

Směrem k Sémantickému webu¹

Jaroslav Pokorný

MFF UK
Malostranské nám. 25, 118 00 Praha 1
Tel: 221 914 265
e-mail: pokorny@ksi.ms.mff.cuni.cz
www: <http://kocour.ms.mff.cuni.cz/layout.php>

Klíčová slova: Sémantický web, XML, RDF, ontologie, logika, agent

Abstrakt: Současné webové vyhledávače založené na technikách vyhledávání informací v textech nejsou schopny využít sémantické znalosti uvnitř webové stránky a tedy nemohou dávat uspokojující odpovědi na uživatelské dotazy. Možným řešením se zdá být tzv. Sémantický web, který koncem 90. let popsal ve své vizi Tim Berners-Lee. Ideou, stojící za Sémantickým webem, je rozšířit webové stránky značkováním, které podchytí alespoň část významu obsahu stránky. Cílem článku je uvést do technologií podporujících vytváření Sémantického webu.

Motto: “Web se dotkl mnoha stránek moderního života, od způsobu nákupu věcí k Nezměnil ovšem způsob jakým programujeme. Je to proto, že web neobsahuje strojově srozumitelné informace. Sémantický web by měl učinit web užitečný pro programy.” (R. Guha, 2004)

1 Úvod

Současný web bývá charakterizován jako web druhé generace. Na rozdíl od první generace, založené na ručně psaných HTML stránkách, druhá generace již pokročila směrem ke stránkám generovaným programy či dokonce ke stránkám aktivním. Typické pro obě generace je, že jak proces vyhledávání na webu, tak práce s webovým rozhraním založeným na vyplňování formulářů jsou řízeny člověkem. Oby typy činností jsou samozřejmě podporovány silnými nástroji, jako jsou webové vyhledávače [6]. Z hlediska architektury webu je úspěch práce s webem postaven na interoperabilitě, která je v obou generacích webu založena zejména na protokolech TCP/IP, HTTP a HTML.

Ještě obecnější požadavek vyjadřuje uvedené motto R. V. Guhy, jednoho z otců jazyka RDF. Zdá se, že jakékoliv další zlepšení současného využívání webu vyžaduje naplnění minimálně čtyř požadavků, které spolu navzájem úzce souvisí:

- (1) realizovat inteligentnější informační webové služby, než je tomu dosud,
- (2) zajistit univerzální vyjadřovací sílu webu, tj. umožnit pracovat s jakoukoliv formou dat,
- (3) podporovat syntaktickou interoperabilitu (znovupoužití softwarových komponent, jako parserů, API pro dotazování apod.),

¹ Práce vznikla v rámci grantového projektu IET100300419 Národního programu výzkumu (Informační společnost).

- (4) realizovat sémantickou interoperabilitu webových zdrojů (umožnit definovat zobrazení mezi pojmy apod.).

Požadavek (1) vychází z kritiky současných webových vyhledavačů založených na Boolovském modelu. Mají některé vážné problémy, jako např. [5]:

- Vysoký koeficient úplnosti a malý koeficient přesnosti. Dokonce i když je získáno hodně relevantních stránek, jsou smíchány s mnoha nerelevantními.
- Nízký nebo nulový koeficient úplnosti. Jde o případ, kdy neobdržíme žádné relevantní stránky, přestože víme, že musí existovat.
- Výsledky jsou závislé na abecedě modelu, tj. někdy je nutné ve výrazu dotazu změnit počáteční množinu termů (klíčových slov), protože v relevantních stránkách může být použita různá terminologie. To odporuje představě, že sémanticky podobné dotazy by měly vracet podobné výsledky.
- Výsledkem vyhledávání jsou celé webové stránky. Jsou-li potřebné informace rozmístěny ve více relevantních dokumentech, je nutné je integrovat ručně. Naopak, celá stránka může být příliš rozsáhlá pro vyhledání potřebné relevantní části.
- Je zřejmé, že zatím nejde o vyhledávání informací, ale spíše o nalezení místa, kde se potřebné informace (možná) nalézají. Problém je totiž v tom, že sémantika obsahu webu není strojově zpracovatelná.

Současné techniky vyhledávání informací (Information Retrieval - IR) nejsou zřejmě schopny sémantické znalosti uvnitř webové stránky využít a tedy nemohou dát uspokojivou odpověď na uživatelské dotazy. Jedním ze způsobů jak řešit tuto situaci je opřít se o techniky umělé inteligence a aplikovanou lingvistiku. Další, technicky skromnější z možných řešení, dnes nabízí vývoj *Sémantického webu*, který koncem 90. let popsal ve své vizi Tim Berners-Lee. Ideou, stojící za Sémantickým webem (viz [7] jako zdroj dalších odkazů), je rozšířit webové stránky značkováním, které podchytí alespoň část významu obsahu stránky. Toto sémantické značkování znamená přidání jistých metadat, která poskytují formální sémantiku obsahu webu.

Značkování v souvislosti s webem již samozřejmě existuje. Napíšeme-li např. v HTML

```
<H2>Vyhledávání informací na webu</H2>
```

specifikujeme, jak má vypadat prezentace daného řetězce na obrazovce. Značkování však nic nevyovídá o sémantice tohoto řetězce.

