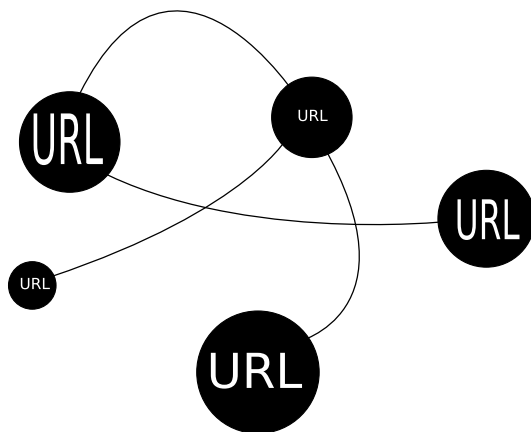


Úvod do sémantického webu

Jakub Vondrák

2. května 2007



Obsah

1	Úvod	3
1.1	Typografické a jazykové konvence	3
2	Teorie a pojmy	4
2.1	Proč klasifikovat?	4
2.2	Pojem koncept	4
2.3	Metadata	5
2.4	Řízený slovník	6
2.5	Taxonomie	7
2.6	Tezaurus	7
2.7	Folksonomie	8
2.8	Ontologie	8
2.9	Sémantika	9
3	Sémantický web	10
3.1	Co je na dnešním internetu špatně?	10
3.2	Sémantické HTML	11
4	Technologie sémantického webu	13
4.1	Dublin Core	13
4.2	RDF	15
4.3	RDF Schema	20
4.4	OWL	23
4.5	FOAF	25
4.6	WOT	27
4.7	SKOS	27
4.8	Další slovníky	27
4.9	SPARQL	28
4.10	RDFa	29
4.11	Mikroformáty	30
4.12	GRDDL	32

4.13 Syndikovaný obsah	33
4.13.1 RSS	34
5 Topic Maps	36
5.1 Topic	36
5.2 Association	38
5.3 Occurrence	39
6 Praktické použití	41
7 Pohled do budoucna	42

1 Úvod

Téměř každý už o sémantickém webu slyšel, ale pro mnohé je to pořád něco vzdáleného, složitého nebo nepoužitelného. Tato publikace si klade za cíl tyto omyly vyvrátit a srozumitelně nastínit aktuální dění kolem sémantického webu. Kromě několika teoretických záletů a vysvětlení situace, která vedla k samotné potřebě sémantického webu, bude stručně přiblíženo třináct technologií, souvisejících s touto problematikou. Velký počet technologií má na svědomí hlavně jejich vzájemná provázanost. Kniha vám tedy poskytne zhuštěný přehled a funguje jako odrazový můstek k dalšímu samostudiu.

U čtenáře se předpokládá znalost HTML, XML a mírná orientace v problematice internetu a programování. Kniha je určena především pro vývojáře webových aplikací se zájmem o budoucí vývoj internetových technologií a dále pak samozřejmě komunikoli, komu nedá neustálé povídání kolem sémantického webu spát.

1.1 Typografické a jazykové konvence

Sémantický web i jeho terminologie je mladá, a proto pro některé výrazy není zažitý nebo ani vytvořený český ekvivalent. Pokud ale takový ekvivalent existovat bude, budu ho v textu používat a při jeho prvním výskytu uvedu do závorky jeho anglický originál. Pokud naopak český výraz existovat nebude nebo nebude dostatečně zaběhnutý, budu v textu používat anglický výraz a při jeho prvním použití uvedu do závorky český překlad.

V celém textu bude pro World Wide Web Consortium používána zkratka w3.org. Adresy URL www stránek budou v textu zkráceny o část udávající protokol HTTP a budou vysazeny **neproporcionálním písmem**. Sousloví „sémantický web“ může být ve výjimečných případech nahrazeno zkratkou s. w.

2 Teorie a pojmy

2.1 Proč klasifikovat?

V informačním věku ve kterém žijeme, je nám k dispozici obrovské množství informací. Čím více máme ale informací, tím je větší problém se v nich orientovat. Tento nedostatek řeší mnoho různých technologií a my si zde popíšeme ty, které souvisejí se sémantickým webem.

Představte si knihovnu, která obsahuje veškeré vědění, ale neexistuje způsob jak určitou informaci v knihovně najít bez toho, aniž by bylo nutné procházet jednu knihu po druhé. Neexistoval by rejstřík a knihy by byly rozmístěny v policích náhodně. Jen vy a nekonečná knihovna. Nejspíš byste v tak obrovské knihovně do konce života nenašli to, co hledáte.

Výše uvedená knihovna by se dala s úspěchem přenést na prostředí internetu. Internet je „divoké“ médium a prozatím všechny snahy o jeho „zaškatulkování“ a usměrnění v globálních měřítkách selhávají. Příčina leží v jeho různorodnosti a nespoehlivosti, dané jeho živelnou podstatou. Jakékoli třídění a řazení se proto zatím děje jen lokálně v rámci serverů nebo jednotlivých služeb. Například fulltextové vyhledávací servery nebo ručně připravované klasifikace podle rozličných kritérií. Je zřejmé, že bez těchto služeb by byl dnešní internet jako informační zdroj nepoužitelný.

2.2 Pojem koncept

Konceptem bude v tomto textu označován význam slova. Jedná se v podstatě o představu, kterou slovo v mysli vyvolává. Různá slova mohou vzvolávat stejnou představu. Například slova Praha, PHA, Prague jsou všechno slova zastupující jednu jedinou myšlenkovou představu. Tato myšlenka, představa nebo význam je

tedy nezávislá na svojí vnější reprezentaci v podobě mluvené či psané formy. V anglické literatuře je, pro zde prezentovaný význam, používán termín koncept (concept) taktéž. V češtině by ale tento význam mohl být klidně vyjádřen například slovním spojením „význam pojmu“.

V sémantickém webu je důležité, aby nejenom lidé, ale i počítače byly schopny za informacemi vnímat koncepty. Slova Praha a PHA by tedy pro počítač, měly být stejně jako pro člověka identické.

2.3 Metadata

Metadata jsou obecně označována jako informace o informacích. Metadata mají tu pěknou vlastnost, že jsou obsažena přímo v objektu (dokument, kniha, obrázek), který popisují. V případě knihy by metadata mohly být její autor, název, rok vydání, vydavatel, atp. Obrázek v počítači by mohl mít pro změnu uložena metadata o čase expozice, názvu, místu, popisek a mnoho dalších informací.

Další vlastností metadat je, že je obvykle vytváří autor dokumentu, obrázku, atd. Problém, který tím zde vzniká je způsoben lidským faktorem. Člověk si může vymýšlet, dělá chyby (místo „autor“ napíše „autr“), může být líný, nesvědomitý, jeden *koncept* označuje různými slovy, není neutrální atp. (Doctorow, 2001).

Zmíněné problémy vedly k tomu, že se metadaty dnes webové vyhledávače téměř neřídí. Na webu jsou tedy kvalitní metadata (prozatím) utopii. Nic ale nebrání tomu, abyste vy sami kvalitní metadata vytvářeli. Vřele doporučuji se možnostmi metadat zabývat zvláště v případě, že jste majitelem digitálního fotoaparátu. Ušetříte tím nervy nejenom sobě, ale hlavně ostatním.

2.4 Řízený slovník

Řízený slovník (Controlled Vocabulary) se skládá z ručně vytvořeného seznamu slov, kde každé slovo bylo vybráno tak, aby odpovídalo přesně jednomu *konceptu*. Nejvíc aplikací řízeného slovníku najdeme v knihovnictví. Různé aplikace a použití řízeného slovníku mají různé názvy a my si zde popíšeme tezaurus, taxonomie a facets (aspektové třídění).

Každý z nás už s řízeným slovníkem přišel do styku, když listoval žlutými stránkami. Pokud jste hledali například heslo „počítače“, našli jste poznámku, abyste se podívali na heslo „výpočetní technika“.

