

**Univerzita Karlova v Praze**

**Filozofická fakulta**

**Ústav informačních studií a knihovnictví**

Studijní program: informační studia a knihovnictví

Studijní obor: informační studia a knihovnictví

**Jakub Roztočil**

**Sémantický web**

**Bakalářská práce**

Praha 2009-05- 31

Vedoucí bakalářské práce:

Doc. RNDr. Jiří Souček DrSc.

Oponent bakalářské práce:

Datum obhajoby:

Hodnocení:

**Prohlášení:**

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

V Praze, 8. června 2009

.....

podpis studenta

## **Identifikační záznam**

ROZTOČIL, Jakub. *Sémantický web [Semantic web]*. Praha, 2009. 99 s. Bakalářská práce. Univerzita Karlova v Praze, Filozofická fakulta, Ústav informačních studií a knihovnictví. Vedoucí bakalářské práce Jiří Souček.

## **Abstrakt**

Práce se nejdříve stručně zabývá historií webu a jeho architekturou, popisuje také jeho současný stav a nedostatky. Poté vymezuje, co je sémantický web a jaké jsou jeho hlavní myšlenky a cíle. Následně popisuje jeho technologie sémantického webu, zejména RDF a ontologie a také se zabývá propojování jednotlivých zdrojů dat tvořících sémantický web. Nakonec naznačuje, jakým směrem by se mohl ubírat budoucí vývoj.

## **Klíčová slova**

Sémantický web, WWW, sémantika, ontologie, pořádání informací, webové služby, elektronické informační zdroje, RDF, OWL, SPARQL, W3C, linked data.

# Obsah

Předmluva.....	1
1 Úvod.....	4
.1.1 Základní pojmy.....	4
.1.1.1 Sémantika.....	4
.1.1.2 Webový zdroj.....	4
.1.1.3 Newebový zdroj.....	4
.1.2 Historie webu.....	5
.1.3 Současný stav.....	6
.1.4 Nedostatky současného webu.....	7
2 Sémantický web.....	7
.2.1 Historie.....	8
.2.2 Hlavní myšlenky a cíle.....	10
.2.2.1 Cíle sémantického webu.....	10
.2.2.2 Anyone can say Anything about Any topic.....	13
.2.2.3 Ověřitelnost pravdivosti.....	14
.2.2.4 Open world assumption.....	14
.2.2.5 Nejednotné pojmenování.....	14
3 Technologie sémantického webu.....	15
.3.1 RDF.....	15
.3.1.1 RDF trojice.....	16
.3.1.2 Identifikátory.....	17
.3.1.3 Serializace RDF.....	18
.3.1.4 Odvozování.....	22
.3.1.5 RDFa.....	22
.3.2 Ontologie.....	23
.3.3 RDF Schema.....	25
.3.3.1 Třídy.....	26
.3.3.2 Vlastnosti.....	26
.3.4 OWL.....	27
.3.4.1 OWL Full.....	28
.3.4.2 OWL DL.....	28
.3.4.3 OWL Lite.....	29
.3.5 Existující ontologie.....	29
.3.5.1 Dublin Core.....	29

.3.5.2 SKOS.....	30
.3.5.3 FOAF.....	31
.3.6 Linked data.....	32
.3.6.1 Použití URI jako jmen pro věci.....	33
.3.6.2 Použití HTTP URI bylo možné získat informace o věcech.....	33
.3.6.3 Poskytovat užitečné informace při navštívení URI.....	33
.3.6.4 Zahrnout odkazy na další věci.....	34
.3.6.5 Způsoby provázání dat na sémantickém webu.....	34
.3.7 Existující otevřené báze provázaných dat.....	37
.3.7.1 DBpedia.....	38
.3.7.2 GeoNames.....	41
.3.8 Prohlížeče sémantického webu.....	41
.3.8.1 The Tabulator.....	42
.3.9 SPARQL.....	44
4 Závěr.....	45
Seznam použité literatury.....	47

## Předmluva

Web jako platforma pro publikování a sdílení informací dosáhl obrovského úspěchu. Za dvacet let od svého vzniku se rozšířil z vědeckých a akademických institucí do téměř každé domácnosti. Za tímto úspěchem stojí jeho jednoduchost a otevřenost. Dokumenty, které tuto pavučinu tvoří, jsou v převážné většině ve formě hypertextově provázaných HTML<sup>1</sup> dokumentů. HTML poskytuje možnost jednotlivé části textu dokumentu označkovat a takto instruovat uživatelského agenta<sup>2</sup>, jakým způsobem má kterou část vykreslit na zobrazovacím zařízení uživatele. Tímto způsobem je člověku, který dokument čte, umožněno pochopit význam jednotlivých částí textu.

Tuto funkci - zprostředkovávání odkazy provázaných dokumentů v lidem srozumitelném formátu - plní klasický web velmi dobře. Kde naopak selhává, je možnost zprostředkovávat tuto sémantiku ve formě, která by byla srozumitelná také strojům. K dispozici je velice omezené množství způsobů, jak klasické HTML dokumenty obohatit o metadata. Současnému webu chybí sémantika, která by umožnila strojům předat pro ně srozumitelné významy a vztahy mezi nimi. Důvodů pro snahu o strojovou zpracovatelnost je mnoho a budou postupně probrány v následujících kapitolách. Z příčin těchto omezení klasického webu od konce devadesátých let minulého století vzniká pod patronací W3C řada technologií a doporučení, která mají web rozšířit o nástroje k definování strojově zpracovatelné sémantiky.

Se sémantickým webem jsem se poprvé setkal v roce 2004 jako uživatel XML/RDF ve formě RSS kanálů mých oblíbených webů. Ihned mne tato technologie zaujala, protože umožňovala něco, s čím jsem se do té doby nesetkal. Jednalo se možnost automatické periodické kontroly, zda nevyšel nový článek, aniž by bylo nutné web navštěvovat a tuto kontrolu

---

1 HTML, HyperText Markup Language - značkovací jazyk určený pro tvorbu hypertextových dokumentů

2 Uživatelský agent, anglicky User Agent - používá se pro označení softwarového vybavení, s nímž uživatel vstupuje na web a prohlíží si stránky

provádět manuálně. Zajímavé pro mne toto setkání bylo také z hlediska mé profese, protože již od roku 2003 pracuji jako vývojář webových aplikací a mnohokrát jsem se setkal se situací, kdy bylo např. potřeba z velkého množství webových stránek vyextrahovat určité informace a pokaždé bylo nutné naprogramovat ad hoc řešení, které se s daným požadavkem vypořádalo. Tušil jsem, že právě sémantický web a jeho přidaná strojově srozumitelná sémantika by mohl podobný typ problémů řešit. Tématem sémantického webu se od té doby zabývám a sleduji jeho vývoj.

Sémantický web úzce souvisí s předmětem mého studia na ÚISK, neboť se jedná o metodu organizace znalostí. Dotýká se také oblastí jako klasifikace (ontologie, SKOS, ...), identifikační popis (identifikátory entit a vlastností pomocí URI) atd.

Výše uvedené důvody mne vedly k rozhodnutí zvolit si sémantický web za téma své bakalářské práce.

Cílem této práce je stručné seznámení s problematikou sémantického webu. Budou v ní popsány důvody pro jeho vznik, jaké technologie byly a stále jsou pro jeho potřeby vyvíjeny a také jaké jsou jeho cíle. Vzhledem ke komplexnosti tématu není cílem práce hlubší analýza, ale spíše široký záběr popisovaných aspektů sémantického webu.

Jako zdroje mi posloužily téměř výhradně anglicky psané zdroje, neboť v českém jazyce kvalitních a originálních dokumentů příliš není. Většina použitých zdrojů je dostupných online, neboť vzhledem k tématu autoři relevantních textů používají jako médium pro publikaci svých prací zejména web. Citace jsou uváděny pomocí průběžných poznámek v textu podle norem ČSN 690 a ČSN 690-2.

Práce je rozdělena do čtyř kapitol. První kapitola se stručně zabývá historií webu a jeho architekturou. Popisuje také současný stav webu a některé jeho problémy a nedostatky, které by bylo možné za pomoci technologií sémantického webu odstranit. Druhá kapitola je věnována



sémantickému webu obecně. Shrnuje jeho historii, hlavní myšlenky a cíle. Ve třetí kapitole jsou popsány hlavní technologie sémantického webu, např. RDF a ontologie. Čtvrtá kapitola se věnuje propojování různých zdrojů dat sémantického webu, které je známé pod pojmem Linked Data. Závěr shrnuje předchozí kapitoly a pokouší se odhadnout budoucí vývoj sémantického webu. Struktura se mírně odlišuje od původního zadání, protože toto zadání vznikalo v době, kdy jsem o sémantickém webu hlubší neměl dostatečné znalosti k vytvoření adekvátní struktury.

# 1 Úvod

Předtím než bude možné popsat sémantický web, který má být rozšířením webu současného, je vhodné vysvětlit základní pojmy a stručně popsat aktuální stav a historii WWW, protože z jeho historie sémantický web vychází a na současný stav navazuje.

## **.1.1 Základní pojmy**

### **.1.1.1 Sémantika**

Vzhledem k obecnosti pojmu sémantika je potřeba vymezit, jak je v kontextu webu tento termín používán. Formální studium sémantiky se protíná s mnoha obory. V lingvistice se sémantika zabývá významem slov. V případě sémantického webu se hovoří o pro stroje použitelném, případně pochopitelném či čitelném obsahu. Slovo „pochopení“ je zde myšleno v přeneseném slova smyslu, neboť se nejedná o kognitivní - tedy myšlení a rozum zahrnující - proces, jak je tomu v případě lidského chápání. Pochopení zde znamená to, že v datech, která počítačový program zpracovává, může pomocí identifikátorů, předdefinované ontologie a dalších nástrojů sémantického webu rozpoznat entity reálného světa, o kterých data nesou informace a data správně interpretovat.

### **.1.1.2 Webový zdroj**

Webovým zdrojem se myslí nějaká entita, která je na webu nejenom identifikovatelná pomocí URI, ale též na něm přímo existuje. Jedná se například o webové dokumenty.

### **.1.1.3 Newebový zdroj**

Sémantický web rozšiřuje pojem zdroj na veškeré entity, které jsou na webu pomocí systému URI identifikovatelné. Newebové zdroje je označení pro zdroje, které na webu přímo neexistují.

## .1.2 Historie webu

Historie webu začíná v březnu roku 1989, kdy Tim Berners-Lee předává managementu švýcarského CERNu<sup>3</sup>, kde v té době pracuje, návrh<sup>4</sup> na distribuovaný hypertextový systém, který zatím označuje výrazem Mesh. Tento systém měl za úkol pomoci vyřešit problémy se sdílením a zpětným vyhledáváním informací, které bylo v té době v CERNu panovaly. V listopadu 1990 s pomocí Roberta Cailliau publikuje formálnější návrh<sup>5</sup> na vytvoření hypertextového projektu, v té době již nazývaný WorldWideWeb, popsaný jako síť uzlů s daty uloženými v hypertextových dokumentech – webových stránkách. Tyto stránky měly být prohlíženy za použití webových prohlížečů skrze počítačovou síť. Měsíc po zveřejnění tohoto návrhu také naprogramoval první webový prohlížeč a první web server a vytvořil první webové stránky. Koncept hypertextu, který ve svých návrzích Tim Berners-Lee představil, vycházel ze starších projektů ze 60. let. Jeho hlavní přínos spočíval v propojení myšlenky hypertextu s Internetem. Web byl tedy původně zamýšlen především jako způsob zpřístupňování a sdílení pro vědecké pracovníky.