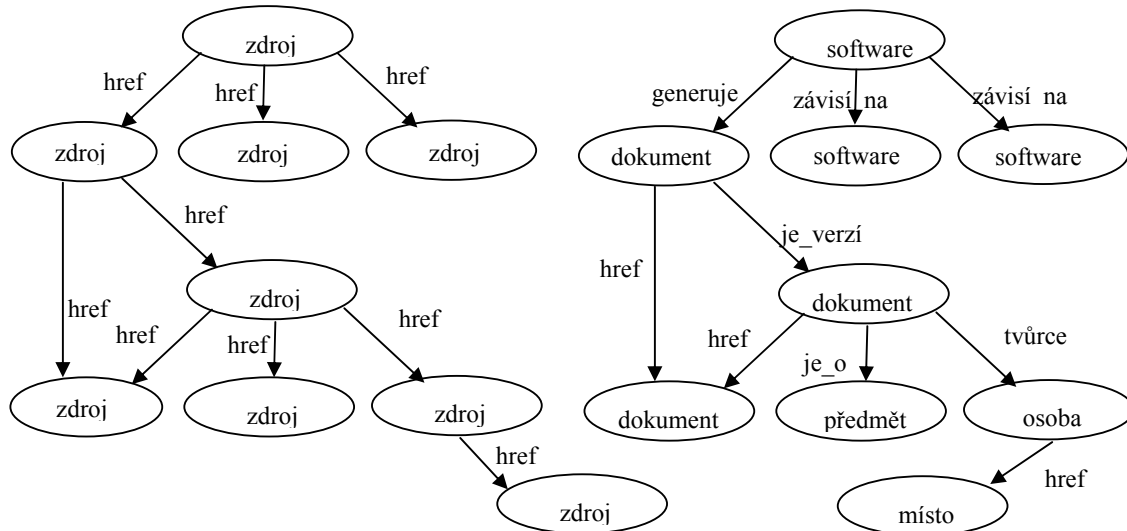
Sémantický web bývá charakterizován jako „web s významem“ v tom smyslu, že program (inteligentní agent), který má zpracovávat data na webu, se může tento význam dozvědět a využít ho. Jinými slovy řečeno jde o koncepci webu, jehož obsah může být organizován spíše sémanticky než za pomoci syntaktických a strukturálních metod. Současně je ovšem třeba vyvinout automatické prostředky, které mohou nalézat a sbírat znalosti anotované na stránkách s cílem porozumět těmto znalostem a využít je pro vyhodnocování dotazů. V takovém webu již nevyhledáváme stránky obsahující podobná slova, ale podobné pojmy. Rozdíl mezi obyčejným webem a Sémantickým webem ukazuje obrázek 1 převzatý z [2]. Např. zdroj označený jako `software` je možno chápat tak, že k jeho identifikátoru je přiřazen typ zdroje `software`.

Sémantické značkování může být částečně dosaženo pomocí standardizovaného značkovacího jazyka XML [9]. Např. XML text

```
<ADRESA>
  <JMÉNO>Michal</JMÉNO>
  <ULICE>Na hroudě 5</ULICE>
  <MĚSTO>Praha 10</MĚSTO>
</ADRESA>
```

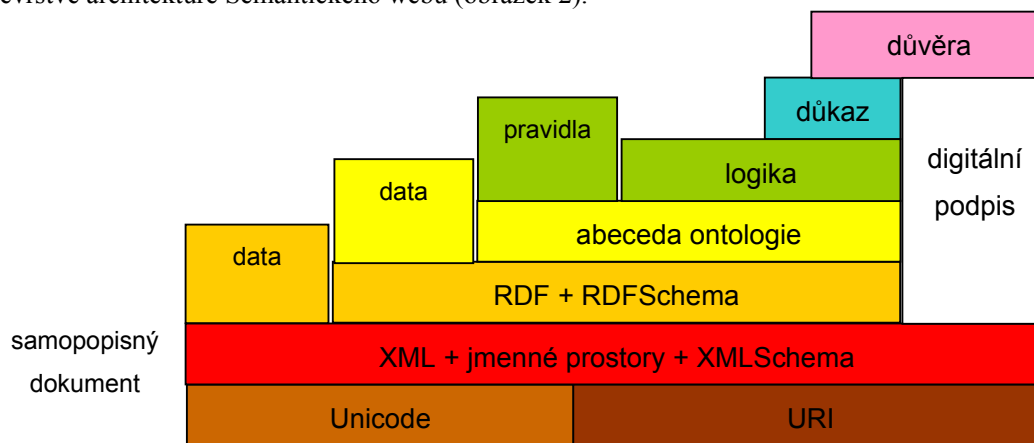
obsahuje značky srozumitelné uživateli, nikoliv však stroji. XML je totiž určen zejména pro specifikaci struktury dokumentu. Poněvadž dokument v jazyku XML je definován pouze na

syntaktické úrovni, není problém určit jednoznačně a korektně význam XML značek obvyklými prostředky řešitelný. Pro Sémantický web je podstatné, že XML a další jazyk XML Schema lze využít pro zápis metadat. Oba jazyky tak slouží k zajištění požadavku (3) na syntaktickou interoperabilitu. Není obtížné naplnit i požadavek (2). Viděli jsme však, že to neplatí pro sémantickou interoperabilitu. Díky tomu, že XML slouží vlastně pro definování gramatiky, cokoliv může být zakódováno v XML. Znakový systém UNICODE použitý v XML kromě toho umožňuje vyjádřit více než 90000 znaků všech jazyků na světě.