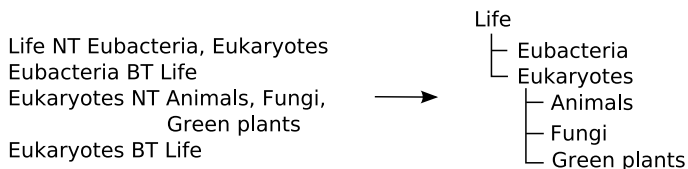
V řízeném slovníku lze mezi slovy, které obsahuje, vytvářet následující vazby:

- **USE** použijeme pro označení preferovaného synonyma: „synonymum“ USE „preferované synonymum“
- **UF** (Used For) pro označení synonyma k preferovanému synonymu: „preferované synonymum“ UF „synonymum“
- **NT** (Narrower Term) upřesňující výraz: „auto“ NT „sportovní auto“
- **BT** (Broader Term) zobecňující výraz: „sportovní auto“ BT „auto“
- **RT** (Related Term) související výraz: „auto“ RT „motor“
- **SN** (Scope Note) zpřesňující popis pojmu, aby se zamezilo jeho špatnému vyložení

Řízeného slovníku lze využít například i při konstrukci vyhledávače na webové stránce. Mezi databází a vstup uživatele vložíme řízený slovník, který k zadanému slovu vyhledá preferované synonymum obsažené v databázi (Leise et al., 2006).

2.5 Taxonomie

Pojem taxonomie je dnes často používán pro jakýkoli hierarchický klasifikační systém. Taxonomie je aplikací řízeného slovníku a vystačí si s BT a NT výrazy. Pojem taxonomie poprvé použil Carl Von Linné, který jím označil hierarchickou strukturu živých tvorů. S pojmem taxonomie se operuje často i v knihovnictví.



Obrázek 1: Ukázka taxonomie – data z tolweb.org

2.6 Tezaurus

Tezaurus je taktéž aplikací řízeného slovníku. Tezaurus je seznam slov, který ke každému slovu může obsahovat seznam jeho synonym, antonym, případně ještě zobecňujících a zužujících termínů a někdy také vysvětlení pojmy. Uvedené vztahy se zaznamenávají pomocí standardních zkratk řízeného slovníku. Následující příklad demonstruje sestavení tezauru.

- dopravní prostředek NT tramvaj, auto
- tramvaj UF šalina
- tramvaj BT dopravní prostředek
- tramvaj RT kolej

- šalina USE tramvaj
- auto UF auťák
- auto BT dopravní prostředek
- auťák USE auto
- kolej SN jezdí po ní tramvaje (není to VŠ kolej)

2.7 Folksonomie

Název *folksonomie* (folksonomy) vznikl složením anglických slov „folk“ a „taxonomy“. Folksonomie je klasifikační technika založená na práci široké komunity a v poslední době se velmi rozšířila. Funguje na principu označování vkládaného obsahu na server klíčovými slovy (tags) *jeho samotnými uživateli*. Pro vkládaný obsah tedy nejsou pevně připravené kategorie (v podobě taxonomie), ale obsah se kategorizuje podle klíčových slov. Příklad folksonomie lze vidět na známém serveru pro sdílení fotografií flickr.com nebo ve většině redakčních systémů – například v Drupalu. Výhody i nevýhody jsou podobné jako u metadat. V komunitách ale funguje tato klasifikace vcelku dobře, protože její členové mají zájem na kvalitních metadatech. Třeba proto, aby se jejich příspěvek dobře našel ostatními.

2.8 Ontologie

Pojem *ontologie* má v různých vědních disciplínách různé významy. V informatice znamená ontologie soubor *tříd*, jejich vzájemných *vazeb* a *atributů* z určité vyčleněné oblasti zájmu. Dnes používaná přesná definice zní: „*Ontologie je formální, explicitní specifikace sdílené konceptualizace*“. *Konceptualizace* je systém pojmů modelující část světa, který musí být specifikován *explicitně*, tj. né skryt v hlavě autora. *Konceptualizace* musí být

sdílená, tzn. že je výsledkem shody zájmové skupiny lidí. Informace o definici ontologie převzaty z (Svátek, 2004).

Zjednodušeně si lze *konceptualizaci* představit jako databázový model. Jednotlivé tabulky jsou třídy, sloupce atributy tříd a cizí klíče vazby. Data uložená v tabulkách jsou potom obsahem ontologie – její instancí. Taktéž je možné přirovnání k definicím tříd a jejich vazeb a atributů, známých z objektového programování. Ale pozor na přílišné zjednodušení. Ontologie může vyjádřit téměř jakýkoli vztah mezi svými prvky a nevystačila by si s možnostmi v databázovém nebo objektovém modelu nebo dokonce s BT a NT výrazy z řízeného slovníku.

Tvorba rozsáhlejší ontologie je komplikovaná záležitost, kterou je dobré nechat v rukou odborníků z oblastí znalostních systémů a umělé inteligence.

2.9 Sémantika

Sémantika je buď nauka o významu jazykových jednotek nebo teorie zabývající se vztahy mezi jazykovými znaky a objekty, k nimž se tyto znaky vztahují (Filipec et al., 2004). V souvislosti se sémantickým webem znamená sémantika význam slov, který je srozumitelný počítači.

3 Sémantický web

Sémantický web tak jak jej vidí w3.org, je web, na kterém budou data *zpracovatelná počítači*. Data na sémantickém webu ale půjdou nejenom strojově zpracovat, ale programově bude možné i odvozovat a automaticky vyhledávat další informace. To umožní zejména *sdílené slovníky a ontologie*. Specifikace sémantického webu se dnes skládá z RDF (Resource Description Framework) a OWL (Web Ontology Language) – obě technologie spadají pod křídla w3.org a poslední revize pochází z roku 2004. Kromě těchto dvou uvedených technologií, pracuje w3.org na dalších specifikacích, které se sémantickým webem souvisejí a o kterých se zmíním. Jsou jimi SPARQL, GRDDL, SKOS a RDFa.

Tři body vize sémantického webu podle w3.org:

- informace na webu mají přesně definovaný smysl
- informace na webu mohou být strojově zpracovány
- počítače budou schopny integrovat informační obsah webových stránek

3.1 Co je na dnešním internetu špatně?

Jedním z nedostatků je malé využití stávající struktury internetu. Člověk vložil nemalé úsilí, aby na internet umístil obrovské množství informací. Přesto jsou ale tyto informace v drtivě většině určeny pouze člověku a počítač je umí tak akorát zobrazit. Počítačové systémy se dnes většinou propojují na úrovni databází nebo jiných složitých technologií. Proč nevyužít toho, že na internetu už většinou tyto informace *jsou* obsažené? Stačí je jen zpřístupnit k počítačovému zpracování. Takový krok by otevřel nové možnosti integrace počítačových systémů.

Web je dnes složen z webových stránek, které představují dokumenty. Dokumenty, které jsou určeny člověku a člověk je taky jediný, kdo se v té záplavě dokumentů umí orientovat. Při hledání určité informace využívá člověk znalostí z oboru, aby správně pokládal vyhledávačům dotazy. Výsledky z vyhledávačů analyzuje a rozhodne se, jak bude postupovat dál. Pro počítač je něco takového dnes nemožné.

Člověk při hledání využívá kontextových informací a podle toho upravuje dotaz. Tyto informace mohou být například znalost autora, doba z které pochází článek, znalost hledané problematiky atd. Člověk totiž ve fulltextovém vyhledávači nevyhledává ve skutečnosti to, co ho zajímá, ale výrazy, o kterých se domnívá, že by se v hledaném textu mohly vyskytnout. Uživatel vyhledávače se vžívá do role autora hledaného textu a tak předvídá, co by v textu mohlo být za slova. Jen z jazykového hlediska je zde problém se synonymy, skloňováním, se zkratkami, se slangem a s různými jazykovými mutacemi. Pokud je výsledků příliš, jde na to člověk odjinud a počítá s tím, že se k výsledku „prokliká“. Je zřejmé, že pro počítač je to příliš silný oříšek, za který může příliš „divoká“ povaha internetu.