Významnou událostí prvních let webu bylo v březnu 1993 oficiální oznámení<sup>6</sup> CERNu, které přineslo informaci o tom, že veškeré CERNem vyvinuté softwarové vybavení (prohlížeč, web server a další technologie) jsou od té chvíle k dispozici zdarma komukoliv, bez jakýchkoliv poplatků (licence Public Domain). Tato událost se v průběhu času ukázala být natolik zásadní, že o ní Tim Berners-Lee později prohlásil, že rozhodnutí CERNu uvolnit základní kameny webu a jeho protokoly zdarma dostupné a bez jakýchkoli dodatečných překážek, bylo klíčové pro existenci webu, a že bez tohoto příspěví by se následné obrovské investice do webových technologií neodehrály a my bychom dnes web neměli<sup>7</sup>. Otevřenost technologií, která v prostředí webu od počátku panuje, tedy byla jedním z

---

3 CERN, European Organization for Nuclear Research

4 BERNERS-LEE, Tim, Information Management: A Proposal

5 BERNERS-LEE, Tim, CAILLIAU, Robert, WorldWideWeb

6 CERN, Ten Years Public Domain for the Original Web Software

7 CERN, CERN celebrates Web anniversary

důvodů enormního úspěchu, kterého web (na rozdíl od podobných, avšak uzavřených projektů) dosáhl. Tato skutečnost se do značné míry odráží i v sémantickém webu.

V srpnu 1994 Tim Berners-Lee zakládá World Wide Web Consortium (W3C). Sdružení se skládá s různých organizací, které spojuje zájem o vytvoření standardů a doporučení za účelem zvýšení kvality webu.

Rok 1996 je ve znamení nárůstu počtu firemních webů a komerčních prezentací na webu. Do té doby byly výhody webu viděny především v možnosti publikovat, ale začíná se projevovat možnost obousměrné komunikace, tedy nejenom směrem od autorům ke čtenářů, ale i naopak. Toto má za následek vznik e-komerce.

V roce 1999 začíná období známé jako *dot-com-boom*, které se postupem času ukáže být spíše *dot-com-bubble*. Za mohutné pomoci investorů vzniká mnoho firem, které provozují většinu svých činností čistě na webu. V roce 2001 tato bublina splaskává a většina těchto firem krachuje.

Od roku 2002 je patrný trend ve zlevňování a zlepšování dostupnosti internetové konektivity, s čímž se výrazně rozšiřuje uživatelská základna webu. Vzniká mnoho úspěšných obchodních modelů. Firmy jako Google, Ebay, Amazon apod. se stávají vysoce ziskové. Web se pomalu přesouvá do fáze označované jako 2.0.

Pro web 2.0 jsou typické vlastnosti jako uživatelsky generovaný obsah, blogy, wiki, sociální sítě, folksonomie a sdílení multimediálního obsahu. Uživatelé jsou nejenom konzumenty, ale také aktivními producenty webového obsahu. V této fázi webu se aktuálně nacházíme.

### **.1.3 Současný stav**

Jak jak již bylo zmíněno, růst webu umožnila jeho otevřenost. Dalším z faktorů, které vedly k současnému stavu, je síťový efekt<sup>8</sup>. Čím více stránek

---

8 ALLEMANG, Dean, et. al., Semantic web for the working ontoloigists

na webu je, tím užitečnějším médiem se stává a motivuje tak další lidi, aby na něm publikovali své dokumenty. Kdokoliv, aniž by musel mít znalosti programování nebo rozumět tomu, jak web technicky funguje, si může vytvořit své webové stránky či založit blog.

Současný web by se dal definovat jako web dokumentů určených pro lidi<sup>9</sup>. Je složen z obrovského množství různorodých hypertextově provázaných dokumentů, jejichž obsah pokrývá téměř jakékoliv myslitelné téma. Obsah je z velké míry vytvářen otevřenou komunitou lidí, kteří dobrovolně popisují témata, jimž rozumí.

Kromě dokumentů jsou na webu různorodé aplikace, které lidem do značné míry ulehčují běžné činnosti, umožňují vzdálenou spolupráci a přinášejí zábavu. Existují online aplikace interaktivních map, plánovačů tras, rezervační systémy letenek, sociální sítě, atd. Inovace přicházejí každým dnem.

### **.1.4 Nedostatky současného webu**

Dnešní webové aplikace mají stále značná omezení, která by mohla být s přidanou sémantikou odstraněna.

Omezení webových aplikací vyplývají především z limitované interoperability. Existují např. velmi sofistikované a užitečné mapové aplikace, které dokáží přesně naplánovat trasy mezi téměř libovolnými místy, ale když si např. na webu nějaké instituce najdeme její adresu a chceme se podívat, jak se na ní dostat, je potřeba jí ručně zkopírovat do vyhledávacího políčka webové aplikace. Není možné např. prostým předáním URL stránky, kde se adresa nachází, zobrazit tuto adresu na mapě. Aplikace jsou ve své funkčnosti limitovány daty, se kterými pracují.

## **2 Sémantický web**

Jedna z neoficiálních definic sémantického webu říká, že se jedná o

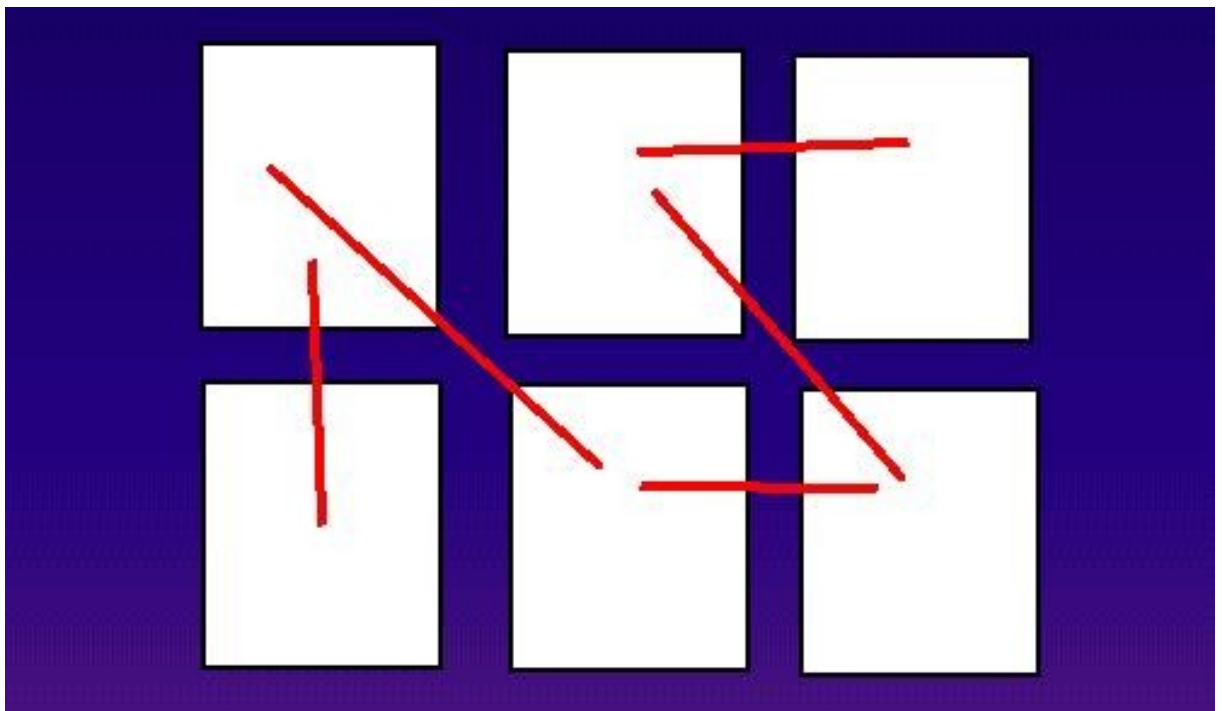
<sup>9</sup> BERNERS-Lee, Tim, Semantic Web, Scientific American

rozšíření dnešního webu dokumentů na web dat<sup>10</sup> a později a na web věcí<sup>11</sup>, která by měla být přístupné pomocí existující prvků architektury webu.

V této kapitole bude dále stručně popsána historie sémantického webu, jeho hlavní myšlenky a také cíle, které si klade.

## .2.1 Historie

Prvky sémantického webu byly patrné již v původních návrzích WWW. Jeden z prvních popisů sémantického webu a potřeby jeho vytvoření podal Tim Berners-Lee na konferenci WWW v Ženevě v roce 1994<sup>12</sup>. Existující web popsal jako síť uzlů propojených odkazy, což demonstroval následujícím schématem:



Zmiňuje užitečnost toto konceptu pro uživatele a dále hovoří o tom, jakým způsobem by přidaná sémantika tuto užitečnost navýšila. Přidání sémantiky má zahrnout umožnění tvorby dokumentů, které obsahují

<sup>10</sup>W3C, W3C Semantic Web FAQ

<sup>11</sup>Web of things - TODO: citace

<sup>12</sup>BERNERS-LEE, Tim. Plenary at WWW Geneva 94

informace ve strojem čitelné podobě a možnost přiřadit odkazům informaci o vztahu odkazovaných entit. Přidání těchto prvků má umožnit naplno využít schopnosti počítačů k tomu, aby informace na webu byly využitelné i jinak, než že je možné je číst.

Důležitým efektem má být také vyvinutí abstraktního prostoru linkovaného s realitou, zejména pomocí systému URI, aby tak počítače mohly za uživatele dělat běžné úkoly, jako je např. nalezení vhodného domu ke koupi a vyjednání převod vlastnictví. Důležitým faktorem je bezpečnost protokolu, pomocí něhož by toto probíhalo.

Později Berners-Lee definoval svojí vizi jasněji:

*„I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web - the content, links, and transactions between people and computers. A ‘Semantic Web’, which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The ‘intelligent agents’ people have touted for ages will finally materialize.“<sup>13</sup>*

V srpnu roku 1997 byla organizací W3C založena pracovní skupina, která měla za úkol vytvořit RDF - obecný rámec pro popis zdrojů<sup>14</sup>, metadatový model a základní stavební kámen sémantického webu. RDF vycházelo z existující specifikace PICS<sup>15</sup>, což byla v té době používaná metoda přidávání metadat do webových stránek za účelem označit dokumenty především podle vhodnosti pro děti s využitím pro rodiče a vzdělávací instituce<sup>16</sup>. První specifikace<sup>17</sup> RDF byla publikována v roce 1999<sup>18</sup> a definovala formální model a gramatiku tohoto jazyka.

---

13 Wikipedia, heslo Semantic Web

14 BOYÉ, RDF : What's in it for us?

15 PICS, Platform for Internet Content Selection

16 W3C, Platform for Internet Content Selection (PICS)

17 W3C nazývá své specifikace jako doporučení (recommendation), ale v podstatě mají status standardu, proto zde bude používán buď termín specifikace nebo standard

18 W3C, RDF Recommendation, 1999

Poprvé byla vize sémantického webu širší veřejnosti představena v roce 2001 ve článku na serveru Scientific American<sup>19</sup>, ve kterém Berners-Lee a další autoři na několika příkladech demonstrují, čeho by mělo být možné pomocí sémantického webu dosáhnout.

V roce 2003 založila organizace W3C pracovní skupinu s názvem Semantic Web Activity<sup>20</sup>, která má za úkol vyvíjet další potřebné standardy a specifikace. V roce 2004 byla vydána první specifikace OWL, což je skupina jazyků, které umožňují tvorbu ontologií.