Obr. 1: Rozdíl mezi současným a Sémantickým webem

Vlastní metadata jsou modelována pomocí dalších standardizovaných speciálních jazyků RDF [3] a RDFS [1]². Dalšími technologiemi jsou ontologie, logika a inteligentní agenti. Budeme o nich hovořit v dalších odstavcích. Zdůrazněme, že ačkoliv Sémantický web využívá současných technik umělé inteligence, není umělá inteligence jeho cílem. Jde o spíše inženýrské řešení, které má zlepšit současné pojetí webu. Opírá se o zmíněné standardy či návrhy standardů. Ty jsou použity ve vícevrstvé architektuře Sémantického webu (obrázek 2).



Obr. 2: Vize Sémantického webu

Mezi prostředky pro práci se Sémantickým webem patří zejména jazyky umožňující dotazování a aktualizaci dat i metadat na webu. Síla těchto jazyků zahrnuje nejen klasické dotazování či

² Oby jazyky se označují RDF/S.

aktualizaci semistrukturovaných dat, ale i odvozování, umožňující uvažování nad daty a metadaty webu, tj. získávání nových tvrzení.

Jak jsme již zmínili, podchycení a využití sémantiky zdrojů představuje pouze část problému stávajícího webu. Vícevrstvá architektura Sémantického webu je navržena tak, aby mohla být realizována interoperabilita či dokonce sémantická interoperabilita (viz požadavek (4)). Pak lze Sémantický web na technické úrovni chápat jako systém specifikací, metod a prostředků pro publikování (meta)dat spolu s možnostmi odvozování, a to za použití datových modelů a formátů umožňujících interoperabilitu [9]. Protože se jedná o kvalitativní změnu ve vývoji webu, je možné v této souvislosti hovořit o třetí generaci webu.

V sekcích 2. – 4. vysvětlíme stručně základy jazyků RDF/S, principy ontologií, použití logiky a softwarových agentů, kteří zajišťují dynamicky procesy spojené s vyhledáváním v Sémantickém webu. Vynechány jsou problémy důvěry související se speciálními agenty, certifikačními agenturami a uživatelskými organizacemi. V sekci 5 budeme diskutovat jednu z možných architektur inteligentních vyhledávacích strojů. V závěrech nastíníme některé problémy současného vývoje Sémantického webu.

2 Jazyky RDF/S

Technologickou základnou pro reprezentaci metadat v Sémantickém webu je jazyk RDF (Resource Description Framework), doporučený konsorciem W3C (World Wide Web Consortium).

Řečeno databázovými pojmy, je RDF druhem datového modelu. Základními objekty zájmu jsou tzv. *zdroje*. Pojem zdroj je ovšem míněn obecněji a neznamená pouze webovou stránku. Zdroj je identifikován pomocí *URI* (Uniform Resource Identifier). URI může být přiřazeno jakémukoliv objektu, abstraktnímu pojmu, osobě apod. Fakticky může jít o URL, ale i identifikace jako dobře známé ISBN, rodné číslo apod. Při tvorbě URI se uplatní prostory jmen a notace předpona:přípona. Např. v prostoru urn, kde URN označuje Uniform Resource Names, mohou existovat ISBN, takže např. urn:isbn:80-246-0837-5 identifikuje skripta Základy implementací souborů a databází autorů Pokorného a Žemličky. Jediný požadavek na URI kladený je, aby bylo jednoznačné.

Zdroje jsou charakterizovány *vlastnostmi* (nebo *atributy*) a popisují se pomocí výroků. *Výrok v RDF* asociuje se zdrojem dvojicí vlastnost:hodnota; jinými slovy jsou výroky trojice složené z předmětu (zdroje), predikátu (vlastnosti) a objektu (hodnota vlastnosti).

- *Předmět* je zdroj, na který se můžeme odkazovat pomocí URI. Jinak je zdroj *anonymní* a běžně se označuje ve tvaru `_ :n`, kde `n` je přirozené číslo.
- *Predikát* definuje binární relaci mezi zdroji a/nebo atomickými hodnotami poskytovanými definicemi primitivních datových typů v XML. Také predikát může být zadán pomocí URI.
- *Objekt* specifikuje pro nějaký předmět hodnotu vlastnosti. Charakteristiky webových dat jsou tedy dány pomocí těchto objektů. Objekt je buď opět nějaký zdroj nebo literál.