Někdy v budoucnu, až budou informace na webu dávat počítači smysl a bude nad nimi vybudována ontologie, mohly by sémantickým webem cestovat tzv. *agenti*, kteří by vyhledávali za člověka. Nejenom to, agent by mohl zjistit i další související informace a ty potom uživateli poskytnout.

3.2 Sémantické HTML

Sémantické HTML není sémantický web. Oblast zájmu sémantického HTML je jen a pouze HTML. Sémantické HTML je o používání HTML značek pro účel k jakému se mají používat. HTML značky mají totiž svůj význam. Existuje značka pro nadpis, odstavec atd.

Význam HTML elementů by se neměl násilně měnit kaskádovými styly (např. z nadpisu udělat běžný text). Také není dobré formátovat stránku jen s použitím `div` a `span` a spoustou vlastních tříd. Pokud pro něco existuje významová značka v HTML, měla by se použít. Pokud tak neučiníme, omezujeme možnost automatického zpracování struktury HTML programem (například automatická osnova dokumentu). Znevýhodnění budou také speciální prohlížeče pro nevidomé atd. Používání HTML elementů, podle jejich navrženého významu je tedy nanejvýš vhodné a je to základ pro přístupné stránky.

Sémantické HTML tedy není nic složitého. Stačí se seznámit s HTML elementy a správně je používat. Mezi sémantické elementy patří `<a>`, `<abbr>`, `<acronym>`, `<address>`, `<blockquote>`, `<cite>`, `<dt>`, `<dd>`, `<dfn>`, ``, `<h1>`, ..., `<h6>`, ``, `<label>`, `<legend>`, ``, `<p>`, `<q>`, `<samp>`, `<sub>`, `<sup>`, `<table>` (pro tabulku, ne pro layout), `<tbody>`, `<td>`, `<tfoot>`, `<th>`, `<thead>`, `<tr>`, ``, `<var>` a možná další.

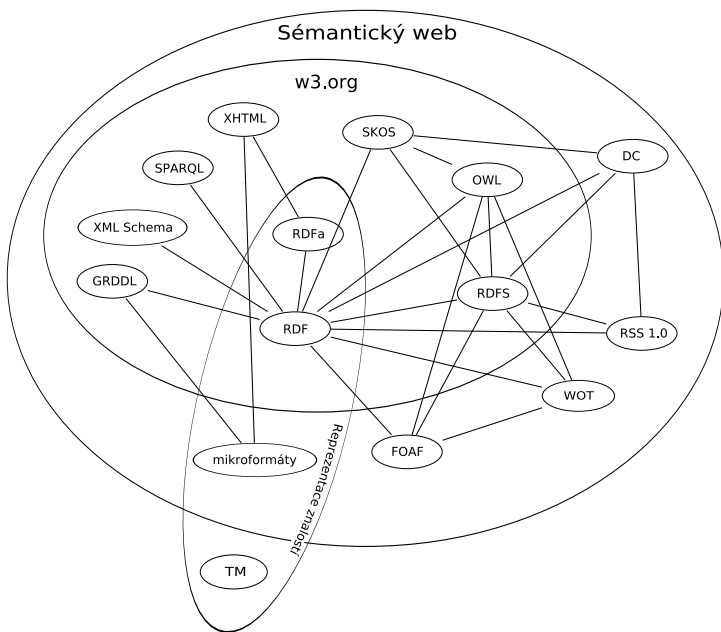
Mezi sémantické atributy patří: `caption`, `class`, `data`, `href`, `hreflang`, `id`, `lang`, `rel`, `rev`, `scope`, `summary` a `type`.

Sémantické a validní HTML začíná nabírat na důležitosti s příchodem mikroformátů a v budoucnu i RDFa. Validita stránek je dnes problém a mikroformáty se budou špatně dolovat z nevalidních struktury. Ani u XHTML neplatí to, co bylo při vzniku XHTML zásadní – XHTML je podmnožina XML a tudíž je vždy validní. Není. Většina XHTML dokumentů má chyby a tudíž je na tom stejně jako HTML. Možná i hůř, protože u HTML se takové chování od začátku předpokládalo.

Jako odpověď na tento problém přišlo od tvůrců mikroformátů doporučení POSH¹, které by autoři stránek pro jejich správnou funkci měli dodržovat.

¹microformats.org/wiki/posh

4 Technologie sémantického webu



Obrázek 2: Některé technologie s. w. a jejich souvislosti

4.1 Dublin Core

Dublin Core (dále jen DC) je soubor standardizovaných názvů popisek pro metadata. Základy pro DC byly položeny v Dublinu v roce 1995 a dodnes je soubor prvků rozvíjen. DC je dnes široce

využívané v RDF², HTML a dalších technologiích. Návrh DC byl totiž udělán nezávisle na konkrétním použití. Mezi velké servery, které Dublin Core využívají patří:

- www.ibm.com
- www.root.cz
- www.lupa.cz
- interval.cz

Následující ukázka je převzata přímo z HTML na serveru www.root.cz.

```
<meta name="DC.Identifier"  
      content="(SCHEME=ISSN) 12128309" />
```

Standardní jádro, tvoří 15 popisků: Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, Rights.

Dublin Core se ale neustále vyvíjí a obsahuje už i další prvky, které jsou ale specifikované mimo jádro. Další informace můžete najít na dublincore.org.

Pokud chcete na první pohled vidět, zda stránka obsahuje DC metadata a používáte Mozilla Firefox, stáhněte si rozšíření Dublin Core Viewer.

V roce 2005 proběhl výzkum³, ve kterém se sledovalo, jestli použití DC metadat zlepší umístění stránek ve známých vyhledávacích. Výsledek byl negativní.

²RDF slovník naleznete na adrese purl.org/dc/elements/1.1

³www.webology.ir/2005/v2n2/a13.html

4.2 RDF

Zkratka RDF znamená *Resource Description Framework* a jedná se o doporučení w3.org z roku 2004 určené pro standardizovanou výměnu dat v prostředí internetu. Jelikož je RDF určeno pro internet, respektuje jeho specifika jako je nespolehlivost a distribuovanost. Specifikace RDF zahrnuje jak abstraktní model, tak formáty pro jeho fyzické uložení, mezi které patří RDF/XML, N-Triples, Turtle a další.

Základním stavebním prvkem RDF jsou tzv. *triplety* – trojice. Triplet má za úkol vyjádřit o *informačním zdroji* nějakou informaci. Jeho části se nazývají *subjekt*, *predikát* a *objekt*. Jinými slovy také *předmět*, *vlastnost* a *hodnota* a toto označení bude dále používáno. Množina tripletů tvoří orientovaný graf, v kterém se mohou vyskytnout dva základní objekty: *uzly* (nodes) a *literály* (literals).

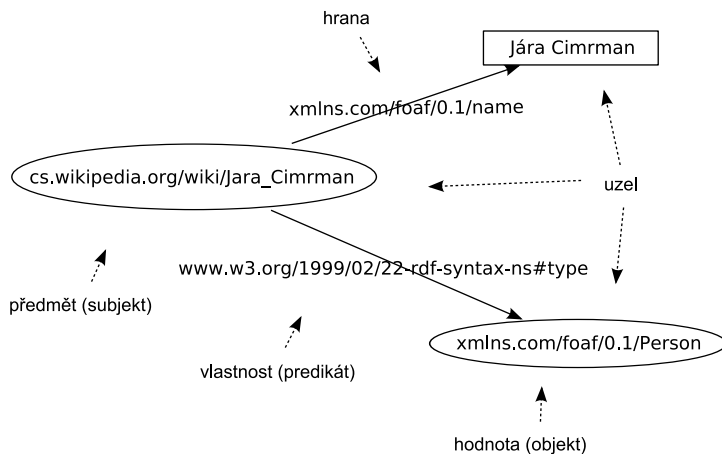
Přípustné hodnoty předmětu, vlastnosti i hodnoty jsou omezené. Předmět může být buď prázdný nebo URI⁴, vlastnost může být pouze URI a hodnota může být URI nebo literál.

