

## Digitální knihovny v prostředí sémantického webu

Jaroslav Pokorný\*

jaroslav.pokorny@mff.cuni.cz

**Abstrakt:** Digitální knihovny (DK) přispívají k rozvoji Sémantického webu a současně mohou využívat jeho technologické prvky. Lze tak docílit kvalitnějšího řízení dat v DK a snazší integrace více DK, jakož i a zvýšení možnosti interakce s dalšími informačními zdroji. Ideou, stojící za Sémantickým webem, 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. Projekty Sémantického webu vycházejí z technologií, které jsou vyvíjeny jako standardy. Patří sem jazyky XML, XML Schema, RDF a RDF Schema. Tyto jazyky slouží pro zápis metadat, z nichž část se organizuje v ontologiích. Další úroveň Sémantického webu využívá jazyky logiky. Základ zpracování v takto pojatém webu poskytují programy – softwaroví agenti. Cílem článku je uvést do technologií Sémantického webu a ukázat jejich uplatnění při vytváření DK.

**Klíčová slova:** Digitální knihovna, Sémantický web, XML, RDF, ontologie, logika, agent

### 1 Úvod

Současný rozvoj webu bývá označován jako přechod ke třetí generaci webu. Od ručně psaných stránek v HTML (1. generace), přes stránky generované programy (2. generace) jsou dalším cílem architektury, které se opírají se o techniky umělé inteligence a aplikovanou lingvistiku. Technicky skromnější řešení dnes nabízí vývoj *Sémantického webu*, který koncem 90. let popsali ve své vizi Tim Berners-Lee a který je řízen konsorciem W3C (World Wide Web Consortium). Ideou, stojící za Sémantickým webem (viz [13] jako zdroj dalších odkazů), je rozšířit webové stránky značkováním, které podchytí alespoň část významu obsahu stránky. Netřeba zdůrazňovat, že musíme také podchytit význam značek. V obou případech to znamená přidat k datům na webu jistá metadata, která poskytují formální sémantiku obsahu webu.

Pro uživatele je v každé generaci webu nejzajímavější vyhledávání informací podporované silnými nástroji, jako jsou webové vyhledávače a odpovídající prohlížeče [9]. Nezanedbatelným rysem jakéhokoliv vyhledávání je jeho kvalita a přesnost vyhledávaných informací. Již dlouho je zřejmé, že vyhledávání v dnešním webu má mnoho neduhů, jako:

- 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 vyhledávání, 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.

---

\* MFF UK, Malostranské nám. 25, 118 00 Praha

- 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.

V této souvislosti představují digitální knihovny (DK) s dobře organizovanými informacemi ve webovém prostředí ostrůvky neobvyklé kvality. Jejich prostřednictvím je možný přístup nejen k elektronickým informacím, ale i k dalším fondům (historické dokumenty, staré ročníky časopisů, výstřižky apod.), které byly dříve přístupné pouze uživatelům klasické knihovny.

Vyhledávače DK se ovšem liší od webových v několika aspektech:

- DK jsou spíše uzavřené systémy, do nichž jsou zdroje cílevědomě přidávány, takže činnost robota, který prolézá webové stránky není nutná,
- kromě indexace ve stylu zpracování plných textů (fulltext) jsou často využívány explicitní popisy zdrojů, což zvyšuje kvalitu vyhledávání,
- uživatelé DK mají jasnější představu o cíli, ke kterému chtějí vyhledáváním dospět, než běžní uživatelé webu.

Na druhé straně v DK přetrvávají problémy řízení dat a integrace DK [9].

Jakékoliv zlepšení daného stavu využívání webu nebo DK vyžaduje naplnění minimálně čtyř požadavků, které spolu navzájem úzce souvisí:

- (1) realizovat inteligentnější informační služby, než je tomu dosud,
- (2) zajistit univerzální vyjadřovací sílu webu resp. DK, tj. umožnit pracovat s jakoukoliv formou dat,
- (3) podporovat syntaktickou interoperabilitu (znovupoužití softwarových komponent, jako jsou parsery označovaných dat, aplikační programová rozhraní, jako jsou rozhraní pro dotazování, apod.),
- (4) realizovat sémantickou interoperabilitu informačních zdrojů (umožnit definovat zobrazení mezi pojmy apod.).