Příkladem výroku v RDF je třeba

```
<osobní:Jarda, osobní:email, osobní: jarda.p@mail.cz>
```

kde `osobní` je předpona označující hypotetický prostor jmen, který se týká personalistiky. V přirozeném jazyce odpovídá danému výroku věta: „Informační zdroj identifikovaný jako `osobní:Jarda` má emailovou adresu `jarda.p@mail.cz`“. Příkladem jiného jmenného prostoru může být standard Dublin Core³ (česky také Dublinské jádro), používaný pro vytváření metadat dokumentů. Obsahuje elementy jako TITLE, CREATOR, SUBJECT, DESCRIPTION,

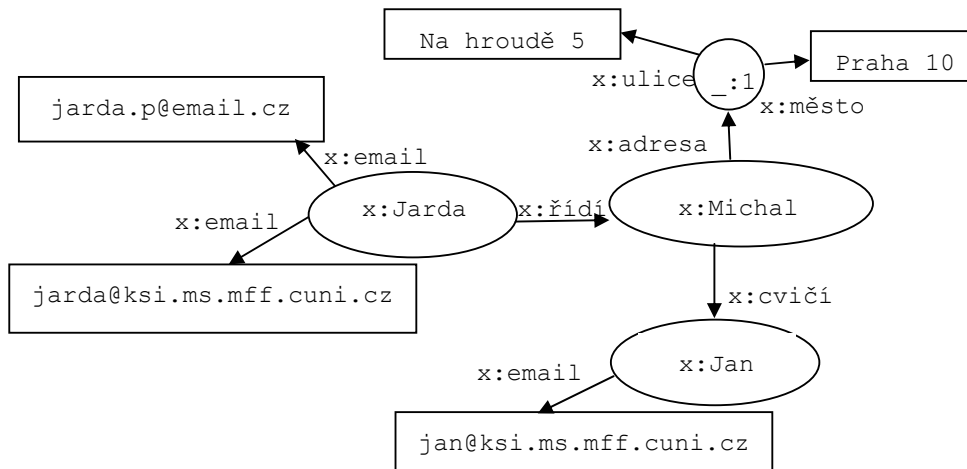
³ <http://dublincore.org/>

PUBLISHER, FORMAT a několik dalších s definovaným významem. Význam je specifikován krátkou anglickou větou. Popíšeme-li tedy v trojici predikát Titul s předponou dc označující prostor jmen daný Dublinským jádrem, jedná se titul ve významu daném na webové stránce standardu Dublin Core. Tam se lze dočíst definici TITLE: A name given to the resource.

Terminologie RDR používá speciální termín – *RDF graf* definovaný jako množina zmíněných trojic. Někdy se také hovoří o *RDF modelu* nebo o *RDF datech*. Pro vyměňování a zpracovávání RDF dat může být výhodná jejich serializace v XML (nazývá se často *RDF/XML*). V RDF/XML se pro elementy RDF používá předpona rdf, pro vyjádření prostorů jmen pomocí atributů XML se používá předpona xmlns apod. Prostor jmen pro Dublin Core se pak specifikuje jako XML atribut xmlns:dc = 'http://purl.org/DC/'. Z hlediska modelování jde v tomto pohledu o semistrukturovaná data.

V grafickém vyjádření lze *RDF graf* popsat jistým ohodnoceným orientovaným multigrafem. Dva uzly mohou totiž být spojeny více hranami. V grafové reprezentaci se nicméně objevují jisté potíže, které vyžadují poněkud přesnější přístup, jehož detaily přesahují cíl tohoto článku. Ovály se používají pro zdroje, obdélníky pro literály a orientované označené hrany pro vlastnosti. Zdroje mohou mít více vlastností, literály se ve výrocích vyskytují pouze jako objekty. Vlastnosti, jako označení hran, jsou vyjádřeny pomocí URI. Pomocí x: pro jednoduchost vyjadřujeme libovolné URI.

RDF-graf reprezentující náš příklad je na obrázku 3. Strukturovaný údaj, jakým je adresa, se zde vyjadřuje pomocí anonymního zdroje. Nepředpokládá se, že by se na něj odněkud odkazovalo. V opačném případě by bylo třeba konstruovat URI a zdroj pro adresy.



Obr. 3: RDF data graficky

RDF poskytuje možnosti modelovat kontejnery pomocí konstruktů bag (neuspořádaná multimnožina), sequence (uspořádaná multimnožina) a alternativy pomocí alternative. Možná je i tzv. *reifikace*, kdy se výrok stává sám objektem, tj. je možné vytvářet výroky o výrocích.

Všimněme si, jak lze jednoduše slévat RDF data. Jestliže jsou některé zdroje označeny URI, pak toto URI je globálně jednoznačné. Uzly se stejným URI označují též zdroj a mohou tedy sloužit jako body pro slévání.