Informaci o tom, že informační zdroj cs.wikipedia.org/wiki/Jara_Cimrman označuje člověka se jménem Jára Cimrman lze vyjádřit pomocí dvou tripletů (RDF graf je na obrázku 3). V zápisu N-Triples by triplety vypadaly následovně:

```
<http://cs.wikipedia.org/wiki/Jara_Cimrman>  
<http://xmlns.com/foaf/0.1/#firstName>  
"Jára Cimrman" .
```

```
<http://cs.wikipedia.org/wiki/Jara_Cimrman>  
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>  
<http://xmlns.com/foaf/0.1/Person> .
```

⁴URI (Universal Resource Identifier) je řetězec, který identifikuje či pojmenovává určitý informační zdroj. URL (Uniform Resource Locator) a URN (Uniform Resource Name) jsou jeho dvě podmnožiny.



Obrázek 3: RDF graf s dvěma triplety; zdroje znázorněny ovály, literály obdelníky

a v zápisu RDF/XML takto:

```
<rdf:RDF
  xmlns:rdf=
    "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/" >

  <rdf:Description rdf:about=
    "http://cs.wikipedia.org/wiki/Jara_Cimrman">
    <foaf:name>Jára Cimrman</foaf:name>
    <rdf:type
      rdf:resource="xmlns.com/foaf/0.1/Person" />
  </rdf:Description>

</rdf:RDF>
```

Jistě jste si všimli, že vlastnosti byly nahrazeny zvláštní URL adresou. To má svůj účel, protože pro označení jména i pro označení typu již existují předdefinovaná *slovíčka*. Tato slovíčka – tedy URL adresy jsou zapisována do *slovníků*, veřejně zpřístupňována a následně lépe či hůře veřejností přijímána.

Jestliže budeme pro označení vlastností, případně předmětů nebo hodnot popuzívat zaběhnutá slovíčka, tak zajistíme, že se data v RDF souborech mohou sloučit. Pokud by existoval kromě našeho příkladu na druhé straně internetu RDF soubor, kde by pro předmět bylo použito také URL `cs.wikipedia.org/wiki/Jara_Cimrman`, tak by se informace z obou dvou sloučily a dozvěděli bychom se o Járovi Cimrmanovi další zajímavé informace.

V následujícím výčtu uvádím atributy nebo elementy, na které bylo nebo bude možné narazit v příkladech při použití RDF/XML.

Základní atributy RDF/XML:

- `rdf:Description` je XML element pro označení uzlu
- `rdf:about` označuje informační zdroj – tedy předmět
- `rdf:ID` má podobný význam jako `id` v běžném XML. Jednoznačně identifikuje uzel; lze se na něj odvolávat pomocí `#hodnotaID`
- `rdf:nodeID` slouží pro identifikaci prázdného uzlu – viditelný pouze v rámci RDF souboru
- `rdf:resource` atribut udává hodnotu vlastnosti, vyjádřené jako informační zdroj
- `rdf:li` označuje položku ve výčtu hodnot v kontejneru

RDF nespécifikuje jen jak zacházet s triplety, ale taktéž definuje⁵ několik standardních slovíček, která lze v tripletech používat a dále uvádím několik důležitých:

- `rdf:type` označuje instanci třídy
- `rdf:Bag` třída označující množinu elementů (members), kde nezáleží na pořadí
- `rdf:Seq` třída označující seznam hodnot, kde záleží na pořadí
- `rdf:Alt` třída označující seznam hodnot alternativních elementů
- `rdf:Property` označuje slovo, které se smí použít jako vlastnost předmětu – používané v slovnících

Příklad použití třídy a instance jsme už viděli na obrázku 3. Jen připomenou, že podle konvencí se třídy píší s velkým počátečním písmenem a často se používá zkrácený formát zápisu, kdy se místo `rdf:Description` uvede přímo konkrétní třída. Náš příklad by pak vypadal následovně a je ekvivaletní k předchozí ukázce.

```
<rdf:RDF
  xmlns:rdf=
    "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/#" >

  <foaf:Person rdf:about=
    "http://cs.wikipedia.org/wiki/Jara_Cimrman">
    <foaf:name>Jára Cimrman</foaf:name>
```

⁵RDF Schema pro RDF: www.w3.org/1999/02/22-rdf-syntax-ns#

```
</rdf:Description>
```

```
</rdf:RDF>
```

V následujícím příkladu je vidět použití `rdf:Bag`. Příklad zachycuje seznam tvůrců postavy Járy Cimrmana. RDF graf je možné vidět na obrázku 4.

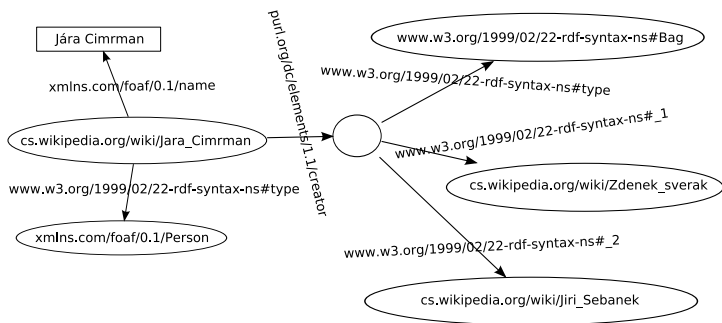
```
<rdf:RDF
  xmlns:rdf=
    "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/#"
  xmlns:dc="http://purl.org/dc/elements/1.1/# >

  <foaf:Person rdf:about=
    "http://cs.wikipedia.org/wiki/Jara_Cimrman">
    <foaf:name>Jára Cimrman</foaf:name>
    <dc:Creator>
      <rdf:Bag>
        <rdf:li rdf:resource=
          "http://cs.wikipedia.org/wiki/Zdenek_sverak"/>
        <rdf:li rdf:resource=
          "http://cs.wikipedia.org/wiki/Jiri_Sebanek"/>
      </rdf:Bag>
    </dc:Creator>
  </rdf:Description>

</rdf:RDF>
```

Na obrázku 4 je dále vidět použití prázdného uzlu.

Pokud budete chtít vaše RDF data umístit na internet, můžete tak učinit umístěním odkazu na RDF soubor z běžné HTML stránky a vyhledávače si jej najdou. Soubor by ale měly mít koncovku `.rdf` a MIME typ `application/rdf+xml`. V přípravách je



Obrázek 4: RDF graf znázorňující použití kontejneru `rdf:Bag`

ale technologie RDFa, o které ještě bude řeč a která má umožnit reprezentovat triplety v běžném XHTML. Tyto informace byly čerpány ze stránek www.w3.org/RDF – (Manola et al., 2004).

4.3 RDF Schema

V předchozí kapitole byla prezentována základní slovíčka (URL identifikátory), která RDF specifikace obsahuje. Taktéž byla v příkladech použita slovíčka o kterých nebylo nic řečeno. To bude napraveno v této a dalších kapitolách, věnovaných slovníkům. RDF Schema je základní prostředek pro definici slovníků. Přináší koncept tříd a dědičnosti a nabízí i slovní zásobu pro vyjádření omezujících podmínek na hodnoty vlastností a zaznamenání různých metadat o slovu ve slovníku. V dalším textu bude slovo „slovník“ používáno pro konkrétní aplikaci RDF Schema.

Slovník se vytváří, pokud existuje potřeba sdílet a vyměňovat data pomocí RDF s více lidmi. Slovníkem lze zajistit požadované slučování RDF dat.