V roce 2007 byla založena iniciativa Linking Open Data, která má za úkol vypracovat doporučení pro publikaci RDF dat na webu. Cílem iniciativy je udržovat informace o otevřených ontologiích za účelem jejich znovu používání a kombinování raději než tvorba vlastních *ad hoc* řešení. Doporučení se týkají architektury pro URL identifikátory entit, způsobu jejich zpřístupňování apod.

## **.2.2 Hlavní myšlenky a cíle**

Žádná oficiální definice sémantického webu bohužel neexistuje.<sup>21</sup> Sémantický web poskytuje obecný rámec, který má umožnit to, aby data mohla být sdílena a znovu používána napříč aplikacemi, komerční sférou a komunitami.<sup>22</sup> Sémantický web je komplexní problematika, proto bude nejlépe ji popsat skrze jeho cíle, přístupy, teoretické a praktické příklady použití a konkrétní aplikace.

### **.2.2.1 Cíle sémantického webu**

Z již dříve citované vize ředitele W3C a dalších autorů<sup>23</sup> vyplývá, že na počátku byla myšlenka webu, který jde ve své užitečnosti mnohem dál, než je tomu v dnešní době a to i při využití informací, které jsou na něm

---

19 BERNERS-LEE, Tim et. al., The Semantic Web

20 W3C, Semantic Web Activity

21 W3C, W3C Semantic Web Activity

22 W3C, W3C Semantic Web FAQ

23 BERNERS-LEE, Tim, et. al., The Semantic Web



již k dispozici.

V příkladu uvedeném ve výše zmíněném článku je toto demonstrováno na situaci, kdy se dva sourozenci snaží pro svojí matku najít lékaře. Jsou vymezena určitá kritéria, která musí lékař splňovat. Hledají lékaře vhodného z hlediska jeho zaměření, dostatečně kvalifikovaného, působícího ve vhodné lokalitě a s volným termínem, který by vyhovoval všem zúčastněným. Jedná se informační potřebu, kterou lze uspokojit i v podmínkách webu nesémantického. Klasický postup by byl pravděpodobně zhruba takový:

- Na základě informační potřeby jsou zvolena klíčová slova vymezujících kritéria lokality a typu lékařské specializace a vhodně se spojí operátory selekčního jazyka.
- Takto formulovaný dotaz zadá některému z webových vyhledávačů.

Výsledky vyhledávání, které dnešní vyhledávače na svém výstupu poskytují, jsou založeny především na indexu vytvořeném na základě strojové analýzy webových stránek. Tato analýza pracuje s webovými zdroji na úrovni dokumentů a textů v nich obsaženým, proto se společně s odkazy vedoucími na relevantní webové prezentace lékařů mohou objevit i takové, které mají relevanci buď velmi nízkou nebo naprosto nulovou, a naopak některé vysoce relevantní dokumenty mohou kvůli absenci klíčových slov zcela chybět.

Dalším krokem by bylo zjištění ordinačních hodin v rámci webových prezentací jednotlivých lékařů a emailové či telefonické ověření, zda je v kalendáři lékaře volné místo, které nekoliduje s kalendáři všech zúčastněných.

Tato situace demonstruje nevyužitý potenciál webu. Přesto, že téměř všechny potřebné informace jsou již na webu strojům i lidem dostupné, pořád je nutné, aby to byl právě člověk, který vyhledá a vyhodnotí relevantnost webových zdrojů a posléze zkombinuje informace z více

zdrojů k uspokojení své informační potřeby.

Důvodem, proč je lidský faktor v tomto procesu stále nezbytný, je to, že dnešní web je při své enormní velikosti a obsažnosti pouze na minimální úrovni strojově pochopitelný. Specifikace jazyka HTML<sup>24</sup> definuje pouze obecnou, a pro automatizované řešení úkolu, jako je ten z příkladu naprosto nevyhovující, sémantiku. HTML umožňuje texty strukturovat a přiřazovat jednotlivým částem textu v kontextu jednoho dokumentu důležitost a další obecná vymezení významu pomocí značek jako H1, STRONG apod. Vyhledávače a další automatické programy to např. informuje o relativní významnosti daného úryvku textu, což umožňuje při vytváření indexu přiřadit těmto textům větší selekční váhu. Dále je webový prohlížeč instruován, jak má graficky danou část textu vykreslit. Tak tedy např. nadpisy jsou zobrazeny větším písmem, citace odsazené atd. Tato sémantika HTML má sice velký význam pro uživatele, který se tak - i v případě absence definice vzhledu dokumentu pomocí např. CSS<sup>25</sup> - může v textu za použití své inteligence a zkušeností snadno zorientovat a pochopit významy jednotlivých částí, ale nepředává to žádnou strojově čitelnou informaci o obsahu těchto elementů ve formě, která by umožňovala identifikovat popisované objekty reálného světa a další entity, o nichž dokument pojednává a pochopit vztahy mezi nimi.

Když se nyní opět vrátíme k příkladu a představíme si, že sémantický web je již existujícím a plně využívaným rozšířením webu klasického, tak by postup při uspokojování dané informační potřeby vypadal např. následovně<sup>26</sup>:

Jeden ze sourozenců by pověřil svého agenta sémantického webu, aby nalezl lékaře odpovídající požadovaným kritériím. Ten by pomocí informací

24 HTML, HyperText Markup Language - hypertextový značkovací jazyk. V tomto textu jsou, pokud není explicitně uvedena verze, pod tímto označením myšleny všechny verze HTML a XHTML, které byly do dnešního data publikovány jako doporučení W3C, nebere tedy v úvahu např. HTML 5, které je stále ve fázi dramatického vývoje a kde je situace mírně odlišná.

25 CSS, Cascading StyleSheets - kaskádové tabulky stylů. Umožňují definovat vzhled písma, barvu pozadí a další vizuální atributy HTML a XML dokumentů pro různá zobrazovací zařízení.

26 Volně podle BERNERS-LEE, Tim et. al., Semantic Web

získaných z různých zdrojů vyhledal ty lékaře, kteří daná kritéria splňují nejlépe. Spojil by se s jejich agenty a zkusil by se vyhledat termín, který by nejlépe vyhovoval všem zúčastněným.

V předchozích odstavcích byl popsán jeden významný cíl sémantického webu, tj. umožnit pomocí strojově čitelných dat počítačům využít potenciál jejich a také potenciál webu jakožto globálního a otevřeného prostoru pro sdílení informací za účelem uspokojování informačních a dalších potřeb uživatelů. Dosažení toho cíle je možné pomocí otevřených a sdílených ontologií, jejichž problematika bude popsána dále.

### **.2.2.2 Anyone can say Anything about Any topic**

Tento slogan, který je znám též jako AAA slogan a v českém překladu zní „kdokoliv může říct cokoli na jakékoliv téma“, odráží důležitý aspekt webu<sup>27</sup>. Vyjadřuje myšlenku, že kdokoliv se může vyjádřit o jakéhokoliv tématu a přispět tak ke všem již existujícím výrokům na toto téma. Jedná se o důležitou vlastnost klasického webu a sémantický web jí podporuje a dále s ní pracuje.

Jestliže se může vyjádřit skutečně každý a to na jakékoliv téma, tak to nutně znamená, že ne všechny informace publikované na webu jsou zcela přesné a některé z nich mohou být z různých důvodů nepřesné, zavádějící či záměrně nepravdivé.

Pro rozpory mezi jednotlivými tvrzeními může existovat mnoho důvodů, mezi nimiž jsou např. tyto<sup>28</sup>:

- Autoři tvrzení mohou zásadně nesouhlasit na dané téma.
- Autor nepravdivého tvrzení může úmyslně lhát.
- Autor nepřesného výroku může vycházet z nepravdivých informací.

---

<sup>27</sup> ALLEMANG, Dean, et. al., Semantic web for the working ontologists

<sup>28</sup> tamtéž

- Autor může vycházet ze zastaralých informací.

### **.2.2.3 Ověřitelnost pravdivosti**

Klasický web neposkytuje žádné univerzální mechanismy pro ověření platnosti informací na něm prezentovaných. Je na uživateli, aby sám za pomoci kritického myšlení zhodnotil důvěryhodnost zdroje a informací, které informační zdroj zpřístupňuje.

Tuto situaci neřeší ani sémantický web. Její popření by znamenalo také popření AAA sloganu. Snaha o její eliminaci by znamenala omezení jednoho ze základních principu webu, jeho otevřenost, která skrze síťový efekt umožnila jeho růst do dnešních rozměrů. Z toho vyplývá, že sémantický web musí toto vzájemné protiřečení si napříč jednotlivými výroky brát v potaz a přizpůsobit se mu. Tím, že výroky učiní globálně dostupné a na základě různých vztahů přiřaditelné k entitám, o nichž pojednávají, dává možnost téma zhodnotit v širších souvislostech a vybrat si ze všech dostupných výroků ty, které považujeme za nejdůvěryhodnější. Kromě toho sémantický web neposkytuje žádné speciální nástroje pro ověření výroků z hlediska jejich případného rozporu s realitou.

### **.2.2.4 Open world assumption**

Předpoklad otevřeného světa říká, že v žádném okamžiku nemůžeme nabýt jistotu, že máme všechny informace, které jsou na dané téma dostupné. V prostředí webu to může být způsobeno např. výpadkem serveru, na němž jsou umístěny dokumenty, které obsahují relevantní informace k danému tématu. Tento předpoklad lze částečně ignorovat v uzavřených systémech, ne však v kontextu globálního webu.<sup>29</sup>

### **.2.2.5 Nejednotné pojmenování**

Ačkoliv je přiřazení jednoznačných identifikátorů zdrojům a jejich jednotné používání jedním z cílů sémantického webu, je zřejmé, že toho

---

<sup>29</sup> ALLEMANG, Dean, et. al., Semantic web for the working ontoloigists

nelze dosáhnout v úplnosti. Producenti informací na sémantickém webu nebudou nutně koordinovat svoje činnosti, což bude mít, a v dnešní době již má, za následek to, že stejné zdroje jsou známé pod různými identifikátory. Tento problém, známý jako „Nonunique naming assumption“, je dalším aspektem webu, který musí sémantický web brát v potaz. Např. na úrovni OWL je to řešeno vlastností *sameAs*, která umožní označit nějakou třídu či vlastnost za ekvivalentní ke třídě jiné.

## 3 Technologie sémantického webu

### 3.1 RDF

RDF je jazyk pro reprezentaci informací o zdrojích na WWW. Je speciálně zamýšlen pro reprezentaci metadat o webových zdrojích, jako např. datum, titulek nebo datum aktualizace webové stránky, informace o autorských právech webového dokumentu apod. RDF je však díky zobecnění pojmu „webový zdroj“ možno použít pro prezentaci informací na webu pouze identifikovatelných (newebové zdroje), přestože tyto nejsou skrze web získatelné. Je to založeno na systémů globálně unikátních identifikátorů URI, zejména jejich speciální typ URL.

Tyto entity se dělí následovně:

- webové zdroje - věci, které na webu existují, tudíž mají již přidělené URL
- newebové zdroj - věci, které na webu přímo neexistují
  - fyzické objekty
  - pojmy (koncepty)
  - vlastnosti

Doporučovaná praxe je i pro entity, které na webu neexistují, poskytnout

při získání jejich URL jejich popis.

RDF kromě poskytování modelu pro data slouží i jako jazyk pro tvorbu ontologií na sémantickém webu používaných.

### **.3.1.1 RDF trojice**

RDF má velice jednoduchou a flexibilní strukturu, postavenou na centrálním konceptu RDF výroku, který se skládá z trojice členů. RDF trojice jsou základní datová struktura, kterou RDF jako model sémantického webu poskytuje. Jsou založeny na myšlence, že popisované zdroje mají vlastnosti a ty mají hodnoty<sup>30</sup>.