Již jsme zmínili, že tvrzení v RDF lze vyjádřit v XML. Odpovídající syntaxe je ovšem několikrát. Existují minimálně dvě doporučené varianty – hnížděnými elementy a pomocí atributů. Pro data z obrázku 3 využijeme dva prostory jmen: rdf (podle standardu RDF) a uni (podle schématu Univerzita) a hníždění elementů.

```

<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/TR/REC-rdf-syntax/#"
  xmlns:uni="http://www.kocour.mff.cuni.cz/univerzita#">
  <rdf:Description about="http://www.mostarna.cz/Jarda">

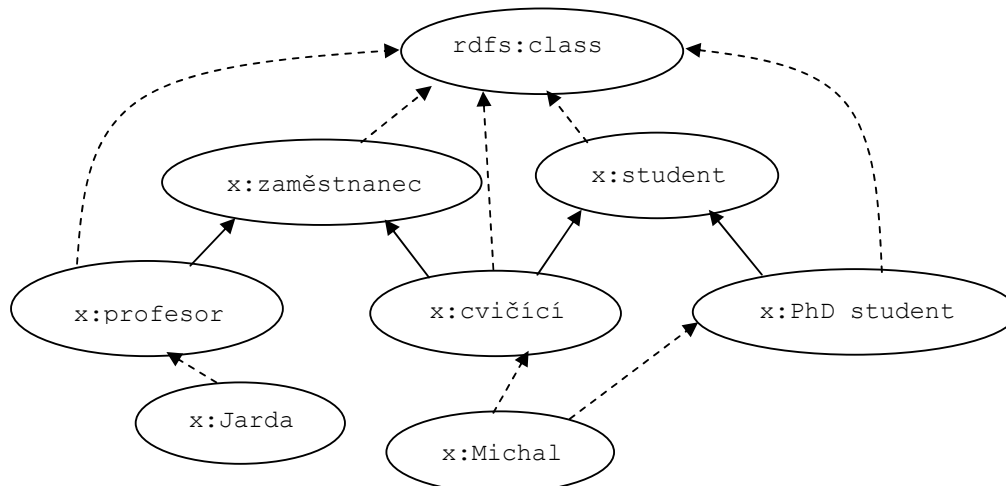
```

```

<uni:email> jarda.p@email.cz </uni:email>
<uni:email> jarda@ksi.ms.mff.cuni.cz </uni:email>
<uni:řídí>
  <rdf:Description
    about="http://www.rybarna.cz/Michal">
    <uni:adresa>
      <rdf:Description about="_:1">
        <uni:ulice>Na hroudě 5</uni:ulice>
        <uni:město>Praha 10</uni:město>
      </rdf:Description>
    </uni:adresa>
    <uni:cvičí>
      <rdf:Description
        about="http://www.student.cz/Jan">
        <uni:email>jan@ksi.ms.mff.cuni.cz</uni:email>
      </rdf:Description>
    </uni:cvičí>
  </rdf:Description>
</uni:řídí>
</rdf:Description>
</rdf:RDF>

```

Dalším jazykem je RDF Schema (zkráceně RDFS). Pomocí RDFS se popisují koncepty použité v RDF datech a typová omezení na objekty a hodnoty ve trojicích. Takový popis lze po vzoru databází nazvat *RDF schématem*⁴. Máme-li k dispozici nějaké RDF schéma, je možné RDF data validovat vzhledem k tomuto schématu.



Obr. 4: Specifikace tříd pomocí RDFS (vlastnosti *subClassOf* a *instanceOf*)

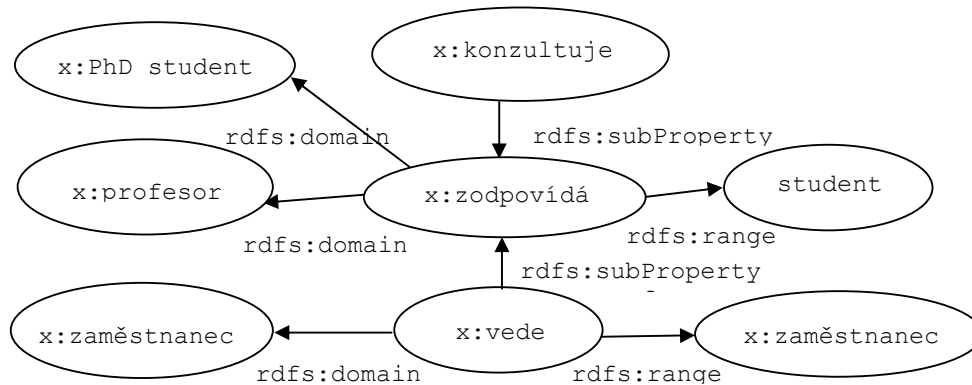
Jazyk RDFS vychází z objektově orientovaných přístupů. Je založen na třídách a ISA-hierarchiích tříd s množinovou sémantikou. Na obrázku 4 jsou plnou čarou vyjádřeny vztahy specializace (cvičící JE zaměstnanec, PhD student JE student), čárkovaně vztahy instanciace (Michal JE INSTANCÍ cvičícího). Specifikujeme-li, že cvičící JE student, znamená to, že množina instancí třídy cvičící je podmnožinou instancí třídy student.

Na nejvyšší úrovni konceptů jsou v RDFS kategorie `rdfs:Resource`, `rdfs:class` a `rdfs:property`. Třídy v RDF Schema se ovšem nedefinují pomocí atributů a nemají asociace

⁴ Díky nevhodným původním názvům, je třeba pečlivě odlišovat jazyk RDF Schema od specifikace v tomto jazyku, zde nazývané RDF schéma. Podobně je třeba rozlišovat XML Schema jako jazyk a XML schéma v jazyku XML Schema. Protože jde opět o text v XML, označuje se někdy toto schéma jako XSD (XML Schema Document).

k dalším třídám. Atributy (zde vlastnosti) jsou objekty samy o sobě. V jazyku RDFS je povoleno vícenásobné dědění.

Ukázka části schématu, která popisuje pouze vlastnosti, je na obrázku 5.