RDF Schéma je obyčejné RDF, v kterém jsou ale dohodnu-

tým způsobem zaznamenány definice slov. Každé slovo má vymezené své použití podle toho, zda je to *třída* (class) nebo *vlastnost* (property). Třída může být hodnotou atributu `rdf:type` nějakého předmětu a tak se stane předmět instancí této třídy.

```
<rdf:type
  rdf:resource="xmlns.com/foaf/0.1/Person" />
```

Následujícím příklad ukazuje použití vlastností `dc:creator` a `dc:date`.

```
<rdf:Description rdf:about="w3:RDF">
  <dc:creator rdf:resource="w3:#" />
  <dc:date>2004</dc:date>
</rdf:Description>
```

U definice každého slova ve slovníku je nejdůležitější jeho adresa, která jej jednoznačně identifikuje. Poté následují další atributy jako komentář, rozsah platných hodnot v případě vlastnosti nebo v případě třídy její rodič atp. Dále uvádím podstatná slovíčka z RDF Schematu. Samotný slovník slovíček pro specifikaci RDF Schematu je vlastně RDF Schema.

- `rdfs:Class` označuje slovo, které se smí použít jako třída
- `rdfs:subClassOf` označuje předka naší třídy, předek musí být samozřejmě typu třída
- `rdfs:subPropertyOf` označuje předka vlastnosti
- `rdfs:comment` komentář předmětu – jakýkoli řetězec
- `rdfs:label` vlastnost označuje název předmětu v lidské, čitelné formě

- `rdfs:domain` domána hodnoty vlastnosti; například číslo, informační zdroj atp., doména musí být třída
- `rdfs:range` rozsah hodnoty vlastnosti - taktéž musí být vyjádřen třídou
- `rdfs:seeAlso` odkaz na další informační zdroj vztahující se k předmětu
- `rdfs:isDefinedBy` odkaz na RDF Schema, kde je tato třída nebo vlastnost definovaná
- `rdfs:Literal` třída představující literál
- `rdfs:Container` třída, od které musí dědit všechny kontejnery. Například `rdf:Bag`, `rdf:Seq` dědí od této třídy

Nejdůležitějším slovíčkem je `rdfs:Class`, které zavádí koncept tříd. Ukázka definice slova `rdf:Bag`, třídy, které známe z RDF:

```
<rdfs:Class
  rdf:about=
    "http://www.w3.org/1999/02/22-rdf-syntax-ns#Bag">

<rdfs:isDefinedBy
  rdf:resource=
    "http://www.w3.org/1999/02/22-rdf-syntax-ns#" />

<rdfs:label>Bag</rdfs:label>

<rdfs:comment>
  The class of unordered containers.
</rdfs:comment>
```

```
<rdfs:subClassOf
  rdf:resource=
    "http://w3.org/2000/01/rdf-schema#Container"/>
</rdfs:Class>
```

Informace byly čerpány z (Brickley et al., 2004)⁶, kde naleznete další informace.

4.4 OWL

Již jsme si něco řekli o RDF a RDF Schema, ale jejich slovní zásoba není pořád dostačující k tomu, aby bylo ve slovníku možné podchytit nejrůznější vztahy mezi slovy, jako například kardinalitu a všemožné typování. Jinými slovy nám chybí prostředky pro plnohodnotnou definici ontologie. Pro tento účel tedy vznikla specifikace OWL.

OWL, neboli Web Ontology Language, staví na svém předchůdci DAML + OIL a od roku 2004 je w3.org doporučením. OWL přidává slovní zásobu pro zachycení dalších vztahů mezi třídami, pro definici kardinality (rodič může mít několik dětí, dítě musí mít rodiče), rovnosti, neslučitelnosti (žena nemůže být muž), rozšířené typování vlastností a jejich charakteristik a umí vyjádřit výčet tříd (enum). Pokud bude pro jistou oblast definována ontologie, znamená to, že bude možné strojově odvozovat mnohá další fakta.

OWL nabízí tři slovníky s různě bohatou slovní zásobou. OWL Lite, OWL DL a OWL Full seřazeno od nejjednoduššího po nejobsáhlejší.

Definování plnohodnotné promyšlené ontologie není jednoduché a proto zde bude uvedeno jen pár příkladů z (Smith et al., 2004)⁷.

⁶www.w3.org/TR/rdf-schema

⁷w3.org/TR/owl-guide

OWL je opět obyčejné RDF, respektive RDF Schema. Zapišme tedy definováním jmenných prostorů:

```
<rdf:RDF
  xmlns="http://.../REC-owl-guide-20040210/wine#"
  xml:base="http://.../REC-owl-guide-20040210/wine#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:rdf="http://.../1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
```

Dále následuje hlavička ontologie:

```
<owl:Ontology rdf:about="">
  <rdfs:comment>
    An example OWL ontology
  </rdfs:comment>

  <owl:priorVersion
    rdf:resource="http://...20031215/wine"/>

  <owl:imports
    rdf:resource="http://...20040210/food"/>

  <rdfs:label>Wine Ontology</rdfs:label>
</owl:Ontology>
```

V ukázce hlavičky si všimněte prázdného řetězce v atributu `rdf:about`. Prázdný řetězec je možné zadat, když definujeme výchozí jmenný prostor. Dále je v hlavičce podstatný element `owl:imports`, s jehož pomocí lze importovat jiné ontologie a třeba je jen rozšířit pro vlastní specifické potřeby, což je velice podstatná vlastnost, který zajistí možnost tvorby distribuovaných ontologií. Hlavička slouží tedy pro definici metadat ontologie.

Pod hlavičkou už následuje samotné definování obsahu ontologie. Ten může obsahovat tři základní stavební prvky: třídy, vlastnosti a *individua*. Individuum je člen (instance) nějaké třídy.

```
<owl:Class rdf:ID="Winery"/>
<owl:Class rdf:ID="Region"/>
<owl:Class rdf:ID="ConsumableThing"/>
```

Zatím jsme nadefinovali pouze jména tříd a nic jiného nevíme. Pokud chceme definici dále rozšířit, odkážeme se jen na definici pomocí `rdf:resource` – můžeme tak učinit třeba v jiném souboru, kam výše uvedená jména tříd jen naimportujeme.

```
<owl:Class rdf:ID="PotableLiquid">
  <rdfs:subClassOf rdf:resource="#ConsumableThing" />
</owl:Class>
```

Hotová OWL ontologie, je stejně jako RDF Schema, slovník, který lze dále využívat. Narozdíl od RDF Schema ale ontologie může zachytit mnohem více informací. Ontologie mohou mít příponu `.owl` nebo `.rdf`.

Program, který umí pracovat s ontologií, tzv. resoner může odvozovat další fakta. Takovým programem je například open source nástroj SWOOP⁸ – Semantic Web Ontology Editor, který momentálně spravuje Google. Pomocí tohoto programu si můžete pohodlně z GUI prostředí sestavit vlastní ontologii nebo si prohlížet jiné.

4.5 FOAF

Friend of a Friend (FOAF) je slovník, přesněji ontologie, protože využívá některé OWL konstrukce. Cílem FOAF je dát k dispozici slovíčka, která lze použít při vytváření RDF zachycujícího

⁸lze stáhnout z code.google.com/p/swoop

informace o tom kdo koho zná. Využití proto nachází v komunitních projektech a je to zřejmě zatím nejrozšířenější aplikace sémantického webu a RDF. Google na hledaný výraz „foaf.rdf“ vrátil 1 380 000 výsledků. Pokud chcete i vy využít FOAF k zveřejnění informací o sobě, existuje už množství generátorů, které vám RDF vytvoří automaticky⁹.