Jedná se o podobný koncept jako způsob uložení dat v relačních databázích. V databázích jsou data reprezentována pomocí tabulek skládajících se ze sloupců a řádků. Tabulka představuje jednu třídu. Řádky reprezentují instance a každý sloupec reprezentuje jednu vlastnost instance daného řádku. RDF trojice je ekvivalentem buňky v tabulce.<sup>31</sup>

Jedna trojice může vyjadřovat jednu vlastnost zdroje, zápis v přirozeném jazyce by vypadal např. takto:

Webový dokument `http://www.w3.org/Proposal.html` má titulek „WorldWideWeb: Proposal for a HyperText Project“.

Jednotlivé prvky trojice jsou tyto:

1. subjekt (`http://www.w3.org/Proposal.html`)
2. predikát („má titulek“)
3. objekt („WorldWideWeb: Proposal for a HyperText Project“)

RDF má za cíl umožnit vytvářet strojově pochopitelné výroky. Pro splnění

---

<sup>30</sup> Heslo Resource Description Framework ve Wikipedii

<sup>31</sup> ALLEMANG, Dean, et. al., Semantic web for the working ontoloigists

tohoto cíle jsou zapotřebí dvě věci<sup>32</sup>.

1. Systém strojově zpracovatelných identifikátorů pro identifikování subjektu, predikátu a objektu bez možnosti záměny s podobně vypadajícím identifikátorem, který by mohl být již použit jinde na webu.
2. Strojově zpracovatelný jazyk pro reprezentaci těchto výroků a jejich výměnu mezi stroji.

Subjekt je vždy zdroj identifikovaný pomocí URI. Predikát je jeho konkrétní vlastnost, která má rovněž přiřazený svůj URI. V roli objektu může stát buď webový zdroj nebo literál. Literál je textový řetězec, který reprezentuje datové hodnoty. Textový řetězec nemůže stát v roli subjektu ani vlastnosti.

V roli objektu je v příkladu výše webový dokument umístěný na URL <http://www.w3.org/Proposal.html>, predikátem je výraz „má titulek“ a objektem je řetězec „WorldWideWeb: Proposal for a HyperText Project“. Vlastnost „má titulek“ musí být v RDF také identifikovaná pomocí URI, musí tedy vycházet z nějaké ontologie. V tomto by bylo vhodné použít např. ontologii vyvinutou v rámci Dublin Core Metadata Initiative, která bude představena později.

### **.3.1.2 Identifikátory**

Jako identifikátory architektura webu poskytuje URL. URL je globálně unikátní, proto nemůže dojít ke záměně s jinými URL. Kromě toho umožňuje přístup k webovému zdroji tím, že definuje jeho umístění na síti. Pro potřeby RDF je důležité, aby bylo možné identifikovat i zdroje, které nejsou síťově dostupné. K tomuto účelu slouží obecnější systém identifikátorů URI. Rozdíl mezi URL a URI je v tom, že URL kromě identifikace umožňuje i přístup ke zdroji. Vzhledem k širšímu pojetí pojmu zdroj jako cokoliv, co je potřeba uvést v nějakém výroku, používá RDF speciální typ URI, a to URI reference. URI reference, označovaná také

---

<sup>32</sup> W3C, RDF Primer

jako URIref, je URI doplněná a volitelný fragment, který je od URI oddělen znakem „#“.

Důležité je vymezit rozdíl mezi zdrojem, který na webu není dostupný a webovým dokumentem, který o něm pojednává. Např. autor tohoto textu má domovskou stránku dostupnou na URL:

```
http://jakub.roztocil.name/
```

Tato URL je vhodná pro identifikaci dokumentu, který se na ní nachází (tedy webového zdroje), je ale nevhodná pro identifikaci entity, o které pojednává (newebového zdroje). Toto je jeden z problémů, který URI reference řeší. Příklad URIref, která identifikuje autora tohoto textu přidáním fragmentu:

```
http://jakub.roztocil.name/#me
```

Takto je pomocí RDF možné identifikovat prakticky cokoliv.

### **.3.1.3 Serializace RDF**

RDF je určené pro stroje, aby jim umožnilo pochopit to, co je lidem zřejmé. Specifikace RDF ho definuje jako abstraktní model a lze ho vyjádřit pomocí několika formátů z nichž nejvýznamnějším je RDF/XML, které je založeno na XML<sup>33</sup>.

XML je obecný a rozšiřitelný značkovací jazyk a bylo vytvořeno za účelem umožnit komukoliv vytvořit svůj formát pro dokumenty a poté pomocí něj tyto dokumenty publikovat. V XML jsou jednotlivé části dokumentů uzavřeny do značek, což umožňuje programům porozumět jejich obsahu a správně ho interpretovat.

XML poskytuje mechanismus jmenných prostorů, které umožňují pomocí URI vymezit oblast jmen pro elementy použité v dokumentu. Elementy tak mohou nést stejné názvy, ale pokud se nacházejí v různých jmenných

---

33 W3C, RDF Primer



prostředích, tak nedochází ke kolizi či záměně. Tato vlastnost je pro potřeby sémantického webu velice důležitá, protože umožňuje kombinovat data z různých zdrojů.

Jednoduchost výroků, které lze pomocí RDF tvořit a použití globálních identifikátorů umožňuje RDF vyjádřit také formou grafu. Graf obsahuje šipkami propojené uzly, které vyjadřují zdroje, vlastnosti a jejich hodnoty. Kombinování výroků z více zdrojů se graf rozrůstá a podává více informací o zdrojích<sup>34</sup>.



Obrázek 1: ukázka RDF grafu popisující Erica Millera (zdroj: W3C, RDF primer)

Graf ilustruje použití URI k identifikaci entit:

<sup>34</sup> W3C, RDF Primer

- „Věci“ - v tomto případě Eric Miller, identifikován jako <http://www.w3.org/People/EM/contact#me>
- Druhy věcí - v tomto případě je to Person (osoba), identifikovaná jako <http://www.w3.org/2000/10/swap/pim/contact#Person>
- Vlastnosti věcí, např. elektronická adresa, identifikovaná jako <http://www.w3.org/2000/10/swap/pim/contact#mailbox>
- Hodnoty vlastností, např. pro emailovou adresu je to URI <mailto:em@w3.org> a pro vlastnost jméno je to řetězec „Eric Miller“

Tato množina výroků se pomocí RDF/XML vyjádří takto:

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:contact="http://www.w3.org/2000/10/swap/pim/contact#">
  <contact:Person rdf:about="http://www.w3.org/People/EM/contact#me">
    <contact:fullName>Eric Miller</contact:fullName>
    <contact:mailbox rdf:resource="mailto:em@w3.org"/>
    <contact:personalTitle>Dr.</contact:personalTitle>
  </contact:Person>
</rdf:RDF>
```

O

### *brázek 2: RDF/XML reprezentace množin výroků*

Kromě RDF/XML, který je preferovaným způsobem serializace RDF v prostředí webu, existují i další formáty, které umožňují reprezentovat RDF ve strojově čitelné podobě a které nemají na XML založenou syntaxi. Jsou tím pádem ve své syntaxi mnohem jednodušší, proto jsou vhodnější pro ukázky a nadále v této práci budou příklady uváděny pomocí nich. Jedním z těchto formátů je Notation3. V té by se ten samý RDF graf vyjádřil takto:

```

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix contact: <http://www.w3.org/2000/10/swap/pim/contact#> .