Požadavek (1) vychází z kritiky současných vyhledávačů založených na Booleovském modelu. Jejich inovace pomocí současných technik vyhledávání informací (Information Retrieval) mají své hranice, protože sémantiku webového zdroje nelze jednoduše využít.

Cílem článku je uvést do technologií Sémantického webu podporujících vytváření DK. Požadavky (1)-(4) platí i pro DK. Řešení (3) a (4) tvoří pro integraci DK a webu dokonce smysluplnou nutnou podmínku.

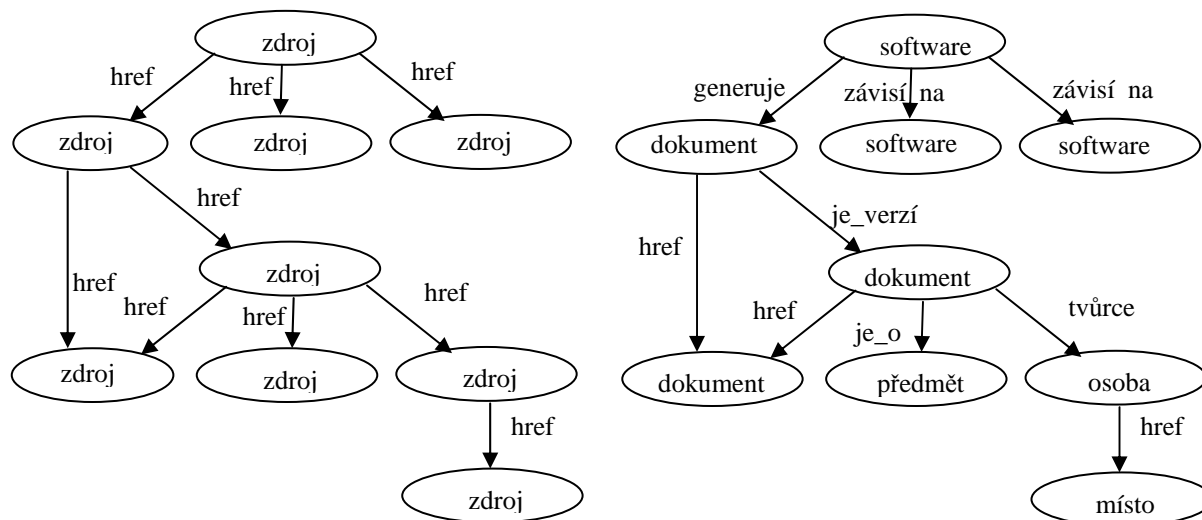
V sekci 2 stručně vysvětlíme základy Sémantického webu. Sekce 3 ukazuje, jak lze v jazycích jazyků RDF a RDFS reprezentovat metadata. O ontologiích pojednává sekce 4. Použití logiky a softwarových agentů, kteří zajišťují procesy spojené s vyhledáváním v Sémantickém webu, jsou zmíněny v sekci 5. V sekci 6 budeme diskutovat možnosti DK v prostředí Sémantického webu. V závěrech zmíníme některé trendy rozvoje DK.

## 2 Základy Sémantického webu

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. Současně je ovšem třeba vyvinout automatické prostředky, které mohou nalézat a

sbírat znalosti anotované na stránkách (obecněji zdrojích) s cílem porozumět těmto znalostem a využít je pro vyhodnocování dotazů. V takovém webu již nevyhledáváme stránky obsahující podobná slova, ale podobné pojmy. Za zmínku stojí, že o rozvoj Sémantického webu se částečně zasloužila i komunita rozvíjející DK, která měla dobré zkušenosti s vyhledáváním v klasických knihovnách a chtěla budoucí DK do webu vhodně zapojit.