Obr. 5: Specifikace vlastností pomocí RDF

ISA-hierarchie se tedy v RDFS modelují pomocí jistého typu vlastnosti. Základní typy vlastností jsou:

- *instanceOf* a *subClassOf*: *instanceOf* definuje vztah mezi zdrojem a prvkem třídy *Class*, *subClassOf* definuje vztah mezi dvěma prvky ze třídy *Class*. O vlastnosti *subClassOf* se předpokládá, že je tranzitivní.
- *Constraint* je podtřídou *Property*. Má dvě základní instance *range* a *domain* aplikovatelné k typům vlastností, které mají třídu jako hodnotu. *Range* a *domain* definují obor hodnot resp. doménu typů vlastností.

Povšimněme si, že vyjadřovat metadata v RDF/S a zapisovat je v XML ručně je zřejmě časově náročné a složité. Pro jejich generování je třeba využít odpovídajících softwarových nástrojů. Pro zpracování těchto metadat se využívá databázových řešení a speciálních dotazovacích jazyků.

3 Ontologie

Při zpracování dat na webu je často třeba překonat terminologické problémy. Co je např. pro jednu komunitu uživatelů typ výrobku, může být pro jiného kategorie zboží. Naopak jeden term může mít ve dvou komunitách různý význam. Např. term *informatika* může znamenat někde obor ukončení studia na vysoké škole, jinde jméno předmětu. Data obsažená v RDF modelech obvykle již nemají žádné další schéma, které by je blíže popisovalo. Místo toho je význam dat specifikován pomocí ontologií.

Pojem ontologie pochází z filosofie, kde označuje nauku o bytí, která řeší dvě základní otázky: (1) co existuje a (2) jestliže to, co existuje, je rozložitelné na části, co jsou tyto části a jaké jsou mezi nimi vztahy. V informatice je možné transformovat tyto otázky do otázek (1') jaké objekty a pojmy se předpokládají v kontextu uvažované domény zájmu a (2') jaké jsou mezi nimi vztahy. *Ontologie* je tedy explicitní specifikace abecedy pro nějakou doménu. Sdílené ontologie [4] pomáhají učinit sémantické značkování kompatibilní mezi jednotlivými informačními zdroji a dotazy. Z hlediska uživatelů ontologie to znamená mít prostředek pro společné chápání dané domény.

Tento přístup připomíná z minulosti známé pojmy jako konceptuální modelování a integrace heterogenních databází. Ontologie a jejich sdílení znamená rozšíření těchto koncepcí do prostředí webu. Jedno z jejich využití je jistě v inteligentních vyhledávacích strojích. Nezbytné jsou dále ve zcela praktických úlohách, jako je např. e-byznys. Umožňují strojově zajišťovanou komunikaci mezi prodávacem a zákazníkem, vertikální integraci trhů apod.

Ontologie jsou výsledkem dohody, což v praxi znamená vyvinout hodně úsilí a rozsáhlé časové investice pro jejich vytvoření. Protože rozsah různosti dat na webu je značný, ontologie by měly být specifické vzhledem k nějaké doméně, nebo personalizované tak, aby vyjadřovaly specifické zájmy osob nebo komunit. V praxi se používají dva způsoby vytváření ontologií: shora-dolů a zdola-nahoru. V prvním případě se začíná od obecných pojmů a postupuje se ke speciálním. Pro Sémantický web je typičtější druhý případ, kdy se doménové ontologie integrují do větších celků.

Ontologie obvykle zahrnuje definice tříd, vztahů mezi třídami, funkcí a omezení. Třídy odpovídají pojmům, např. učitelé, studenti, předměty a přednášky. Typicky obsahuje ontologie hierarchický popis důležitých pojmů dané domény, obvykle pomocí mechanismu atribut - hodnota. Obrázky 4 a 5 naznačují, jak mohou ontologie vypadat v grafickém vyjádření. V přístupech založených na logice se k těmto definicím ještě přidávají axiomy.

Ontologie jsou též reprezentovány pomocí RDF schématu, jehož interpretace definuje sémantiku ontologie. Bohužel, jak RDF, tak RDFS jsou stále ještě omezené ve srovnání s propracovanějšími jazyky vyvinutými dříve pro reprezentaci znalostí.

Ontologie jsou užitečné zejména z hlediska vyhledávání na webu. Cílem je nalézt stránku obsahující ne klíčové slovo, nýbrž daný pojem. K tomu slouží jednak slévání ontologií, jednak modifikace dotazu směrem k obecnějším pojmům (je-li nízký koeficient úplnosti), nebo směrem ke speciálním pojmům (je-li nízký koeficient přesnosti). Tato představa je reálná pro ontologie, které se příliš nepřekrývají, což samozřejmě v praxi není možné dodržet. Pojmy týkající se hradů a zámků z hlediska cestovní kanceláře budou spíše odlišné od pojmů používaných památkáři starajícími se o tyto objekty. Je tedy třeba řešit důležitou úlohu překladu či transformace ontologií.