<http://www.w3.org/People/EM/contact#me> a contact:Person;
  contact:fullName "Eric Miller";
  contact:mailbox <mailto:em@w3.org>;
  contact:personalTitle "Dr." .

```

Obrázek 3: Stejná množina výroků jako na předchozím obrázku vyjádřená pomocí syntaxe Notation3. Pro převod z RDF/XML byl použit program CWM ("Closed World Machine").

Z obou příkladů je patrné, že ačkoliv subjekt, predikát a někdy také objekt musí být identifikovány pomocí URI, není nutné plnou URI vždy vypisovat. Jak v případě RDF/XML, tak v případě Notation3 je možné prefixu URI dát jméno a pomocí něj pak na prefix odkazovat. Prefix se s lokálním jménem spojí pomocí dvojtečky, čímž vznikne kvalifikované jméno (Qname). V RDF/XML je toto dosaženo definicí prefixu - jmenného prostoru - pomocí atributu `xmlns:názevPrefixu="URI-jmenného-prostoru"`, čímž je daný URI prefix dostupný pro element nesoucí tento atribut a všechny jeho potomky. V Notation3 se prefixy definují jejich deklarací za použití klíčového slova `@prefix`.

RDF/XML obsahuje sadu elementů, které umožňují reprezentaci RDF výroků.

- `<rdf:RDF>` - kořenový element RDF elementu. Obvykle má atribut definující základní XML jmenný prostor pro RDF dokumenty, kterým je URI `http://www.w3.org/1999/02/22-rdf-syntax-ns#`. Tento prostor je k dispozici pod prefixem `rdf`. Dále obsahuje definice jmenných prostorů použitých ontologií.
- `<rdf:Description>` - reprezentuje jeden zdroj, který je identifikován pomocí jeho URI uvedeném v atributu `rdf:about`. Obsahem elementu jsou vlastnosti zdroje a jejich hodnoty. Názvy

vnořených elementů, které se nenachází ve jmenném prostoru RDF, tak definují libovolné množství libovolných vlastností a jejich hodnot definovaných připojenými ontologiemi, do jejichž jmenného prostoru spadají. Hodnotou je buď v elementu uzavřený textový řetězec, nebo zdroj, který je identifikován pomocí URI uvedeném jako hodnota atributu `rdf:resource`. Hodnoty vlastností mohou být též vyjádřeny pomocí atributů elementu `<rdf:Description>`. V tom případě mohou obsahovat pouze hodnoty v podobě řetězců.

- `<rdf:Bag>` - může být přímým potomkem elementů vlastností uvnitř `<rdf:Description>` a jeho obsahem je výčet více hodnot jedné vlastnosti uzavřených v elementech `<rdf:li>`.
- `<rdf:Seq>` - pro tento element platí to samé, jako pro element `<rdf:Bag>`, s tím rozdílem, že při jeho použití je pořadí hodnot vlastností důležité.
- `<rdf:Alt>` - má stejné umístění a stejný obsah jako `<rdf:Bag>` a `<rdf:Seq>`, ale výčet hodnot vlastnosti je považován za seznam alternativních variant, z nichž je uživatelem vybírána pouze jediná.

#### **.3.1.4 Odvozování**

Odvozování (angl. Inferencing) je proces, při kterém jsou nové trojice systémově přidávány do grafu na základě vzorů v explicitně vyjádřených výrocích.<sup>35</sup> Jedná se o koncept podobný např. hierarchii tříd v objektově orientovaném programování. Instance jedné třídy je zároveň instancí třídy nadřazené. Odvozování se provádí na základě pravidel definovaných ontologií.

#### **.3.1.5 RDFa**

Dnešní web je postaven tak, aby byl užitečný lidem. Přesto, že se na webu již objevují strojově čitelná data, tak ta jsou typicky distribuována

---

<sup>35</sup> ALLEMANG, Dean, et. al. Semantic web for the working ontologists

odděleně od HTML stránek, které ta samá data reprezentují ve formě určené pro lidi. Mezi verzí pro lidi a pro stroje je velmi slabé spojení, takže ve výsledku je prohlížeč odkázán na verzi určenou pro lidi, která poskytuje informace pouze o vizuální prezentaci dat.<sup>36</sup>

RDFa (Resource Description Framework in Attributes) má za úkol tento problém řešit. Specifikace RDFa<sup>37</sup> definuje sadu XHTML atributů které umožňují obohatit prezentační data o strojově čitelné doplňky. RDFa umožňuje do HTML vložit jednoduchá i komplexní RDF data, která učiní data a odkazy povodně určené lidem čitelné i pro stroje bez nutnosti jejich opakování v externích souborech. Tímto se snižuje redundance dat v systémech zpřístupňující jak data určená lidem, tak i strojům.

### **.3.2 Ontologie**

Slovo ontologie pochází z řeckého *ontos* (jsoucno) a *logos* (slovo) a znamená dvě věci:<sup>38</sup>

1. Větev metafyziky zabývající se podstatou a vztahy jsoucna.
2. Teorii o charakteru jsoucna a druhů existence.

Jedna z definic ontologie, která je v souvislosti se sémantickým webem často zmiňována, zní: „Ontologie je formální, explicitní specifikace sdílené konceptualizace“<sup>39</sup>.

*Konceptualizace* zde znamená abstraktní model; *explicitní* znamená, že musí být jasně definována; a *formální* indikuje, že specifikace musí být strojově zpracovatelná. Na základě této definice vznikla definice konkrétnější, která zní: „Ontologie je reprezentace znalostí z určité oblasti zájmů, kde objekty a jejich vztahy jsou popsány slovníkem termínů.“<sup>40</sup>

---

36 W3C, RDFa Primer

37 W3C, *RDFa in XHTML: Syntax and Processing*

38 Heslo Ontology v Merriam-Webster : Online Dictionary

39 GRUBER, A Translation Approach to Portable Ontology Specifications

40 BREITMAN, et. al., Semantic Web : Concepts, Technologies and Applications

Další z definic zní: „Ontologie je studie o kategoriích věcí, které existují nebo mohou existovat v nějaké doméně zájmu. Produkt této studie, nazývaný též ontologie, je katalogem typů věcí, o kterých je předpoklad, že existují v doméně zájmu D, z perspektivy osoby, který používá jazyk L pro účely komunikace o D“<sup>41</sup>.

Sebraná kolekce několika definic termínu ontologie, které se objevila v literatuře, vypadá takto<sup>42</sup>:

- Ontologie je filozofická disciplína.
- Ontologie je formální konceptualizace systému.
- Ontologie je formální sémantický popis.
- Ontologie je specifikace konceptualizace.
- Ontologie je reprezentace konceptuálního systému za použití logických teorií.
- Ontologie je slovník termínů používaný v nějaké logické teorii.

Ontologie se používají v různých oblastech počítačové a informační vědy, jako např.:

- umělá inteligence
- reprezentace znalostí
- zpracování přirozeného jazyka
- softwarové inženýrství

To je jedním důvodů pro neexistenci zcela univerzální definice tohoto termínu. Jedna z dalších definic termínu ontologie používaná v oblasti sémantického webu vzešla z World Wide Web konzorcia a zní:

---

41 SOWA, Principles of ontology

42 BREITMAN, et. al., Semantic Web : Concepts, Technologies and Applications

„Ontologie je termín převzatý z filozofie a poukazuje na vědu popisování druhů entit ve světě a to, jak spolu tyto entity souvisí.“<sup>43</sup>

Tato organizace dále navrhuje, aby ontologie používané na sémantickém webu poskytovaly popisy následujících elementů:

- Třídy nebo „věci“ z různých domén zájmů.
- Vztahy mezi entitami.
- Vlastnosti, které mohou entity mít.

Většina existujících jazyků pro reprezentaci ontologií jsou s touto definicí konzistentní<sup>44</sup>. Jde zejména o RDF Schema a OWL, které budou podrobněji popsány později.

Panuje obecná shoda, že nezávisle na těchto definicích je nutné porozumět tomu, že ontologie se používají k popisům velice různorodých modelů. Od jednoduchých taxonomií typu katalog s kategoriemi až po velice sofistikované modely vyjádřené pomocí konstruktů logik prvního řádu.<sup>45</sup>

### **.3.3 RDF Schema**

RDF samo o sobě tvoří datový model pomocí kterého reprezentuje data, která lze vyjádřit pomocí grafu. RDF schema (RDFS) poskytuje způsob, jak tento graf použít strukturovaným způsobem. Slouží k popisu konceptů použitých v RDF. RDF schema samo je vytvořeno pomocí RDF.<sup>46</sup>

Podobně jako jiné jazyky pro tvorbu schémat, také RDF schema hovoří o datech, která jsou v systému reprezentována. Klíčová myšlenka RDF schématu je v tom, že by mělo poskytnout informace o významu dat, a používá proto koncept odvozování, který umožňuje dozvědět se o datech a

---

43 BREITMAN, et. al., Semantic Web : Concepts, Technologies and Applications

44 tamtéž

45 tamtéž

46 ALLEMANG, Dean, et. al., Semantic web for the working ontoloigists

jejich významu více, než je v nich explicitně zaneseno.

Významnou vlastností RDFS je, že definuje doménu (*domain*) a rozsah (*range*) pro predikáty v RDF trojicích. Toto má spíše než funkci validační funkci odvozovací. Doména určuje, jakého typu je zdroj stojící v roli objektu RDF trojice. Rozsah naopak definuje typ zdroje stojícího v roli hodnoty vlastnosti. Pomocí tohoto mechanismu je možné, je-li využít, z jedné trojice zjistit, jaké třídy jsou subjekt a objekt, které jsou ve vztahu definovaném danou vlastností.

### **.3.3.1 Třídy**

RDFS poskytuje konstrukt `rdfs:Class`, jehož pomocí lze definovat třídy. Např. zjednodušená definice třídy `foaf:Person` pomocí RDFS vypadá takto:

```
foaf:Person rdf:type rdfs:Class ;  
  rdfs:label "Person" ;  
  rdfs:comment "A person." .
```

Je patrné, že definice třídy je úzce provázaná se samotným RDF. Využívá jeho vlastnost `rdf:type` pro přiřazení typu `rdfs:Class` objektu v trojici. Instance třídy jsou provázány s třídou opět pomocí `rdf:type`:

```
<http://jakub.roztocil.name/#me> rdf:type foaf:Person .
```

Hierarchie tříd se určuje pomocí `rdfs:subClassOf`. Následující RDF trojice určuje, že každá instance třídy `foaf:Person` je také instancí nadřazené třídy `foaf:Agent`:

```
foaf:Person rdfs:subClassOf foaf:Agent .
```

### **.3.3.2 Vlastnosti**

RDFS rozšiřuje možnosti RDF pro definování vlastností. Pomocí konstrukt `rdfs:subPropertyOf` se vyjadřuje hierarchie vlastností. Vlastnost `rdfs:subPropertyOf` je instancí třídy `rdf:Property`.



Je-li vlastnost P podřazenou vlastností P', pak jsou všechny páry zdrojů, které jsou spojeny vlastností P', rovněž spojeny vlastností P.

Vztah vlastností ke třídám se definuje pomocí vlastností `rdfs:domain` a `rdfs:range`.

Pomocí vlastnosti `rdfs:domain` se ke třídě vlastnosti přiřadí jedna nebo více tříd, jejichž instance se mohou objevit v roli subjektu při použití dané vlastnosti jako predikátu v RDF trojici. Vlastnost `rdfs:range` plní podobou roli s tím rozdílem, že pro RDF trojice, kde je vlastnost použita, deklaruje typy zdrojů v roli objektu. Např. v ontologii FOAF nalezneme vlastnost `rdfs:workplaceHomepage`, která má doménu a rozsah definován následovně:

```
rdfs:workplaceHomepage
  rdfs:domain foaf:Person;
  rdfs:range foaf:Document .
```

Z toho vyplývá, že v RDF trojicích, které jsou spojeny predikátem `rdfs:workplaceHomepage`, stojí v roli subjektu instance třídy `rdfs:domain foaf:Person` a v roli objektu instance třídy `foaf:Document`.

RDFS je základní jazyk pro tvorbu ontologií pro prostředí sémantického webu. Disponuje poměrně omezenou schopností vyjadřovat složitější vztahy a definovat omezení, což je důvod pro existenci dalšího jazyka rozšiřující možnosti RDF schma - OWL.

### **.3.4 OWL**

OWL (Web Ontology Language<sup>47</sup>) je rodina jazyků, která rozšiřuje možnosti RDFS popisovat třídy a vlastnosti. Bylo vytvořeno za účelem reprezentace bohatých znalostí o entitách, skupinách entit a vztazích mezi entitami. OWL bylo vytvořeno organizací W3C. Existují tři varianty

---

<sup>47</sup> Pořadí písmen je autory úmyslně prohozeno, aby tak tvořila lépe vyslovitelnou zkratku, *owl* taktéž v angličtině znamená *sova* a implikuje moudrost, kterou má OWL pomoci zprostředkovávat.

OWL:

1. OWL Full
2. OWL DL
3. OWL Lite

Přičemž platí<sup>48</sup>:

- Každá platná OWL Lite ontologie je zároveň platnou OWL DL ontologií.
- Každá platná OWL DL ontologie je zároveň platnou OWL Full ontologií.
- Každý platný OWL Lite závěr je zároveň platným OWL DL závěrem.
- Každý platný OWL DL závěr je zároveň platným OWL Full závěrem.

### **.3.4.1 OWL Full**

Je založeno na sémantice OWL DL a OWL Lite. Bylo navrženo tak, aby zachovávalo kompatibilitu s RDFS. Např. s třídami lze nakládat jak jako s kolekcí instancí, tak jako s instancemi samotnými. OWL Full umožňuje rozšířit významy RDF výrazů. Považuje se za nepravděpodobné, že by nějaký software byl schopný plně zpracovávat veškerá odvozování podle OWL Full.<sup>49</sup>

### **.3.4.2 OWL DL**

OWL DL je založeno na deskriptivní logice. Jeho cílem je poskytnout co nejvyšší možnou vyjadřovací schopnost při zachování garance, že veškeré odvozování je možné strojově zpracovat v konečném čase. OWL DL obsahuje všechny konstrukty z OWL, ale ty mohou být použity pouze s určitými omezeními, aby bylo možno zachovat zpracovatelnost

---

<sup>48</sup>Heslo Web Ontology Language ve Wikipedii

<sup>49</sup>tamtéž

odvozování.<sup>50</sup>

### **.3.4.3 OWL Lite**

OWL Lite bylo původně zamýšleno pouze pro tvorbu klasifikačních hierarchií s možností definovat jednoduchá omezení. Jedná se o podskupinu komplexnějšího OWL DL a taktéž garantuje možnost vypočítat veškeré odvozování.

## **.3.5 Existující ontologie**

### **.3.5.1 Dublin Core**

Ontologie Dublin Core definuje sadu elementů metadat sloužící pro popis informačních zdrojů. Dublin Core je široce používané k popisu digitálních materiálů jako video, zvuk, obraz, text a složená média jako webové stránky. Dublin Core je ISO standardem a kromě použití jako ontologie pro RDF data má i další využití.<sup>51</sup>

Sémantika Dublin Core byla vytvořena a je udržována mezinárodní interdisciplinární skupinou profesionálů z oblastí jako knihovnictví, počítačová věda, muzejnictví a z příbuzných oblastí.

Slovo „core“ v názvu indikuje, že se jedná pouze o základní sadu elementů která může být rozšířena na základě potřeb systému, ve kterém je Dublin Core používáno.

Dublin Core existuje ve dvou verzích. Dublin Core Simple definuje sadu 15 základních elementů. Každý z nich je volitelný a může se opakovat s různými hodnotami vícekrát. Qualified Dublin Core těchto 15 elementů dále rozšiřuje. Základní elementy definované Dublin Core jsou tyto<sup>52</sup>:

1. Title – jméno zdroje.

---

<sup>50</sup> tamtéž

<sup>51</sup> Heslo Dublin Core ve Wikipedii

<sup>52</sup> DCMI, Soubor metadatových prvků Dublin Core

2. Creator - entita zodpovědná za tvorbu obsahu.
3. Subject - předmět dokumentu.
4. Description - popis obsahu zdroje.
5. Publisher - entita zodpovědná za zpřístupnění zdroje.
6. Contributor - entita, která přispěla k obsahu zdroje.
7. Date - datum události, která se vyskytla během existence zdroje..
8. Type - povaha nebo žánr (druh) obsahu zdroje.
9. Format - fyzické nebo digitální provedení zdroje.
10. Identifier - jednoznačný odkaz na zdroj v rámci daného kontextu.
11. Source - odkaz na zdroj, ze kterého je popisovaný zdroj odvozen.
12. Language - jazyk intelektuálního obsahu zdroje.
13. Relation - odkaz na příbuzný zdroj.
14. Coverage - rozsah nebo záběr obsahu zdroje.
15. Rights - informace o právech vztahujících se k popisovanému zdroji.

Dublin Core je na sémantickém webu často využívaná ontologie. Obvykle se používá v kombinaci ještě s dalšími ontologiemi reflektující specifika daného systému.

### **.3.5.2 SKOS**

SKOS (Simple Knowledge Organization Systems) je rodina formálních jazyků vytvořená za účelem reprezentace tezaurů, klasifikačních schémat, taxonomií, systémů předmětových hesel a různých strukturovaných řízených slovníků. Hlavním cílem je umožnit snadnou publikaci řízených

slovníků pro prostředí sémantického webu.<sup>53</sup> SKOS je vyvíjen organizací W3C a stále se nachází pouze ve stádiu *Candidate Recommendation* (kandidát na doporučení).

Užitečným prvkem formátu SKOS je možnost vyjádřit shodu nebo podobnost mezi pojmy (koncepty) různých znalostních systémů. Například pomocí elementu `<skos:exactMatch>` lze odkázat na ekvivalentní pojem v jiném znalostním systému, na nějž se lze jednoznačně odkazovat s použitím URI. Takže s pomocí těchto elementů můžeme v hesle "internet" odkázat na ekvivalentní (nebo blízce podobné) heslo "Internet" z Library of Congress Subject Headings, protože toto heslo je jednoznačně identifikováno pomocí URI. Vazby na další pojmy mohou být vytvářeny na všechny znalostní systémy, které jsou k dispozici ve formě prolinkovaných dat (např. DBpedia, Open Calais nebo Freebase).

### **.3.5.3 FOAF**

FOAF (Friend of a Friend) je projekt, který se zabývá tvorbou ontologie pro popisy lidí, vztahy mezi nimi a tím, čím se zabývají. Ontologie FOAF umožňuje např. tvorbu profilů, kde jsou uvedeny informace o osobě, kontaktních údajích, zájmech, přátelích a dalších charakteristikách. Např. zjednodušený profil vytvořený za pomoci RDF a FOAF v syntaxi Notation3 pro autora tohoto textu vypadá následovně:

```
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
<http://jakub.roztocil.name/#me> rdf:type foaf:Person ;
    foaf:surname "Roztočil" ;
    foaf:firstName "Jakub" ;
    foaf:title "Mr" ;
    foaf:gender "male" ;
    foaf:mbox <mailto:jakub@roztocil.name> ;
    foaf:jabberID "jakub@roztocil.name" ;
```

---

<sup>53</sup>Heslo Simple Knowledge Organization System na Wikipedii

```
foaf:homepage <http://jakub.roztocil.name/> ;
foaf:openid <http://openid.webkitchen.cz/> ;
foaf:workplaceHomepage <http://www.webkitchen.cz/> ;
foaf:schoolHomepage <http://uisk.ff.cuni.cz/> .
foaf:knows <http://data.semanticweb.org/person/jan-
zemanek> .
```

V příkladu lze vidět použití odkazů jak na webové zdroje (foaf:homepage, foaf:schoolHomepage) tak i na zdroje newebové (foaf:knows). Takto strukturovaná data o osobách mají mnoho praktických využití, mezi něž může patřit:

- Aplikace typu správce kontaktů mohou na jejich základě automaticky vytvořit seznam kontaktů.
- Jakákoliv registrace může být zjednodušena poskytnutím potřebných kontaktních a dalších údajů osob ve strojově čitelné podobě.
- Sociální sítě typu Facebook mohou na základě FOAF profilu nalézt přátele, kteří jsou v síti registrováni a nabídnout tak uživateli automatické přidání mezi přátele v rámci dané sociální sítě.

RDF data využívající ontologii FOAF jsou např. indexována vyhledávačem Google, který třetí stranám poskytuje bezplatné API (application programming interface) pro práci s nimi. Google Social Graph API umožňuje RDF data popsaná pomocí FOAF využívat pro účely popsané výše.

### **.3.6 Linked data**

Linked data je označení pro sadu doporučení a pravidel, které mají za cíl propojit jednotlivé báze dat publikované na webu jako RDF. Berners-Lee definuje<sup>54</sup> čtyři předpoklady, které jsou zapotřebí dodržet, aby mohl web růst. Tyto předpoklady platí pro web dokumentů a také pro sémantický

---

<sup>54</sup>BERNERS-LEE, Linked data

web.

1. Použití URI jako jména pro věci.
2. Použití HTTP URL, aby lidé mohli získat informace věcech.
3. Pokud někdo navštíví danou URL, tak poskytnout užitečné informace.
4. Zahrnout odkazy na další zdroje, aby mohli být objeveny další informace.

Dodržení těchto předpokladů není nezbytně nutné ke zpřístupnění provázaných dat, ale jejich nedodržení znemožňuje jejich znovupoužití neočekávanými způsoby, což je považováno za přidanou hodnotu webu.<sup>55</sup> Jednotlivé předpoklady budou dále detailněji rozebrány.

### **.3.6.1 Použití URI jako jmen pro věci**

Použití URI jako jmen pro věci je základním předpokladem pro přispívání do sémantického webu. Pokud nejsou používány globální identifikátory URI, nejedná se o sémantický web.

### **.3.6.2 Použití HTTP URI bylo možné získat informace o věcech**

System HTTP URI umožňuje podle jména zdroje získat informace i o věcech, které na webu neexistují. Je tak možné získat popis newebového zdroje identifikovaného pomocí dané URI přesto, že se s ní uživatel setká v kontextu, kdy neví, k čemu se váže. Není vhodné používat jiná schémata než HTTP, jako například DOI nebo URN. HTTP adresování používá systém doménových jmen, který díky standardům a národních organizacím umožňuje delegovat autoritu a zaručuje nezávislost na jediné organizaci.<sup>56</sup>

### **.3.6.3 Poskytovat užitečné informace při navštívení URI**

---

<sup>55</sup> tamtéž

<sup>56</sup> tamtéž

Praxe poskytnout užitečné informace o zdrojích a vlastnostech identifikovaných pomocí URI je přinejmenším u velkých ontologií obvykle dodržovaná.<sup>57</sup> Důvodem pro toto doporučení je to, aby bylo možné získat informace popisující newebové zdroje na základě jejich URI a to jak ve strojově, tak lidem čitelném formátu.

#### **.3.6.4 Zahrnout odkazy na další věci**

Tato praxe je na webu běžná a umožňuje prozkoumávání souvislostí. V případě RDF dat toto umožňuje rozšířit graf tvořený z více zdrojů a jejich vztahů. Hodnota informací je do určité míry dána také tím, jakým způsobem odkazují na další relevantní zdroje.<sup>58</sup>

#### **.3.6.5 Způsoby provázání dat na sémantickém webu**

Následující příklad ukazuje možnost, jak učinit data provázanými. V RDF souboru na adrese <http://example.org/smith> jsou uvedeny tyto výroky:

```
<#albert> fam:child <#brian>, <#carol>.
```

Trojice v tomto souboru podávají informace o osobě Albert, který má dvě děti - Briana a Carol. Albert je identifikován pomocí fragmentu `#albert`, což vzhledem k umístění souboru expanduje do plného URI identifikátoru <http://example.org/smith#albert>.

Díky tomu může kdokoliv ve svých RDF datech za pomoci této URI na Alberta odkázat a poskytnout o něm další informace. Může to provést např. následujícím způsobem:

V souboru <http://example.org/jones> jsou uvedeny následující trojice:

```
<#denise> fam:child <#edwin>,  
<http://example.org/smith#carol>.
```

Tímto je vyjádřeno, že osoba jménem Denise má syna Edwina a dceru

---

<sup>57</sup>BERNERS-LEE, Linked data

<sup>58</sup>tamtéž



Carol. Carol je ve výroku identifikována pomocí absolutní URI. Je tak možné po odstranění fragmentu začínajícího znakem # a získáním souboru na dané URI zjistit další informace, v tomto případě, že jejím otcem je Albert.

Je-li URI jménem pro věc, která není na webu dostupná, pak by při jejím získání mělo dojít k přesměrování za pomoci mechanismu protokolu HTTP zvaném content-negotiation (vyjednávání o obsahu).

Klient serveru předá HTTP GET požadavek na URL identifikující danou entitu. Společně s informací o požadované URL předává v HTTP hlavičce Accept také informaci o požadovaném MIME typu pro obsah. Je-li agentem prohlížeč sémantického webu, pak HTTP hlavička obsahuje informaci v této podobě:

```
Accept: application/rdf+xml
```

Jedná-li se o klasický webový prohlížeč nebo jiný program nezpracovávající RDF, pak posílá požadavek na HTML stránku jako:

```
Accept: text/html
```

Na základě této informace server určí, ve kterém formátu informace o newebovém zdroji klientovi vrátí. URL, která je identifikátorem newebového zdroje, by neměla přímo obsahovat žádný obsah, namísto toho by server měl pomocí HTTP hlavičky přesměrovat na URL, která reprezentuje danou věc v příslušném formátu. Tímto se zamezí případné záměně newebového zdroje se zdrojem webovým, který o newebovém zdroji poskytuje informace.

Doporučovaný způsob pro reprezentaci newebových zdrojů je poskytnutí celkem tří URL:

1. URL identifikující zdroj, které slouží jako jeho jméno.
2. HTML verzi obsahující popis newebového zdroje.

3. RDF verzi obsahující trojice definující vlastnosti zdroje a vazby na další webové či newebové zdroje.

Existuje několik variant, jak navrhnout architekturu pro URL systému, který poskytuje identifikátory a zpřístupňuje informace o newebových zdrojích, jedním z nich je odlišení pomocí části cesty k souboru na serveru:

1. Identifikátor zdroje:

`http://dbpedia.org/resource/Berlin`

2. HTML stránka s informacemi pro lidi:

`http://dbpedia.org/page/Berlin`

3. RDF reprezentace určená pro stroje:

`http://dbpedia.org/data/Berlin`

Příklad komunikace mezi klientem a serverem v případě, že je klientem RDF agent vyžadující informace o Berlínu z DBpedia:

1. Klient posílá HTTP požadavek na URL

`http://dbpedia.org/resource/Berlin` v této podobě:

```
GET /resource/Berlin HTTP/1.1
```

```
Host: dbpedia.org
```

```
Accept: text/html;q=0.5, application/rdf+xml
```

Požadavek obsahuje jak MIME typ HTML, tak RDF, ale HTML má v tomto případě nižší prioritu vyjádřenou modifikátorem `q=0.5`.

2. Server vrací HTTP odpověď obsahující informaci o skutečném umístění RDF reprezentace zdroje pomocí následujících HTTP hlaviček:

```
HTTP/1.1 303 See Other
```

```
Location: http://dbpedia.org/data/Berlin
```

```
Vary: Accept
```

3. Klient na základě odpovědi posílá nový požadavek na URL specifikovanou v hlavičce Location.
4. Sever vrací reprezentaci zdroje v RDF.

Tímto způsobem je umožněno jak lidem tak strojům získat dostupné informace o entitě podle její URL.

### **.3.7 Existující otevřené báze provázaných dat**

Sémantický web se pomalu stává realitou. Dnes již uzavřená pracovní skupina *Semantic Web Education and Outreach (SWEO) Interest Group* založená organizací W3C, registrovala již v roce 2007 přes dvě miliardy RDF trojic propojených více než dvěma miliony odkazů.<sup>59</sup> Linked Data Community<sup>60</sup> je projekt, který shromažďuje a udržuje informace o zdrojích provázaných RDF dat. Linking Open Data je stále pokračující projekt, který má za cíl podporovat sdílení provázaných dat publikovaných jako RDF.

---

<sup>59</sup>Wikipedia, heslo Linked Data

<sup>60</sup>Linked Data community. Linked Data : Connect Distributed Data across the Web

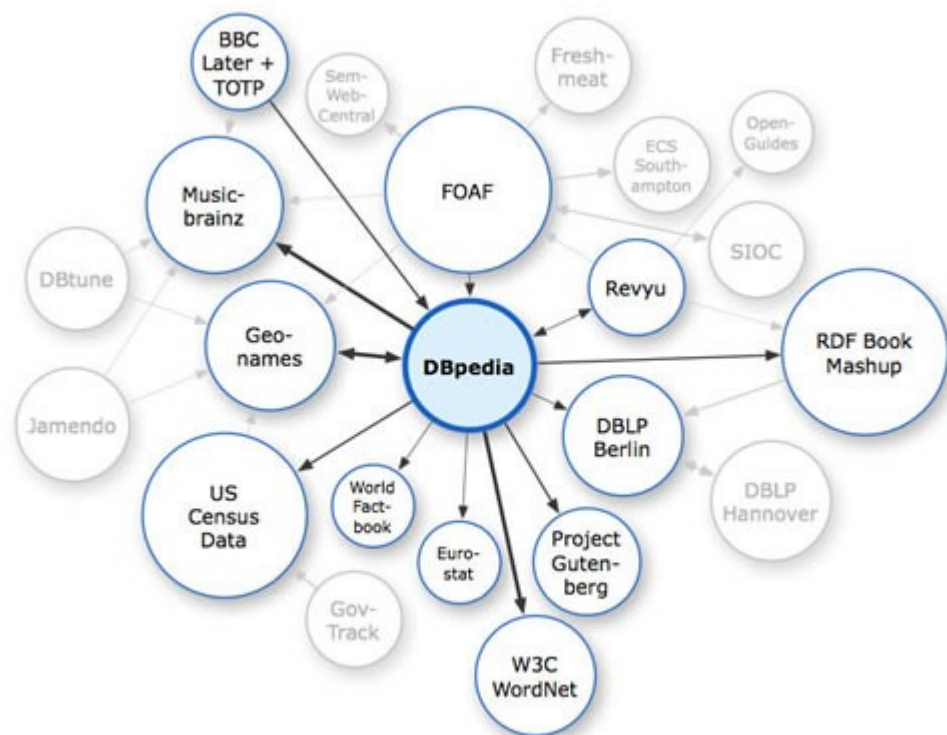


- 213.000 osob
- 328.000 míst
- 57.000 hudebních alb
- 36.000 filmů
- 20.000 společností

Bázi celkově tvoří 274 milionů RDF trojic, které zahrnují:

- popisy entit ve více než 30 jazycích
- 609.000 vazeb na obrázky
- 3.150.000 odkazů na externí webové zdroje
- 4.878.100 odkazů na data v jiných RDF bázích

Data z DBpedia jsou kromě Wikipedie provázána s dalšími bázemi RDF dat, mezi něž patří například Freebase, OpenCyc, UMBEL, GeoNames, Musicbrainz, CIA World Fact Book, DBLP, Project Gutenberg, DBtune, Jamendo, Eurostat a další.



Entitám v DBpedii je přiřazeno 415.000 kategorií z Wikipedie. Struktura dat je z velké části založena na údajích z postranních infoboxů, které lze na Wikipedii často nalézt.

Díky způsobu organizace dat jako RDF trojice je možné se skrze DBpedii Wikipedie dotazovat podobně, jako kdyby data byla uložena v relační databázi. Není tedy nutné se omezovat na vyhledávání založené na klíčových slovech.

Kromě již zmíněných funkcí plní DBpedia funkci neoficiální autority pro zdroj identifikátorů. Existuje zřetelná snaha producentů RDF dat znovu používat identifikátory pro entity z DBpedie, pokud jim v rámci ní již jednou byly přiděleny. Díky tomu do výše vyobrazeného schématu stále přibývají nové položky a globální síť sémantického webu se rozšiřuje.

Přístup k datům DBpedie je možný třemi způsoby:

1. Rozhraní pro kladení dotazů pomocí selekčního jazyka pro

sémantický web, jímž je SPARQL.

2. Linked Data rozhraní pro prohlížeče sémantického webu, které umožňují následováním odkazů bázi dat procházet.
3. Kompletní obsah DBpedie lze stáhnout ve formátech RDF/XML nebo CSV.

### **.3.7.2 GeoNames**

Databáze GeoNames obsahuje přes 8 milionů geografických jmen pro 6,5 milionům unikátních geografických míst. Vše je organizováno do 9 kategorií jména mají přiřazen 645 kódů dále určujících jejich charakter. Místa mají též přiděleny souřadnice pomocí WGS84 systému (World Geodetic System 1984).<sup>62</sup>

Každá entita v databázi GeoNames je reprezentována pomocí stabilního URI a vzhledem k tomu, že geografické entity jsou newebového charakteru, tak je pomocí HTTP content-negotiation poskytnut obsah podle požadovaného MIME typu.

GeoNames používá vlastní ontologii<sup>63</sup>, která popisuje vlastnosti zdrojů. Kategorizace je v RDF reprezentaci definována skrze SKOS. Dále jsou data provázána na relevantní zdroje v DBpedii.

Data jsou zpřístupňována také pomocí SPARQL koncového bodu.

### **.3.8 Prohlížeče sémantického webu**

Prohlížeč sémantického webu je speciální software, který umožňuje procházet sémantický web podobně, jako webové prohlížeče umožňují procházet web klasický. Je to však pouze jeden ze způsobů jak procházet sémantický web. Hlavní funkcionalita spočívá v možnosti zobrazit RDF data v lidem snadno čitelném formátu a následovat odkazy na další

---

<sup>62</sup> Heslo GeoNames ve Wikipedii

<sup>63</sup> WICK, Marc. GeoNames Ontology

zdroje. Některé prohlížeče sémantického webu jsou schopni též kombinovat data z více zdrojů. Mezi hlavní prohlížeče sémantického webu patří<sup>64</sup> The Tabulator – obecný prohlížeč dat a editor.

### **.3.8.1 The Tabulator**

The Tabulator je prohlížeč sémantického webu, jehož hlavní funkcí je zobrazení RDF data v podobě tabulky a také dalších formách zobrazení. Kromě zobrazování umožňuje také editaci dat a následné odeslání SPARQL požadavku na aktualizaci zpět na server. Dále je možné nad zobrazenými daty klást SPARQL dotazy. The Tabulator byl vytvořen organizací W3C za účelem demonstrace a otestování možných způsobů zobrazování RDF dat v prohlížečích sémantického webu a možností sémantického webu obecně. K dispozici jsou dva způsoby, jak jej používat:

1. Rozšíření (plug-in) pro prohlížeč klasického webu Mozilla Firefox.
2. Samostatná webová aplikace přístupná online.

Tabulátor ve verzi pro Firefox funguje tak, že po načtení dokumentu, který je pomocí HTTP hlaviček odeslaných serverem identifikován jako RDF/XML nebo Notation3, převede RDF data v něm obsažená do tabulkového zobrazení. Bez toho rozšíření Mozilla Firefox zobrazuje RDF/XML soubory ve formě vyobrazení stromové struktury XML dokumentu a v případě Notation3 obsahu zobrazí jako prostý text.

---

<sup>64</sup>Heslo Linked Data ve Wikipedii



Tento XML soubor nemá připojeny žádné informace o vzhledu prvků. Strom XML

```
- <rdf:RDF>
- <foaf:Person rdf:about="http://jakub.roztocil.name/#me">
  <foaf:name>Jakub Roztočil</foaf:name>
  <foaf:surname>Roztočil</foaf:surname>
  <foaf:family_name>Roztočil</foaf:family_name>
  <foaf:firstName>Jakub</foaf:firstName>
  <foaf:nick>Oswald</foaf:nick>
  <foaf:title>Mr</foaf:title>
  <foaf:gender>male</foaf:gender>
  <foaf:mbox rdf:resource="mailto:jakub@roztocil.name"/>
  <foaf:jabberID>jakub@roztocil.name</foaf:jabberID>
  <foaf:homepage rdf:resource="http://jakub.roztocil.name"/>
  <foaf:openid rdf:resource="http://openid.webkitchen.cz"/>
  <foaf:workplaceHomepage rdf:resource="http://www.webkitchen.cz/">
  <foaf:schoolHomepage rdf:resource="http://uisk.ff.cuni.cz"/>
```

▼ <http://www.webkitchen.cz/foaf.rdf>

Jakub Roztočil	masters	cz
		en
	type	Person
	family_name	Roztočil
	firstName	Jakub
	gender	male
	homepage	<a href="http://jakub.roztocil.name/">http://jakub.roztocil.name/</a>
	jabber ID	<a href="mailto:jakub@roztocil.name">jakub@roztocil.name</a>
	knows	Jan Zemánek
	personal mailbox	<a href="mailto:jakub@roztocil.name">mailto:jakub@roztocil.name</a>
	name	Jakub Roztočil
	nickname	Oswald
	openid	<a href="http://openid.webkitchen.cz/">http://openid.webkitchen.cz/</a>
	schoolHomepage	<a href="http://uisk.ff.cuni.cz/">http://uisk.ff.cuni.cz/</a>
	Surname	Roztočil
	title	Mr
	weblog	PUBLOG
	workplace homepage	<a href="http://www.webkitchen.cz/">http://www.webkitchen.cz/</a>
PUBLOG	type	Document
	name	PUBLOG
Jan Zemánek	type	Person
	name	Jan Zemánek

Find All

Pomocí ovládacích prvků lze přepínat mezi různými způsoby zobrazení dat.

### .3.9 SPARQL

SPARQL (SPARQL Protocol and RDF Query Language) je dotazovací jazyk pro RDF. Jedná se o klíčovou technologii sémantického webu. SPARQL byl navrhnut pracovní skupinou World Wide Web konsorcia DAWG (Data Access Working Group) a v lednu 2008 byla jeho specifikace vydána jako

doporučení.<sup>65</sup>

SPARQL svojí strukturou připomíná SQL (Structured Query Language), používaný pro dotazování relačních databází, a také používá některá stejná klíčová slova (např. SELECT, FROM, WHERE, ORDER BY, DISTINCT a další) ke stejnému účelu.

Na rozdíl od uzavřených systémů relačních databází platí pro sémantický web již zmíněný předpoklad otevřeného světa. Ten říká, že se nelze na základě nedostupnosti informace potvrzující pravdivost podmínky dojít k závěru, že podmínka není splněna.

Jednoduchý dotaz vyjádřený pomocí SPARQL vypadá takto:

```
SELECT ?person
WHERE { ?person foaf:name "John Smith" }
```

Přeloženo do přirozeného jazyka dotaz zní „vyber všechny osoby, které se jmenují John Smith“. Procesor jazyka SPARQL prohledá všechny dostupné RDF trojice, včetně trojic odvozených, a vrátí vazbu na zdroje, které odpovídají podmínce.

Kromě samotného dotazovacího jazyka existuje také protokol *SPARQL Protocol for RDF*<sup>66</sup>, který pomocí WSDL (Web Services Description Language) umožňuje vzdálené dotazování tzv. SPARQL koncových bodů, které zpřístupňují báze RDF dat.

## 4 Závěr

Množství neustále vznikajících standardů a doporučení signalizuje i neustálý živý vývoj sémantického webu jakožto rozšíření webu klasického. Vznikají také nové významné báze prolinkovaných RDF dat a již i velcí hráči na poli vyhledávacích a dalších online služeb začínají těžit z dostupnosti strukturovaných a provázaných dat určených pro stroje.

---

<sup>65</sup> W3C, SPARQL Query Language for RDF

<sup>66</sup> W3C, SPARQL Protocol for RDF

Sémantický web se tak pomalu vymaňuje z fáze, kdy producentům dat chybí motivace data poskytovat v RDF, protože nejsou k dispozici aplikace taková data vyžívající a naopak pro autory aplikací je množství dostupných dat motivací tato data využívat.

Tim Berners-Lee poté, co byl selekční jazyk pro sémantický web SPARQL vydán ve finální verzi jako standard, prohlásil, že všechny zásadní prvky architektury sémantického webu jsou již standardizovány a sémantický web je tak připraven i pro využívání v komerční sféře a dalších neakademických prostředích.

Vývoj sémantického webu bude mít i nadále pravděpodobně podobu spíše pozvolné evoluce než dramatické revoluce.

## Seznam použité literatury

1. ALLEMANG, Dean, HENDLER, Jim. *Semantic web for the working ontologists : Effective modeling in RDFS and OWL*. Burlington : Morgan Kaufmann, 2008. 330 s. ISBN 0-12-373556-4.
2. BERNERS-LEE, Tim, CAILLIAU, Robert. *WorldWideWeb : Proposal for a HyperText Project [online]*. 1999 [cit. 2009-05-09]. Dostupný z WWW: <<http://www.w3.org/Proposal.html>>.
3. BERNERS-LEE, Tim, HENDLER, James, LASSILA, Ora. *The Semantic Web. Scientific American [online]*. 2001 [cit. 2009-05-20]. Dostupný z WWW: <<http://www.scientificamerican.com/article.cfm?id=the-semantic-web>>.
4. BERNERS-LEE, Tim. *Information Management: A Proposal [online]*. March 1989, May 1990 [cit. 2009-05-09]. Dostupný z WWW: <<http://www.w3.org/History/1989/proposal.html>>.
5. BERNERS-LEE, Tim. *Linked Data [online]*. 2006 [cit. 2009-05-20]. Dostupný z WWW: <<http://www.w3.org/DesignIssues/LinkedData.html>>.
6. BERNERS-LEE, Tim. *Plenary at WWW Geneva 94 [online]*. 1994 [cit. 2009-05-20]. Dostupný z WWW: <<http://www.w3.org/Talks/WWW94Tim/Overview.html>>.
7. BIZER, Chris, CYGANIAK, Richard, HEATH, Tom. *How to Publish Linked Data on the Web [online]*. 2007- [cit. 2009-05-20]. Dostupný z WWW: <<http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/>>.
8. BIZER, Linked Data community. *Linked Data : Connect Distributed Data across the Web [online]*. 2007- [cit. 2009-05-20]. Dostupný z WWW: <<http://linkeddata.org/>>.
9. BOYE, Janus. *RDF : What's in it for us? [online]*. 1998 , Last Updated 15th March 2009 [cit. 2009-05-20]. Dostupný z WWW: <<http://www.irt.org/articles/js086/index.htm>>.
10. BREITMAN, Karin K., CASANOVA, Marco Antonio, TRUZKOWSKI, Walter. *Semantic Web : Concepts, Technologies and Applications*. London : Springer, 2007. 327 s. Nasa Monographs in System and Engineering Systems. ISBN 978-1-84628-581-3.

11. BRICKLEY, Dan, MILLER, Libby. *FOAF Project [online]*. 2000- [cit. 2009-05-31]. Dostupný z WWW: <<http://www.foaf-project.org/>>.
12. CERN. *CERN celebrates Web anniversary [online]*. 2003 [cit. 2009-05-09]. Dostupný z WWW: <<http://press.web.cern.ch/press/PressReleases/Releases2003/PR04.03EW eb10Years.html>>.
13. CERN. *Ten Years Public Domain for the Original Web Software [online]*. 2003 [cit. 2009-05-09]. Dostupný z WWW: <<http://tenyears-www.web.cern.ch/tenyears-www/Welcome.html>>.
14. DCMI. *Soubor metadatových prvků Dublin Core : referenční popis [online]*. verze 1.1. 2004 , Poslední aktualizace: 20. listopad 2006 [cit. 2009-05-31]. Dostupný z WWW: <[http://www.ics.muni.cz/dublin\\_core/elems.html](http://www.ics.muni.cz/dublin_core/elems.html)>.
15. *Dublin Core*. In *Wikipedia : The Free Encyclopedia [online]*. Wikimedia, 2007 [cit. 2009-05-20]. Dostupný na WWW: <[http://en.wikipedia.org/wiki/Dublin\\_core](http://en.wikipedia.org/wiki/Dublin_core)>.
16. Freie Universität Berlin. *DBpedia [online]*. [2008- ] [cit. 2009-05-31]. Dostupný z WWW: <<http://wiki.dbpedia.org/>>.
17. *GeoNames*. In *Wikipedia : The Free Encyclopedia [online]*. Wikimedia, 2007 [cit. 2009-05-20]. Dostupný na WWW: <<http://en.wikipedia.org/wiki/GeoNames>>.
18. GRUBER, T. R. *A Translation Approach to Portable Ontology Specifications [online]*. 1993 [cit. 2009-05-20]. Dostupný z WWW: <[http://ksl-web.stanford.edu/KSL\\_Abstracts/KSL-92-71.html](http://ksl-web.stanford.edu/KSL_Abstracts/KSL-92-71.html)>.
19. *Linked Data*. In *Wikipedia : The Free Encyclopedia [online]*. Wikimedia, 2007 [cit. 2009-05-20]. Dostupný na WWW: <[http://en.wikipedia.org/wiki/Linked\\_Data](http://en.wikipedia.org/wiki/Linked_Data)>.
20. *Ontology*. In *Merriam-Webster : Online Dictionary [online]*. Merriam-Webster, 1996- [cit. 2009-05-20]. Dostupný na WWW: <<http://www.merriam-webster.com/dictionary/Ontology>>.
21. *Resource Description Framework*. In *Wikipedia : The Free Encyclopedia [online]*. Wikimedia, 2007 [cit. 2009-05-20]. Dostupný na WWW: <[http://en.wikipedia.org/wiki/Resource\\_Description\\_Framework](http://en.wikipedia.org/wiki/Resource_Description_Framework)>.

22. Semantic Web Education and Outreach Interest Group. *Linking Open Data [online]*. [2008- ] [cit. 2009-05-31]. Dostupný z WWW: <<http://esw.w3.org/topic/SweoIG>>.
23. Semantic Web Education and Outreach Interest Group. *Linking Open Data [online]*. [2008- ] [cit. 2009-05-31]. Dostupný z WWW: <<http://esw.w3.org/topic/SweoIG>>.
24. *Semantic Web*. In Wikipedia : The Free Encyclopedia [online]. Wikimedia, 2007 [cit. 2009-05-20]. Dostupný na WWW: <[http://en.wikipedia.org/wiki/Semantic\\_Web](http://en.wikipedia.org/wiki/Semantic_Web)>.
25. *Simple Knowledge Organization System*. In Wikipedia : The Free Encyclopedia [online]. Wikimedia, 2007 [cit. 2009-05-20]. Dostupný na WWW: <<http://en.wikipedia.org/wiki/SKOS>>.
26. SOWA, J. F. *Principles of ontology [online]*. 1993 [cit. 2009-05-20]. Dostupný z WWW: <<http://ksl-web.stanford.edu/ontology/std/mailarchive/0136.html>>.
27. *Web Ontology Language*. In Wikipedia : The Free Encyclopedia [online]. Wikimedia, 2007 [cit. 2009-05-20]. Dostupný na WWW: <[http://en.wikipedia.org/wiki/Web\\_Ontology\\_Language](http://en.wikipedia.org/wiki/Web_Ontology_Language)>.
28. WICK, Marc. *GeoNames [online]*. 2004 [cit. 209-05-31]. Dostupný z WWW: <<http://www.geonames.org/>>.
29. WICK, Marc. *GeoNames Ontology [online]*. 2004 [cit. 209-05-31]. Dostupný z WWW: <<http://www.geonames.org/ontology/>>.
30. World Wide Web Consicium. *RDF Primer : W3C Recommendation [online]*. 1999 [cit. 2009-05-20]. Dostupný z WWW: <<http://www.w3.org/TR/rdf-primer/>>.
31. World Wide Web Consicium. *RDFa in XHTML: Syntax and Processing [online]*. 2008 [cit. 2009-05-20]. Dostupný z WWW: <<http://www.w3.org/TR/rdfa-syntax/>>.
32. World Wide Web Consicium. *RDFa Primer : W3C Working Group Note [online]*. 2008 [cit. 2009-05-20]. Dostupný z WWW: <<http://www.w3.org/TR/xhtml-rdfa-primer/>>.
33. World Wide Web Consicium. *Resource Description Framework (RDF) Model and Syntax Specification : W3C Recommendation [online]*. REC-

rdf-syntax-19990222. 1999 [cit. 2009-05-20]. Dostupný z WWW:  
<<http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>>.

34. World Wide Web Consicium. *SKOS Simple Knowledge Organization System [online]*. 2004 [cit. 2009-05-20]. Dostupný z WWW:  
<<http://www.w3.org/2004/02/skos/>>.

35. World Wide Web Consicium. *SPARQL Query Language for RDF : W3C Recommendation [online]*. 2008 [cit. 2009-05-20]. Dostupný z WWW:  
<<http://www.w3.org/TR/rdf-sparql-query/>>.

36. World Wide Web Consorcium. *OWL Web Ontology Language : Overview [online]*. 2004 [cit. 2009-05-20]. Dostupný z WWW:  
<<http://www.w3.org/TR/owl-features/>>.

37. World Wide Web Consorcium. *SPARQL Protocol for RDF : W3C Recommendation [online]*. 2008 [cit. 2009-05-20]. Dostupný z WWW:  
<<http://www.w3.org/TR/rdf-sparql-protocol/>>.

38. World Wide Web Consorcium. *The Tabulator [online]*. [2005] [cit. 2009-05-20]. Dostupný z WWW: <<http://www.w3.org/2005/ajar/tab>>.

39. World Wide Web Consortium. *Platform for Internet Content Selection (PICS) [online]*. 1997 [cit. 2009-05-20]. Dostupný z WWW:  
<<http://www.w3.org/PICS/>>.

40. World Wide Web Consortium. *W3C Semantic Web Activity [online]*. c1994- [cit. 2009-05-20]. Dostupný z WWW:  
<<http://www.w3.org/2001/sw/>>.

41. World Wide Web Consortium. *W3C Semantic Web Frequently Asked Questions [online]*. c1994- [cit. 2009-05-20]. Dostupný z WWW:  
<<http://www.w3.org/2001/sw/SW-FAQ>>.