Rozdíl mezi obyčejným webem a Sémantickým webem ukazuje obrázek 1 převzatý z [5]. Na rozdíl od modelu v HTML (levá strana obrázku), lze v Sémantickém webu rozlišit o jaké zdroje a vazby mezi nimi jde (pravá strana obrázku).



Obr. 1: Rozdíl mezi současným a Sémantickým webem

Jak je to se značkováním dat na webu? Nějaké již samozřejmě existuje. Napíšeme-li např. v HTML

<H2>Digitální knihovny v prostředí sémantického webu</H2>, specifikujeme, jak má vypadat prezentace daného řetězce na obrazovce. Značkování však nic nevypovídá o sémantice tohoto řetězce.

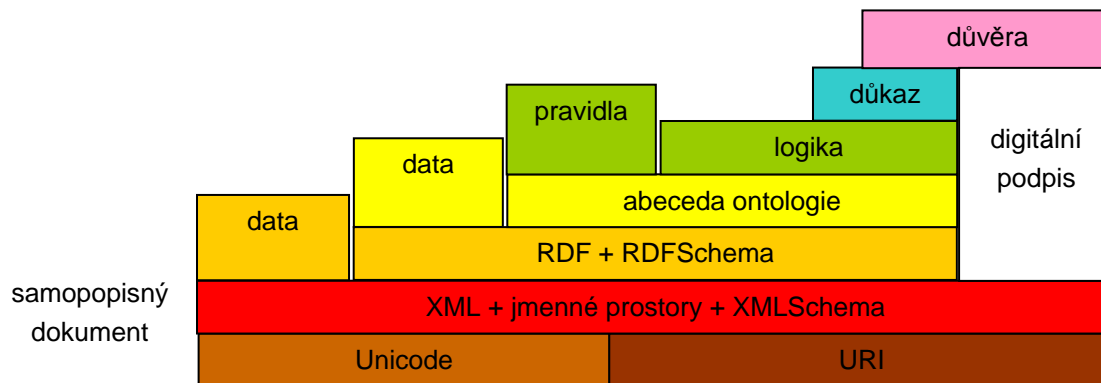
Sémantické značkování může být částečně dosaženo pomocí standardizovaného značkovacího jazyka XML [15]. Např. text v XML

```
<ADRESA>
  <JMÉNO ROD_Č = '480423016'>Jarda</JMÉNO>
  <ULICE>Svojetická 2</ULICE>
  <MĚSTO>Praha 10</MĚSTO>
</ADRESA>
```

obsahuje značky srozumitelné uživateli, nikoliv však stroji. XML je totiž určen zejména pro specifikaci struktury dokumentu. Určit jednoznačně a korektně význam XML značek nelze obvyklými prostředky řešit. XML a další jazyk XML Schema lze ale využít pro zápis metadat. Oba jazyky tak slouží k zajištění požadavku (3) na syntaktickou interoperabilitu. Není obtížné naplnit i požadavek (2). Díky tomu, že XML vlastně slouží pro definování gramatiky, cokoliv může být zakódováno v XML.

Z hlediska architektury webu je úspěch práce s webem postaven na interoperabilitě, která je v obou generacích webu založena zejména na protokolu TCP/IP, HTTP a HTML. Problémem tedy zůstává sémantická interoperabilita. K jejímu dosažení slouží metadata, která jsou modelována pomocí dalších standardizovaných speciálních jazyků RDF (Resource Description Framework) [7] a jazyk pro definici RDF schémat RDFS [3]. Dalšími

technologemi jsou ontologie, logika a inteligentní agenti. Budeme o nich hovořit v dalších odstavcích. Důležitým záměrem W3C je navrhnout a vytvořit všechny podpůrné prostředky jako standardy. 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. Pracují tedy s daty v RDF a XML. Pro vyhledávací stroje v prostředí Sémantického webu znamená tento přístup vznik nové kategorie *sémantických vyhledávacích strojů*, které tyto jazyky využívají.