Specifikace FOAF se nachází na adrese `xmlns.com/foaf/0.1` a ve slovníku lze najít například následující slovíčka: `foaf:Person`, `foaf:name`, `foaf:nick`, `foaf:knows`, atd. Význam jednotlivých slovíček je tedy myslím celkem zřejmý. Automaticky vygenerovaný RDF FOAF soubor může vypadat následovně:

```
<rdf:RDF
xmlns:rdf="http://.../1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://.../2000/01/rdf-schema#"
xmlns:foaf="http://xmlns.com/foaf/0.1/"
<foaf:Person rdf:ID="jan_novak">
<foaf:name>Jan Novak</foaf:name>
...
<foaf:knows>
  <foaf:Person>
    <foaf:name>Petr Novak</foaf:name>
    ...
  </foaf:Person>
</foaf:knows>
</foaf:Person>
</rdf:RDF>
```

Pokud chcete přidat FOAF soubor na svou domovskou stránku, můžete tak učinit v hlavičce HTML následovně:

```
<link rel="meta"
```

⁹FOAF generátor: www.ldodds.com/foaf/foaf-a-matic

```
type="application/rdf+xml"  
title="FOAF" href="foaf.rdf" />
```

4.6 WOT

Web Of Trust je ontologie – slovník¹⁰, zajišťující důvěryhodnost informací prezentovaných v RDF. Dnes se s jeho pomocí digitálně podepisují hlavně FOAF soubory¹¹, ale použít ho lze pro jakýkoli RDF soubor. Podepsání zajistíme přidáním následujících řádků do RDF a pak samozřejmě vytvoříme samotný podpis a distribuujeme ho spolu s RDF souborem.

```
<!-- digital signature for this file -->  
<rdf:Description rdf:about="">  
  <wot:assurance rdf:resource="foaf.rdf.asc" />  
</rdf:Description>
```

4.7 SKOS

SKOS - Simple Knowledge Organisation System je w3.org specifikace, která definuje slovník, jehož slovní zásoba je shodná s řízeným slovníkem a lze jím tedy vytvářet tezaury, taxonomie, folksonomie a ostatní aplikace řízeného slovníku. SKOS specifikace není zatím dokončená a její aktuální RDF Schema lze nalázt na webu w3.org¹².

4.8 Další slovníky

Postupem času budou vznikat další a další slovníky a distribuované ontologie. Než proto začnete tvořit vlastní, podívejte

¹⁰xmlns.com/wot/0.1/index.rdf

¹¹FOAF profil s digitálním podpisem: heddley.com/edd/foaf.rdf

¹²www.w3.org/2004/02/skos/core/history/2006-04-18.rdf

se kolem, jestli už práci někdo neudělal za vás. Můžete hledat například na stránkách www.schemaweb.info, kde je spravován seznam mnoha slovníků. Další možností je využít fulltextový vyhledávač RDF Schémat a ontologií Swoogle, který najdete na adrese swoogle.umbc.edu. V současné době obsahuje přes pět miliónů RDF dokumentů.

4.9 SPARQL

SPARQL Protocol and RDF Query Language je dotazovací jazyk pro RDF data navrhovaný w3.org. Specifikace SPARQL je zatím stále v přípravách, ale má nahradit několik dosavadních nestandardizovaných technologií. Dotaz ve SPARQL funguje na principu trojic. Pozice v trojici, které znáte zadefinujete a o ostatní necháte volné. Výsledkem budou trojice, které této podmínce vyhovují. Následující příklad je převzat od w3.org¹³.

RDF data:

```
<http://example.org/book/book1>  
<http://purl.org/dc/elements/1.1/title>  
"SPARQL Tutorial" .
```

SPARQL dotaz:

```
SELECT ?title  
WHERE  
{  
  <http://example.org/book/book1>  
  <http://purl.org/dc/elements/1.1/title>  
  ?title  
}
```

Výsledek: „SPARQL Tutorial“

¹³www.w3.org/TR/2007/WD-rdf-sparql-query-20070326

4.10 RDFa

Technologie RDFa w3.org umožňuje reprezentovat triplety v běžném XHTML. Konkrétně XHTML1.1, jehož specifikace není, stejně jako samotného RDFa, ještě dokončena.

Následující příklad je převzat z (Adida a Birbeck, 2007)¹⁴ a ukazuje, jak lze pomocí RDFa doplnit sémantické informace do běžného XHTML, v kterém je popsáno konání události.

```
<html>
  <head><title>Jo's Blog</title></head>
  <body>
    ...
    <p>
      I'm giving a talk at the XTech Conference about
      web widgets, on May 8th at 10am.
    </p>
    ...
  </body>
</html>
```

Po doplnění sémantiky pomocí RDFa a slovníku pro iCalendar by vypadalo XHTML následovně:

```
<html xmlns:cal="www.w3.org/2002/12/cal/ical#">
  <head><title>Jo's Blog</title></head>
  <body>
    ...
    <p class="cal:Vevent"
      about="#xtech_conference_talk">
      I'm giving <span property="cal:summary">a talk
      at the XTech Conference about web widgets</span>,
      on <span property="cal:dtstart">
```

¹⁴www.w3.org/TR/xhtml-rdfa-primer

```
        content="20070508T1000+0200">May 8th
    at 10am</span>.
</p>
...
</body>
</html>
```

Atributy `property`, `cal`, `content` jsou atributy specifické pro XHTML1.1, které ještě není dokončeno.

4.11 Mikroformáty

Mikroformáty mají přesně stejný cíl jako RDFa – reprezentovat data v XHTML. Pouze s tím rozdílem, že jsou použity stávající rozšířené technologie a to velice jednoduchým a přístupným způsobem. Za mikroformáty nestojí w3.org, ale nadšenci a bloggeři. Pro mikroformáty se vžilo označení evoluční revoluce.

Seznam existujících mikroformátů:

- **hCard** pro anotaci kontaktních dat, vychází z vCard
- **hCalendar** pro anotaci dat události, vychází z iCalendar standardu
- **rel-license** pro indikaci použité license
- **rel-nofollow** pro označení odkazu, který by neměl být hodnocen vyhledávači
- **rel-tag** pro označení odkazu, který je klíčovým slovem dané stránky
- **VoteLinks** pro označení naší podpory/nesouhlasu s odkazem, který musíme udělat

- **XFN** XHTML Friends Network pro vyjádření vztahu s jinými lidmi pomocí URL
- **XMDP** XHTML Meta Data Profiles pro definici profilu pro metadata
- **XOXO** eXtensible Open XHTML Outlines formát pro anotaci struktury

Následující mikroformáty jsou v přípravě: adr, geo, hAtom, hResume, hReview, rel-dictionary, rel-enclosure, rel-home, rel-payment, Robots exclusion a xFolk.

Uvedu zde příklad mikroformátu hCard. Nejdříve tedy původní VCard zápis:

```
BEGIN:VCARD
VERSION:3.0
N:elik;Tantek
FN:Tantek elik
URL:http://tantek.com/
ORG:Technorati
END:VCARD
```

a po přepsání do hCard mikroformátu:

```
<div class="vcard">
  <a class="url fn"
    href="http://tantek.com/">Tantek elik</a>
  <div class="org">Technorati</div>
</div>
```

Z příkladu je vidět, že přiřazování sémantiky se zde děje pomocí existujícího HTML atributu `class`.

Pro mikroformáty existuje rozšíření do Mozilla Firefox, které indikuje jejich přítomnost a na požádání data zobrazí. Nová

verze Mozilla Firefox 3 by v sobě měla mít podporu pro mikroformáty již integrovanou. Další informace a přesný popis mikroformátů lze nalézt na microformats.org.

4.12 GRDDL

Technologie GRDDL – Gleaning Resource Descriptions from Dialects of Languages umožňuje získat RDF data z XML/XHTML. Aby to bylo možné, je nutné připravit konverzní postup pro náš dokument. Nejčastěji se bude zřejmě jednat o XSLT. Na své XHTML stránce poté technologii GRDDL jen určeným způsobem řekneme, jaké XSLT má pro získání RDF dat použít. GRDDL nalezne využití například při získávání RDF dat z mikroformátů.

Pro vytvoření stránky, podporující GRDDL, jsou třeba tři kroky:

1. napsat XSLT transformaci
2. přidat do elementu `head` atribut `profile` s hodnotou `www.w3.org/2003/g/data-view`
3. specifikovat umístění souboru zajišťujícího transformaci

HTML poté může vypadat například takto:

```
...  
<head profile="http://www.w3.org/2003/g/data-view">  
  <title>Robin's Schedule</title>  
  <link rel="transformation"  
    href="http://www.w3.org/2002/12/cal/glean-hcal"/>  
</head>  
...
```

Ukázka převzata z (Davis a Talis, 2006)¹⁵, kde lze o GRDDL získat další informace.