Nejdůležitějšími jazyky pro budování ontologií jsou XML, XML Schéma a RDF/S. Perspektivní je i další jazyk pro webové zdroje DAML+OIL [8]. K rysům XML a RDF/S jazyků přidává pravidla pro popis různých omezení a vztahů mezi informačními zdroji. S DAML+OIL je možný vývoj inteligentních agentů pro implementaci dotazování v prostředí Sémantického webu.

4 Logika, agenti

Další, vyšší úroveň Sémantického webu využívá jazyky logiky. V nejjednodušším přístupu vystačíme s predikátovým počtem. Uvažujme následující množinu klauzulí:

$$\begin{aligned} \text{člen_katedry}(X) &\Rightarrow \text{zaměstnanec}(X) \\ \text{prof}(X) &\Rightarrow \text{člen_katedry}(X) \\ \text{prof}(\text{Evžen}) & \end{aligned}$$

Pak lze odvodit, že

$$\begin{aligned} \text{člen_katedry}(\text{Evžen}) \\ \text{zaměstnanec}(\text{Evžen}) \\ \text{prof}(X) &\Rightarrow \text{zaměstnanec}(X) \end{aligned}$$

Všimněme si, že se jedná o znalosti dané ontologiemi. Logika tedy pomáhá odkrýt implicitní znalosti z ontologií, v lepším případě neočekávané vztahy nebo nekonzistence.

Softwaroví agenti jsou programy, které pracují autonomně a proaktivně. Příkladem takového agenta může být osobní agent fungující nad Sémantickým webem, který obdrží nějaké úkoly a preference od konkrétní osoby - uživatele, prohledává informace z webových zdrojů, srovnává je s uživatelskými požadavky, provádí výběr a vrací odpovědi uživateli.

Agenti Sémantického webu využívají všechny dosud zmíněné technologie:

- metadata k nalezení a extrakci informací z webových zdrojů,
- ontologie k asistenci při vyhledávání na webu, interpretaci vybraných informací a ke komunikaci mezi agenty,
- logiku pro zpracování získaných informací a pro formulaci závěrů.

Uplatnění logiky je možné i v případech, jde-li o to učinit rozhodnutí. Obchodní agent může rozhodnout dát slevu zákazníkovi, který má věrnostní průkazku (dobrý_zákazník). K tomu účelu může sloužit pravidlo

$$\text{dobrý_zákazník}(X) \Rightarrow \text{sleva}(2\%)$$

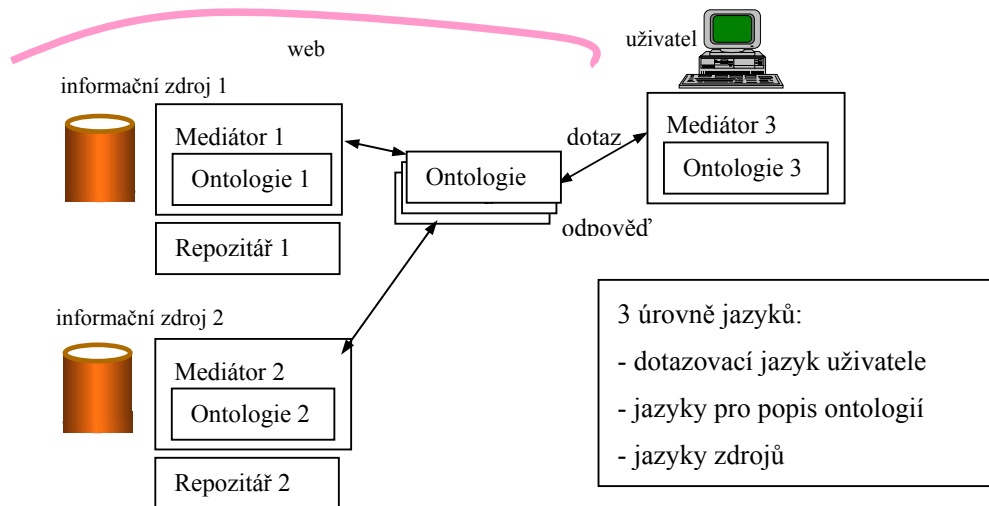
přičemž vlastnosti zákazníka se zjistí z dat nějaké databáze. Jiným praktickým příkladem může být situace, kdy (softwarový) agent má zajistit nákup 10 tiskáren, z nichž 6 objedná u jedné firmy a 4 u druhé. To, že požadavek byl naplněn, je možné zajistit spuštěním jistého pravidla.

Jinou výhodou logiky v prostředí Sémantického webu je její schopnost objasňovat závěry v případě posloupnosti nějakých kroků odvození. Taková vysvětlení pomohou zvýšit důvěryhodnost softwarových agentů pracujících v daném prostředí. To lze uplatnit u činností, kdy agent např. generuje zprávu „Dosud nesplaceny pohledávky v celkovém objemu 15000,- Kč“. Objasňování je žádoucí nejen mezi agentem-programem a agentem-člověkem, ale i mezi agenty-programy navzájem.

Závěrem připomeňme, že pro praktické účely je třeba použít pouze takovou část logiky, která vede k efektivnímu zpracování. Klausule použité v příkladech (říká se jim Hornovy) splňují tuto podmínku.