Jak jsme již zmínili, podchycení a využití sémantiky zdrojů představuje pouze část problému stávajícího webu. Vícevrstvá architektura Sémantického webu je navržena tak, aby interoperabilita či dokonce sémantická interoperabilita mohla být realizována (viz požadavek (4)). Průkopnickou technologií založenou na interoperabilitě jsou dnes např. webové služby. Bohužel, současné nástroje pro realizaci webových služeb nejsou zatím pro zajištění sémantické interoperability dostačující.

Sémantický web lze 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 [13]. Protože se jedná o kvalitativní změnu ve vývoji webu, můžeme hovořit o třetí generaci webu.

### 3 Jazyky RDF a RDFS

RDF je druhem datového modelu. Základními objekty zájmu jsou tzv. *zdroje*. Pojem zdroj je ovšem míněn obecněji než pouhá webová stránka. Zdroj je identifikován pomocí *URI* (Uniform Resource Identifier). URI může být přiřazen 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 identifikaci URI kladený je, aby byla jednoznačná.

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

- *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.

- *Vlastnost* definuje binární vztah mezi zdroji a/nebo atomickými hodnotami (literály). Vlastnost může být zadána pomocí URI.
- *Objekt* specifikuje pro nějaký předmět hodnotu vlastnosti. Charakteristiky webových dat jsou tedy dány pomocí těchto objektů. Objektem je buď opět nějaký zdroj nebo atomická hodnota (literál).

Příkladem výroku v RDF je třeba

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

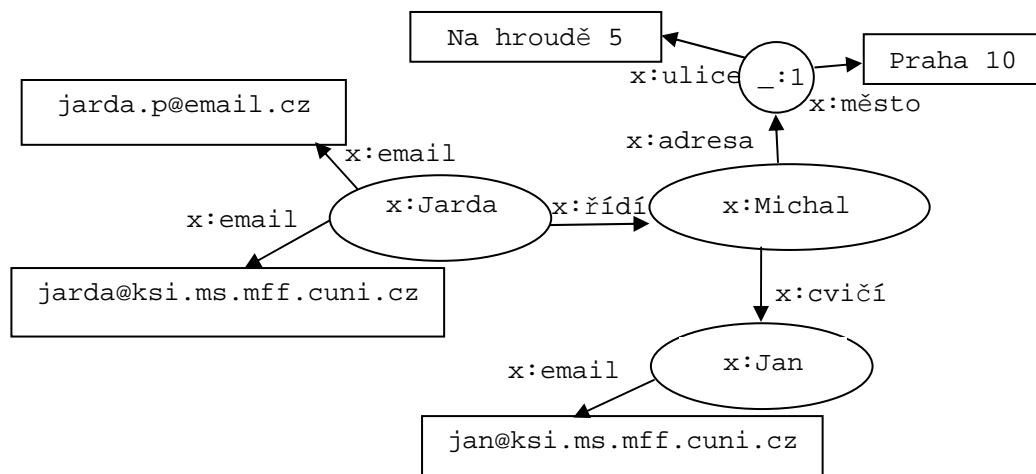
kde `osobní` je předpona označující hypotetický prostor jmen, který se týká personalistiky. V přirozeném jazyce odpovídá danému výroku věta: „Informační zdroj identifikovaný jako `osobní:Jarda` má emailovou adresu `jarda.p@mail.cz`“. Příkladem jiného jmenného prostoru může být standard Dublin Core [1] (česky také Dublinské jádro), používaný pro vytváření metadat dokumentů. Obsahuje elementy jako `TITLE`, `CREATOR`, `SUBJECT`, `DESCRIPTION`, `PUBLISHER`, `FORMAT` a několik dalších s definovaným významem. Význam je specifikován krátkou anglickou větou. Popíšeme-li tedy v trojici vlastnost `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 výhodné jejich zakódování v XML. Pro elementy RDF se používá předpona `rdf`, pro vyjádření prostorů jmen pomocí atributů XML předpona `xmlns` apod. Např. prostor jmen pro Dublin Core specifikuje jako XML atribut

```
xmlns:dc = 'http://purl.org/DC/' .
```

