))>.
52. CASTELLS, Manuel. *The rise of the network society*. Oxford : Blackwell, 2000. 481 s. ISBN 978-0631221401.

53. CASAREZ, Vince et al. *Reshaping your business with web 2.0: using new social technologies to lead business transformation*. 1. ed. [s.l.] : McGraw-Hill Osborne Media, 2008. 259 s. ISBN 978-0071600781.
54. CARDOSO, Jorge; HEPP, Martin; LYTRAS, Miltiadis D. (eds.). *The semantic web: real-world applications from industry*. 1. ed. London : Springer, 2007. 308 s. ISBN 978-0387485300.
55. BUSH, Vannevar. As We May Think. *The Atlantic* [Online]. July 1945 [cit. 2009-08-03]. Dostupný z WWW: <<http://www.theatlantic.com/doc/194507/bush>>.
56. BREITMAN, Karin K.; CASANOVA, Marco A.; TRUSZKOWSKI, Walt T. *Semantic web: concepts, technologies and applications (NASA monographs in systems and software engineering)*. 2. ed. London : Springer, 2007. 327 s. ISBN 9781846285813.
57. BRAY, Tim; POLI, Jean; SPERBERG-McQUEEN, C.M. (eds.). *Extensible Markup Language (XML) 1.0* [online]. February 10, 1998 [cit. 2009-06-27]. Dostupný z WWW: <<http://www.renderx.com/Demos/fo2html/xml.pdf>>.
58. BORST, Willem Nico. *Construction of engineering ontologies for knowledge sharing and reuse*. Enschede : Centre for Telematics and Information Technology, 1997. 231 s. ISBN 978-9036509886.
59. BOJARS, U. et al. *Using the semantic web for linking and reusing data across, web 2.0 communities* [online]. December 04, 2007 [cit. 2009-07-04]. Dostupný z WWW: <http://www.nuigalway.ie/electronic_engineering/documents/jb_j_web_semantics_2008.pdf>.
60. BIZER, C. et al. *DBpedia - a crystallization point for the web of data* [online]. May 25, 2009 [cit. 2009-08-05]. Dostupný z WWW: <<http://www.wiwiss.fu-berlin.de/en/institute/pwo/bizer/research/publications/Bizer-et-al-DBpedia-CrystallizationPoint-JWS-Preprint.pdf>>.
61. BERNERS-LEE, Tim; FIELDING, R.; MASINTER, R. *Uniform Resource Identifier (URI): generic syntax* [online]. January 2005 [cit. 2009-07-06]. Dostupný z WWW: <<http://tools.ietf.org/html/rfc3986>>.
62. BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. The semantic web. *Scientific American Magazine* [online]. May 2001 [cit. 2009-06-24]. Dostupný z WWW: <<http://www.scientificamerican.com/article.cfm?id=the-semantic-web>>.

63. BERNERS-LEE, Tim; FIELDING, R.; MASINTER, L. *Uniform Resource Identifier (URI): generic syntax* [online]. January 2005 [cit. 2009-06-23]. Dostupný z WWW: <<http://labs.apache.org/webarch/uri/rfc/rfc3986.html>>.
64. BAUDRILLARD, Jean. *Simulations*. New York : Semiotext(e), 1983. 159 s. ISBN 978-0936756028.
65. Social media research blog. *Advertising models: from contextual to conceptual/semantic to social* [online]. July 20, 2008 [cit. 2009-07-29]. Dostupný z WWW: <<http://socialmedia.typepad.com/blog/2008/07/advertising-models-from-contextual-to-conceptualsemantic-to-social.html>>.
66. ANTONIOU, Grigoris; Van HARMELEN, Frank. *A semantic web primer*. Cambridge (Mass) : MIT Press, 2004. 238 s. ISBN-10: 0262012103.
67. DBpedia: a nucleus for a web of open data. In *Lecture notes in computer science*. Berlin : Springer, 2008. s. 722-735. ISBN 978-3540762973.
68. D2R Server. *D2R server for the Gutenberg Project* [online]. [cit. 2009-08-04]. Dostupný z WWW: <<http://www4.wiwiss.fu-berlin.de/gutendata/>>.
69. W3C, Semantic Web. *W3C semantic web activity publications, Specifications (recommendations and notes)* [online]. 2009 [cit. 2009-07-06]. Dostupný z WWW: <<http://www.w3.org/2001/sw/Specs.html>>.
70. Princeton University. *About WordNet* [online]. 2009 [cit. 2009-07-29]. Dostupný z WWW: <<http://wordnet.princeton.edu/>>.
71. DBpedia. *Accessing the DBpedia data set over the web* [online]. Last modification 2009-08-06 [cit. 2009-08-07]. Dostupný z WWW: <<http://wiki.dbpedia.org/OnlineAccess>>.
72. DBpedia. *Credits* [online]. Last modification 2009-07-03 [cit. 2009-08-05]. Dostupný z WWW: <<http://wiki.dbpedia.org/Team>>.
73. Semantic Web. *Tools* [online]. Last modified 2008-11-27 [cit. 2009-08-02]. Dostupný z WWW: <<http://semanticweb.org/wiki/Tools>>.
74. ORACLE. *Oracle database semantic technologies overview* [online]. 2008-07-22 [cit. 2009-08-07]. Dostupný z WWW: <http://www.oracle.com/technology/tech/semantic_technologies/pdf/oracle%20db%20semantics%20overview%2020080722.pdf>.
75. ORACLE, Technology network. *Semantic technologies center* [online]. [cit. 2009-08-07]. Dostupný z WWW: <http://www.oracle.com/technology/tech/semantic_technologies/index.html>.