5 Inteligentní vyhledávací stroje

K integraci v rámci Sémantického webu je možné využít známé technologie, jako jsou tzv. *obálky* (wrappers) a *zprostředkovatelé* (mediators, počestně též *mediátory*). První z nich zajišťují skrývání rozdílů pomocí společného datového modelu, ty druhé realizují vlastní integraci jednotlivých komponent. V Sémantickém webu obálky zahrnují ontologie, které zjednodušují práci mediátorům. Obecnější architektura informačních zdrojů zahrnuje v Sémantickém webu také zdroje newebových dat uložené v klasických databázích, případně souborech. V důsledku toho jsou potřeba metadata popisující také tyto zdroje. Klasická schémata databází nejsou v daném případě dostačující.



Obr. 6: Architektura zprostředkovaného vyhledávání na webu

Viděli jsme, že výhoda strojově čitelných metadat, zřejmě v RDF, je také v tom, že mohou být použita k odvozování dalších sémantických relací. Sémantické značkování tedy může pomoci *sémantickému vyhledávání*. *Inteligentní vyhledávače* se vyvíjejí jako následníci současných vyhledávačů nebo metavyhledávačů. Zahrnují sémantiku webu a používají vyhledávací techniky založené např. na strojovém učení, odvozovacích mechanismech, netradičních logikách a nepřesném vyhledávání. Zejména pak rozšiřují a zlepšují běžně používané technologie IR. Takové přístupy umožní v budoucnosti realizaci inteligentních (sémantických) webových informačních služeb, budování personalizovaných webových míst a sémanticky zesílené vyhledávací stroje. Možná zprostředkující architektura vyhledávání na webu s ontologiemi je zobrazena na obrázku 6.

6 Závěr

Techniky Sémantického webu lze použít oblastech jako jsou:

- vyhledávání informací a výměna dat mezi zdroji (tématické portály),
- management znalostí, např. pro komunikaci mezi dvěma inteligentními agenty,
- elektronické obchodování ve variantách B2B nebo B2C.

Rozvoj Sémantického webu je zatím zaměřen hlavně na získávání informací, méně již na postihnání jeho dynamických aspektů. Tím, jak se web neustále vyvíjí, je nutné mít k dispozici jazyky specifikující tento vývoj. Nelze provádět pouze aktualizaci webových zdrojů, ale také propagovat tyto změny k dalším zdrojům, ontologiím. Rozhodne-li se např. agent pro rezervaci nějakého letového spoje, může tato změna mít vliv na rezervaci taxi pro odjezd z letiště. V řeči událostí a triggerů, událost vyjití nové knihy daného zaměření může spustit služby provádějící nabídku konkrétnímu uživateli, který se zajímá o nové knihy toho zaměření.

Jakkoliv se zdá koncepce Sémantického webu futuristická, jedno jisté, že se stále vyvíjí a investují se do ní nemalé prostředky. Často se řeší projekty, které nezahrnují web jako celek, ale jeho jisté části, vedoucí ke koncepcím např. již zmíněných tématických portálů. V souladu s názvem hlavní přednášky Marie-Christine Roussetové, jedné z klíčových postav loňské konference ISWC 2004 v Hirošimě – Malé může být hezké v Sémantickém webu.

Literatura

- [1] Brickley, D., Guha, R.V.: Resource Description Framework (RDF) Schema Specification 1.0. TR W3C, W3C Recommendation. <http://www.w3.org/TR/2000/CR-rdf-schema-2000372/>.
- [2] Koivunen, M.-R., Miller, E.: W3C Semantic Web Activity. In: Semantic web Kick-off in Finland, Ed. E. Hyvönnen, 2002.
- [3] Lassila, O., Swick, R. R.: Resource description framework (RDF) model a syntax specification. TR W3C, 1999. W3C Recommendation. <http://www.w3.org/TR/REC-rdf-syntax>.
- [4] Mena, E., Kashyap, V., Illarramendi, A., and Sheth, A.: Domain Specific Ontology for Semantic Information Brokering on Global Information Infrastructure. In: Proc. International Conf. on Formal Ontology in Information Systems, Trento, Italy, 1998.
- [5] Pokorný, J.: Vyhledávání na webu. In: Sborník příspěvků 19. ročníku konference Moderní databáze, Hotel Amber, Roudnice n. L., KOMIX, pp. 3-14.
- [6] Pokorný, J.: Web Searching and Information Retrieval. *Computing in Science & Engineering*, 2004, Volume 6, Number 4, pp. 43-48.
- [7] W3C: Semantic web. <http://www.w3.org/2001/sw/>.
- [8] W3C: DAML+OIL web Ontology Language. <http://www.w3.org/Submission/2001/12/>.
- [9] W3C: Extensible Markup Language (XML) 1.0 (Third Edition) <http://www.w3.org/TR/REC-xml/>
- [10] W3C: Semantic Web Activity Statement. <http://www.w3.org/2001/sw/Activity>.

Summary

Current web search machines based on techniques of text information retrieval are not able to use semantic knowledge stored inside web pages. Consequently, they can not return satisfiable answers to the user's queries. A possible solution seems to be so called Semantic web, which was described in the vision by Tim Berners-Lee at the end of 90ties. The idea behind the Semantic web is to equip web pages with a mark-up, which captures at least part of meaning of their content. The goal of the paper is to introduce into technologies supporting development of the Semantic web.