V grafickém vyjádření lze *RDF graf* popsat jistým speciálním grafem (viz obrázek 3). Ovály se používají pro zdroje, obdélníky pro literály a orientované označené hrany pro vlastnosti. Zdroje mohou mít více vlastností, literály se ve výrocih vyskytují pouze jako objekty.

*RDF-graf* rozšiřující náš příklad je na obrázku 3. Pomocí `x:` pro jednoduchost vyjadřujeme libovolné URI. 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.



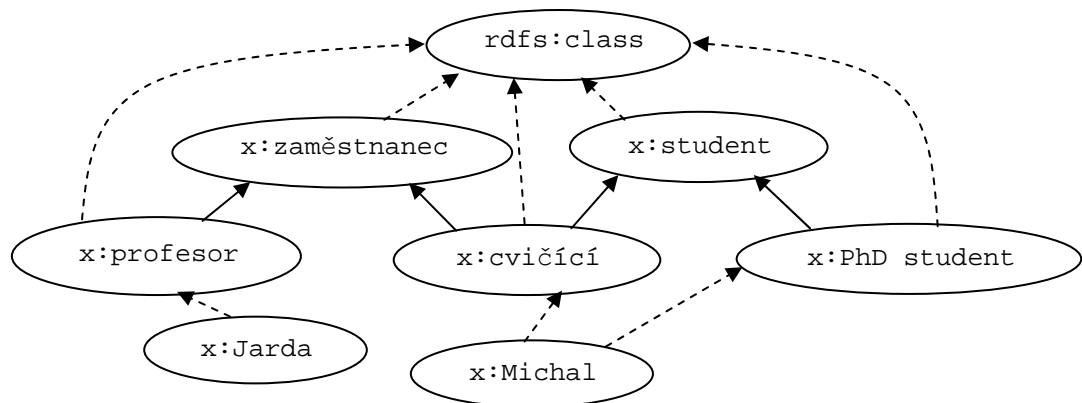
Obr. 3: RDF data graficky

Ukážeme, jak zakódovat tvrzení z obrázku 3 do XML. Využijeme dva prostory jmen, `rdf` (podle standardu RDF) a `uni` (podle schématu Univerzita), a jedno z možných zakódování – 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>
```

RDF data lze jednoduše slévat. 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.

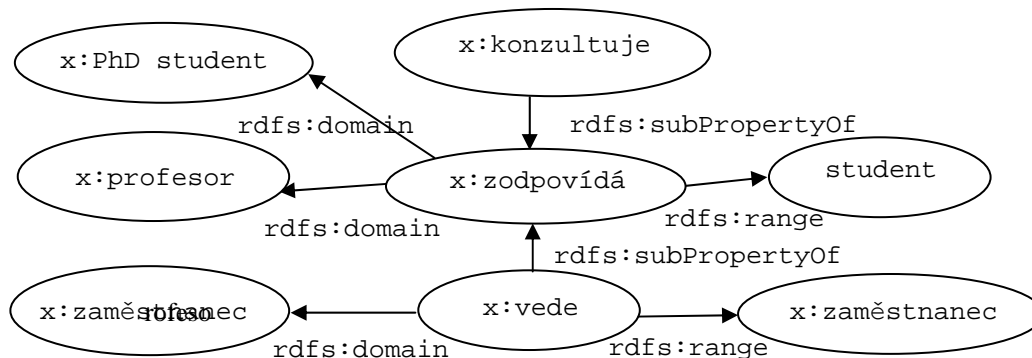
Dalším jazykem je RDF Schema (zkráceně RDFS). Pomocí RDFS se popisují koncepty použité v RDF datech a typová omezení na objekty a hodnoty ve trojicích. Takový popis lze po vzoru databází nazvat *RDF schématem*. Máme-li k dispozici nějaké RDF schéma, je možné RDF data kontrolovat vzhledem k tomuto schématu. Obrázek 4 ukazuje rozvinutí třídy `rdfs:class` (pojem RDFS), tj. její instance (např. třída `x:zaměstnanec`) resp. podtřídy tříd (`x:cvičící` je `x:zaměstnanec`)



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

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

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.



Obr. 5: Specifikace vlastností pomocí RDF

## 4 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 termín může mít ve dvou komunitách různý význam. Např. termín informatika může znamenat někde obor ukončení studia na vysoké škole, jinde jméno předmětu. Význam dat se dnes obvykle specifikuje 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 [8] pomáhají učinit sémantické značkování kompatibilní mezi jednotlivými informačními zdroji a dotazy. Z hlediska uživatelů ontologie to znamená mít prostředky pro společné chápání dané domény.

Tento přístup připomíná v prostředí knihoven známé pojmy jako tezaury či kategorie, pro komunitu od databází zase konceptuální modelování a integrace heterogenních databází. Ontologie a jejich sdílení znamená rozšíření těchto koncepcí do prostředí webu.