76. OpenRDF.org ... home of Sesame. *About* [online]. 2009 [cit. 2009-08-01]. Dostupný z WWW: <<http://www.openrdf.org/about.jsp>>.
77. MyWikiBiz, Author your legacy. *Help:ASK* [online]. August 18, 2009 [cit. 2009-08-04]. Dostupný z WWW: <<http://www.mywikibiz.com/Help:ASK>>.
78. MusicBrainz. *Welcome to MusicBrainz!* [online]. 2009-05-24 [cit. 2009-08-05]. Dostupný z WWW: <<http://musicbrainz.org/>>.
79. METAWEB. *About Freebase* [online]. [cit. 2009-08-04]. Dostupný z WWW: <<http://www.metaweb.com/about/>>.
80. Freebase. *Louvre* [online]. 2009 [cit. 2009-08-06]. Dostupný z WWW: <<http://www.freebase.com/view/en/louvre>>.
81. CORDIS. *Information society technologies: thematic priority under the specific programme "Integrating and strengthening the European research area" (2002-2006)* [online]. 2002-09-03 [cit. 2009-08-09]. Dostupný z WWW: <http://cordis.europa.eu/search/index?fuseaction=prog.document&PG_LANG=EN&PG_RCN=5465040&pid=0&q=148FCB5C0282FB07F9183FCA06C3EC6&type=sim>.
82. ARC RDF Classes for PHP. *Easy RDF and SPARQL for LAMP systems* [online]. Last modified 2009-06-22 [cit. 2009-08-07]. Dostupný z WWW: <<http://arc.semsol.org/>>.
83. Nova Spivack. *Making sense of the semantic web*. San Francisco : Radar networks, [s.a.]. Odkaz na stránky společnosti: <<http://novaspivack.typepad.com/>>.

Použitá literatura

- ALLAN, James (ed.). 2002. *Topic detection and tracking: event-based information organization*. London : Springer, 2002. 280 s. ISBN 978-0792376644.
- CIMIANO, Philipp. 2006. *Ontology learning and population from text: algorithms, evaluation and applications*. London : Springer, c2006. 347 s. ISBN 978-0387306322.
- COCCHIARELLA, Nino, B. 2007. *Formal ontology and conceptual realism*. London : Springer, 2007. xxiv, 332 s. ISBN 978-1402062032.
- COPPIN, Ben. 2004. *Artificial intelligence illuminated*. Boston : Jones and Bartlett Publishers, c2004. xxv, 739 s. ISBN 978-0763732301.

- GREENFIELD, Adam. 2006. *Everyware: the dawning age of ubiquitous computing*. Berkeley (Ca) : New Riders, c2006. iii, 267 s. ISBN 978-0321384010.
- MANNING, Christopher D.; SCHULTZE, Hinrich. 1999. *Foundations of statistical natural language processing*. Cambridge (Mass) : MIT Press, c1999. xxxvii, 680 s. ISBN 978-0262133609.
- POSLAD, Stefan. 2009. *Ubiquitous computing: smart devices, environments and interactions*. Chichester (U.K.) : WileyBlackwell, 2009. xxiv, 473 s. ISBN 978-0470035603.
- POWERS, Shelly. 2003. *Practical RDF*. Beijing; Sebastopol : O'Reilly Media, c2003. xv, 331 s. ISBN 978-0596002633.
- RICHARDSON, Leonard; RUBY, Sam. 2007. *RESTful web services*. Farnham : O'Reilly Media, c2007. xxiv, 419 s. ISBN 978-0596529260.
- SANCHEZ, Elie (ed.). 2006. *Fuzzy logic and the semantic web*. Oxford : Elsevier Science, 2006. 496 s. ISBN 978-0444519481.
- SEGARAN, Toby. 2007. *Programming collective intelligence: building smart web 2.0 applications*. Beijing; Sebastopol : O'Reilly Media, c2007. xxi, 334 s. ISBN 978-0596529321.
- WALTON, Christopher D. 2006. *Agency and the semantic web*. Oxford : Oxford University Press, 2006. xvii, 249 s. ISBN 978-0199292486.
- WILSON, Stephen. 2003. *Information arts: intersections of art, science and technology*. Cambridge (Mass) : MIT Press, 2003. xxiv, 945 s. ISBN 978-0262731584.

Evidence výpůjček

Prohlášení:

Dávám svolení k půjčování této diplomové práce. Uživatel potvrzuje svým podpisem, že bude tuto práci řádně citovat v seznamu použité literatury.

V Praze, 20.08. 2009.

Eduard Piňos

Jméno	Katedra / Pracoviště	Datum	Podpis