4.13 Syndikovaný obsah

Syndikovaným obsahem označujeme informace, který server poskytuje ve strojově zpracovatelné podobě. V případě web feeds se jedná o jakési novinky, a proto této technologie využívají nejčastěji zpravodajské servery nebo blogy.

Obliba web feeds u uživatelů spočívá v rychlému a pohodlnému přístupu k novinkám z jednoho nebo více serverů najednou. A to vše bez reklam a bez zdlouhavého proklikávání se konkrétním webem. Programy a nástroje, které jsou schopny odebírat web feeds jsou nazývány agregátory. V našich podmínkách se lze často setkat také s názvem RSS čtečka, i když ten je zavádějící, protože RSS je jen jeden ze dvou dnes používaných protokolů pro web feeds. V angličtině se těmto programům říká většinou *feed reader* či *news reader*.

Syndikovaný obsah je dobrým příkladem toho, co je možné, jestliže umožníme informace strojově zpracovávat (a o tom je sémantický web). Čtečka syndikovaného obsahu může odebírat novinky ze stovek serverů, tyto novinky může třeba dále filtrovat podle různých kritérií a poté dát uživateli k dispozici jen to, co ho z celého internetu zajímá.

Dnes existují dva standardy pro uložení web feeds. Tím starším je RSS, které má spletitou minulost a novějším a standardizovaným formátem je ATOM. Dále se nicméně budeme zabývat pouze RSS, protože jedna z jeho verzí, konkrétně verze 1.0 je postavena na RDF. Verze RSS 1.0 je také první a poslední, která byla na RDF postavena. Ostatní verze 0.9 i 2.0 už jsou normální XML a s RDF nemají nic společného.

¹⁵www.w3.org/TR/grddl-primer

4.13.1 RSS

RSS je název XML formátu, kterého je využíváno pro distribuci syndikovaného obsahu. Server, který takto svůj obsah zpřístupní vytváří tzv. RSS kanál. Nejpoužívanější verze RSS protokolu je dnes pravděpodobně 0.9 a 2.0. RSS verze se od sebe dosti liší a vzájemně jsou nekompatibilní. Ve verzi RSS 1.0 je RSS aplikací RDF a má tedy svůj slovník¹⁶.

Ukázka RSS verze 1.0:

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf=
    "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns="http://purl.org/rss/1.0/" >
<channel
  rdf:about="http://www.xml.com/xml/news.rss">

  <title>XML.com</title>
  <description>
    XML.com features a rich mix of information and
    services for the XML community.
  </description>

  <items>
    <rdf:Seq>
      <rdf:li resource=
        "http://xml.com/pub/2000/08/09/xslt/xslt.html"/>
      </rdf:Seq>
    </items>
```

¹⁶purl.org/rss/1.0

```
</channel>

<item rdf:about=
  "http://xml.com/pub/2000/08/09/xslt/xslt.html">

  <title>Processing Inclusions with XSLT</title>

  <link>
    http://xml.com/pub/2000/08/09/xslt/xslt.html
  </link>

  <description>
    Processing document inclusions with general
    XML tools can be...
  </description>

</item>

</rdf:RDF>
```

5 Topic Maps

Do českého jazyka se název překládá jako mapa témat, ale v textu se bude objevovat zkratka TM, vytvořená z anglického názvu. Mapy témat svým určením v jistém ohledu konkurují RDF, a proto jsou v této knize zařazeny. Na rozdíl od RDF je jejich použití ale intuitivnější a technologie je navíc průmyslový standard, označený jako ISO/IEC 13250:2000.

Mapu témat je možné uložit do formátu XML nazývaného XTM, která vychází z datového modelu popsaného v (Garshol a Moore, 2006)¹⁷. Propagací a uplatňováním map témat do prostředí internetu se zabývá `topicmaps.org` konsorcium. Fakt, že dnes existují dvě technologie pro stejnou nebo podobnou věc byl zapříčiněn vzájemnou neinformovaností o druhé straně. Dnes se však vyvíjejí nástroje, které by měli dokázat formáty RDF a TM mezi sebou konvertovat.

V mapě témat, stejně jako v RDF, se klade velký důraz na efektivní slučování. Stejně jako v RDF, kde se jednotlivé části tripletu označují URL, existuje i v mapách témat mechanismus pro podchycení konceptu.

V základu se mapa témat skládá z tří fundamentálních částí. Z témat (topics), které představují koncepty, z vazeb (associations), spojující témata a z výskytů (occurrences), které představují adresovatelný informační zdroj nebo konstanty.

5.1 Topic

Téma je zástupcem (proxy) pro koncept, který v mapě témat chceme popisovat. Koncept je nazýván v mapě témat předmětem a může jím být naprosto cokoli. Snahou je, aby téma bylo na předmět navázané co nejpřesněji a nejuniverzálněji. Pro tento účel slouží jeho pojmenování, *identifikátory* a *lokátory*. Pokud

¹⁷www.isotopicmaps.org/sam/sam-model

je předmětem webový dokument, je vše jednoduché, protože přesný identifikátor předmětu, který je univerzální již existuje – URL adresa dokumentu. Tento identifikátor zároveň slouží i k nalezení předmětu a proto může být použit jako lokátor předmětu. V případě věcí reálného světa je třeba postupovat jinak.

Předměty reálného světa nemají svou WWW adresu, ale často mají adresu, kde se o nich píše. Takovou adresu můžeme použít pro identifikaci a nazýváme ji *indikátor předmětu*. Některé předměty nemusí mít své webové stránky, které by sloužily jako indikátor. To vůbec nevadí, adresy nemusí existovat. Skutečné URL adresy je ale výhodné používat, protože máme zajištěnou jejich univerzálnost.

Pokud chceme, aby se témata v různých TM správně slučovala, je potřeba zajistit, aby měly shodné některé parametry. Jedním takovým parametrem je právě indikátor. Jestliže jsou témata označena stejným indikátorem, sloučí se. Z toho důvodu vznikla potřeba nějakým způsobem zajistit, aby co nejvíce autorů map témat používalo stejné indikátory. K tomuto účelu se vztvářejí veřejně přístupné seznamy indikátorů (*PSI*, Published Subject Indicator), které se doporučuje používat¹⁸. PSI tedy mají stejný význam jako slovníky v RDF.

Pozorný čtenář si zde možná všiml prvního rozdílu mezi RDF a TM. V TM lze z principu okamžitě rozlišit, zda popisovaný koncept označuje webový dokument nebo předmět reálného světa. V prvním případě je totiž k identifikaci použit lokátor a v druhém indikátor.

Téma má následující základní atributy: jména (třeba v různých jazycích), výskyty (odkazy na informační zdroje nebo hodnoty vlastností), identifikátory, lokátory. Následující příklad uka-

¹⁸OASIS Technical Committee psi.oasis-open.org spravuje seznamy PSI pro označení zemí a jazyků

zuje definici tématu značícího ZČU.

```
<topic id="zcu">

  <baseName>
    <baseNameString>
      Západočeská Univerzita v Plzni
    </baseNameString>
  </baseName>

  <subjectIdentity>
    <subjectIndicatorRef
      xlink:href="http://www.zcu.cz" />
  </subjectIdentity>

</topic>
```

5.2 Association

Asociace znázorňuje vazbu mezi jedním nebo více tématy. Vazba může být určitého typu a každé téma, které je součástí vazby může mít přiřazenou roli, kterou ve vazbě „hraje“. Například vazba témat „Jára Cimrman“ a „Zdeněk svěrák“ může mít typ „stvořil“ a témata „Jára Cimrman“ a „Zdeněk Svěrák“ budou hrát role „Fiktivní postava“ respektive „Dramatik“. Vazby mohou být tedy na rozdíl od RDF mezi více než dvěma koncepty a platí obousměrně. Příklad ukazuje, jak by mohla vypadat vazba znázorňující autory knihy.