Ontologie jsou výsledkem dohody, což v praxi vyžaduje vyvinout hodně úsilí a rozsáhlé časové investice pro jejich vytvoření. Protože rozsah různosti dat na webu nebo v DK 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. učitel, student, předmět a přednáška. 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 jazycích RDF a RDFS. 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 (nebo zdroj) 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).

Perspektivní pro budování ontologií je kromě XML, XML Schéma a RDF/S i další sémantický značkovací jazyk pro webové zdroje DAML+OIL [14]. 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.

## 5 Logika a agenti

Další, vyšší úroveň Sémantického webu využívá jazyky logiky. Uvažujme následující množinu pravidel:

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

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 při integraci ontologií.

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:

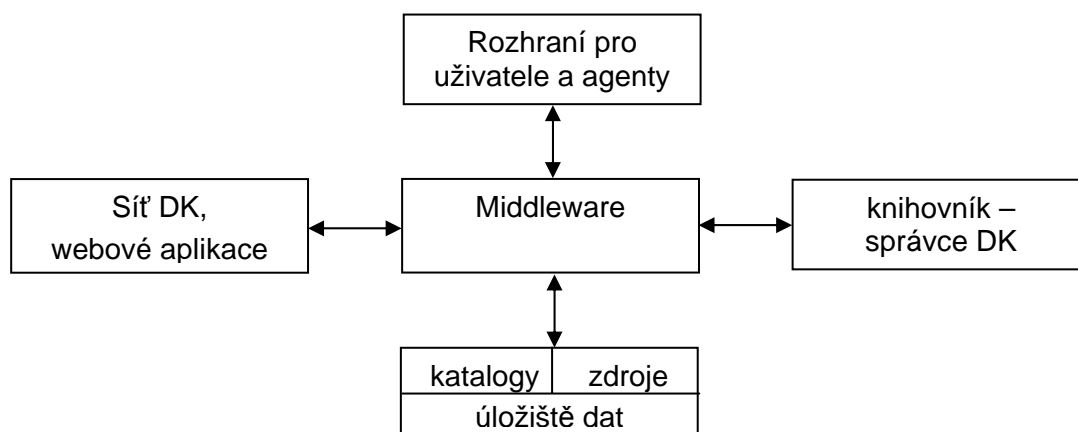
- metadata k nalezení a extrakci informací z webových zdrojů,
- ontologie k asistenci při vyhledávání na webu, interpretaci vybraných informací a ke komunikaci mezi agenty,
- logiku pro zpracování získaných informací a pro formulaci závěrů.

## 6 Digitální knihovna v prostředí Sémantického webu

Ukážeme možnosti DK v prostředí Sémantického webu na aplikaci obecné architektury DK prezentované v [4]. Na obrázku 6 jsou ukázány její tři vrstvy. Knihovník (správce DK) řídí DK po obsahové stránce (např. natahuje zdroje a jejich popisy do datového úložiště), udržuje katalogy. Nabídka zdroje od uživatele prochází přes knihovníka. Komunikační rozhraní (např. podle standardu současných webových služeb SOAP) dovoluje interoperabilitu mezi DK resp. mezi DK a dalšími webovými zdroji. Důležitá je flexibilita pro uživatele. Obsah zdrojů v dané DK může být díky technologii XML poskytován nejen v XHTML (XML verze HTML), ale i např. v PDF nebo RTF formátu. Middleware poskytuje možnosti pro vyhledávání, prohlížení katalogů, dále pak řízení zdrojů a uživatelů.

Pro vyhledávání je nutné použít více technologií a dokonce více modelů vyhledávání (Booleovský, vektorový, nepřesný (přibližný), využívající řazení podle relevance atd.). Kromě běžných fulltextových indexů a bibliografických popisů se ukazuje nutnost zapojit