```
<association id="a133599">

  <instanceOf>
    <topicRef xlink:href="#kniha-autor" />
  </instanceOf>
```

```
<member>
  <roleSpec>
    <topicRef xlink:href="#kniha" />
  </roleSpec>
  <topicRef xlink:href="#numericke_metody" />
</member>
```

```
<member>
  <roleSpec>
    <topicRef xlink:href="#autor" />
  </roleSpec>
  <topicRef xlink:href="#m_prikryl" />
</member>
```

```
<member>
  <roleSpec>
    <topicRef xlink:href="#autor" />
  </roleSpec>
  <topicRef xlink:href="#m_brandner" />
</member>
```

```
</association>
```

Všechny použité hodnoty v typu vazby, rolích i jejich hodnotách musejí být už existující témata.

5.3 Occurrence

Výskyt realizuje vazbu mezi tématem a informací o něm a může mít dvě podoby. Buď informace může být uložena přímo v mapě témat jako vlastnost (stejně jako literál v případě RDF) nebo může výskyt odkazovat na externí informační zdroj. Druhý pří-

pad tedy představuje vazbu mezi znalostní vrstvou a vrstvou informačního obsahu. Výskyty, stejně jako vazby můžeme typovat. Typ vazby je opět tématem v mapě. Můžeme tak mít například rozlišené odkazy na obrázky, video a textové dokumenty. Následující příklad ukazuje použití jak odkazu na informační zdroj, tak data.

```
<topic id="zcu">

  <occurrence>
    <instanceOf>
      <topicRef xlink:href="#legislativa"/>
    </instanceOf>
    <resourceRef xlink:href=
      "http://legislativa.zcu.cz/ven/statut.pdf"/>
    </occurrence>

  <occurrence>
    <instanceOf>
      <topicRef xlink:href="#vznik"/>
    </instanceOf>
    <resourceData>1991</resourceData>
  </occurrence>

</topic>
```

Mapy témat jsou zajímavý a silný nástroj a spoustu jejich dalších vlastností zde nebylo diskutováno. Pro první přiblížení doporučuji dokument www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html.

6 Praktické použití

Běžný uživatel se již dnes setkává s RDF technologií prakticky denně, aniž by si to uvědomoval. Například v RSS 1.0 nebo ve standardu firmy Adobe nazvaného XMP, pomocí kterého lze ukládat metadata do různých formátů obrázků.

Použití mikroformátu je možné vidět na serveru panoramio.com, fotoalba spolupracujícího s Googlem a GoogleEarh, kde je u každé fotky pomocí mikroformátu Geo označena poloha. Další příklad mikroformátů naleznete ve žlutých stránkách¹⁹, kde je mikroformát použit pro kontakty. Uživatel dnes může také těžit z rozšíření prohlížečů, které mu při brouzdání webem nabízejí příjemným způsobem k prohlédnutí Dublin Core metadata nebo mikroformáty. Pro Firefox se jedná o rozšíření Dublin Core Viewer, respektive Tails.

Příkladem snahy v širším měřítku standardizovat pojmenování konceptů je tezaurus EUROVOC, který využívá například portál plzeňského kraje www.kr-plzensky.cz.

Pokročilí uživatelé využívají RDF a slovník FOAF pro publikaci osobních informací. RDF formát lze například použít i pro import dat do Google Base. Google Base je služba, přes kterou uživatel může nabízet své produkty. Při vyplňování mu systém nabízí používané kolonky pro vybranou kategorii zboží, ale lze vytvořit i vlastnost úplně novou. V pozadí tedy existuje jakási ontologie. Možná o Google Base v souvislosti se sémantickým webem ještě uslyšíme.

Pro programátory dnes existuje množství API. Pro Javu je to například Jena²⁰ a Sesame 2²¹, pro PHP RAP²² a ARC²³

¹⁹www.zlatelystranky.cz

²⁰Jena API: jena.sourceforge.net

²¹Sesame 2 API: www.openrdf.org

²²RAP - RDF API: sites.wiwiss.fu-berlin.de/suhl/bizer/rdfapi

²³ARC API: arc.web-semantics.org

a pro .NET Spiral²⁴.

Pro mapy témat bylo navrženo rozhraní TMAPI²⁵, které má implementaci pro Javu v podobě TM4J²⁶ a tinyTIM²⁷ i pro PHP – PHPTMAPI²⁸.

Pěkný prohlížeč RDF dat, je možné nalézt na stránkách univerzity v Salzburgu²⁹.

7 Pohled do budoucna

Otázka, zda se naplní všechny vize w3.org o sémantickém webu zůstává nezodpovězená. Jisté ale je, že některé se plní již dnes a konečné řešení reálných problémů internetu se stále hledá. Jakými prostředky se dosáhne zlepšení není podstatné. Možná se objeví úplně nová technologie, která se dočká globálního rozšíření.

Běžní uživatelé internetu půjdou vždy nejsnazší cestou a budou využívat jen technologie, které jim přinesou reálný užitek. Z toho důvodu dnes nejsou technologie sémantického webu, ani přes krásné vize, využívány nejširší veřejností. Až budou velké internetové vyhledávače zvyhodňovat webové stránky používající tyto nové technologie, bude nástup velice rychlý.

Nezapomínejme, že i samotný internet býval výsadou fyziků a čistě akademického prostředí. Poté následovalo jeho komerční využití a nakonec se připojili i běžní uživatelé. Rozšiřování technologií, stejně jako samotného internetu má exponenciální charakter.

²⁴Spiral RDF framework: semanticplanet.com

²⁵tmapi.org

²⁶www.tm4j.org

²⁷tinytim.sourceforge.net

²⁸phptmapi.sourceforge.net

²⁹semweb.salzburgresearch.at/apps/rdf-gravity/index.html

Na závěr si dovolím krátký pohled do budoucna. Mozilla připravuje v prohlížeči Firefox 3.0³⁰ integrovanou podporu pro mikroformáty a další prohlížeče nebudou jistě zůstat pozadu, protože sám Bill Gates vyjádřil mikroformátům podporu³¹. Není zřejmě daleko doba, kdy prohlížeč bude na stránce detekovat datové objekty a nabízet k nim příslušné akce. Například zkopírování události do kalendáře nebo přidání kontaktu do adresáře. Internet tak bude stále více používán jako nosič dat a to je hlavní myšlenka sémantického webu.

³⁰Firefox 3.0 features: wiki.mozilla.org/Firefox3/Product_Requirements_Document

³¹www.youtube.com/watch?v=Z9X-vHJ_Z-I

Reference

- ADIDA, B., BIRBECK, M. Rdfa primer 1.0. 2007. URL <<http://www.w3.org/TR/xhtml-rdfa-primer>>.
- BRICKLEY, D., GUHA, R., MCBRIDGE, B. RDF vocabulary description language 1.0: Rdf schema. 2004. URL <<http://www.w3.org/TR/rdf-schema/>>.
- DAVIS, I., TALIS. GRDDL primer. 2006. URL <<http://www.w3.org/TR/grddl-primer/>>.
- DOCTOROW, C. Metacrap: Putting the torch to seven strawmen of the meta-utopia. 2001. URL <<http://www.well.com/~doctorow/metacrap.htm>>.
- FILIPEC, J., DANES, F., MACHAC, J., MEJSTRIK, V. *Slovník spisovné češtiny*. Praha : Academica, 3. vyd., 2004. ISBN 978-80-200-1347-7.
- GARSHOL, L. M., MOORE, G. Topic maps data model. 2006. URL <<http://www.isotopicmaps.org/sam/sam-model/>>.
- LEISE, F., FAST, K., STECKEL, M. What is a controlled vocabulary? 2006. URL <http://www.boxesandarrows.com/view/what_is_a_controlled_vocabulary_>.
- MANOLA, F., MILLER, E., MCBRIDE, B., HP Labs. RDF Primer. 2004. URL <<http://www.w3.org/TR/rdf-primer/>>.
- SMITH, M. K., WELTY, C., MCGUINNESS, D. L. OWL web ontology language guide. 2004. URL <<http://www.w3.org/TR/owl-guide/>>.
- SVÁTEK, V. Ontologie a WWW. 2004. URL <<http://nb.vse.cz/~svatek/onto-www.pdf>>.