ontologie [12]. Samotné klasické bibliografické popisy typu MARC nebo BibT<sub>E</sub>X nejsou pro strojové porozumění dostačující. Na druhé straně vytvářené ontologie s nimi musí být kompatibilní. Vzhledem k diverzitě zdrojů však zatím není cílem vytvářet pro DK obecné ontologie. Ontologický přístup se také zatím využívá hlavně pro anotace zdrojů a nikoliv celého jejich obsahu. Dobrým prvním krokem pro tvorbu ontologie může být lexikální databáze WordNet ve verzi pro český jazyk<sup>\*\*</sup>. Velkorysejší řešení ovšem vyžaduje důkladnější propojení sémantického webu a DK. Slibně se rozvíjí návrh iniciativy MarcOnt [6], která takovou ontologii rozvíjí. Pro řadu typů dokumentů či zdrojů dnes již také existují samostatné standardy pro metadata (viz např. [2]). RDF/S popisy mohou stavět na těchto standardech.



Obr. 6: Generická architektura DK

## 7 Další trendy v DK a závěry

Izolovaná DK, jakkoliv svázaná s webem, neznamena ještě cílové řešení. V klasickém pojetí sítě knihoven, kdy uživatel resp. knihovník, vyhledává v dostupných fondech, je taková DK pouhým prvním krokem. Již v druhé generaci webu byl tento problém řešen pomocí metavyhledávačů, kdy bylo možné jeden dotaz posunout a transformovat k více zdrojům. Tato metoda „hrubé síly“ však není pro uživatele DK příliš efektivní. Objevují se architektury typu peer-to-peer (P2P), a to přímo jako specializované řešení v síti digitálních knihoven. Nazývají se L2L (library-to-library). L2L architektury rozšiřují množství a kvalitu dat poskytovaných uživateli. Tím že komunikují přímo mezi sebou a ne prostřednictvím centrálního článku (serveru v P2P).

Specifickou alternativou DK je vytváření sociálních sítí založených na personalizaci knihoven. V jejím rámci lze řešit problém, kdy např. anotace zdroje (zdrojů) nevyhovuje potřebám uživatele (jiná ontologie) a chce používat vlastní. Pro personalizaci využívají DK ze Sémantického webu zcela specifické přístupy, jako je FOAF (Friend of a Friend). Projekt FOAF<sup>††</sup> poskytuje více než 20 užitečných termínů pro popis lidí tak, že jsou strojově čitelné a lze mezi nimi realizovat vazby. Do DK pak lze zapojit funkce realizující autentizační mechanismy (vstupuji-li já, může i můj přítel), přejímání anotací zdrojů v DK a jejich klasifikaci od přítele apod.

<sup>\*\*</sup> <http://nlp.fi.muni.cz/projekty/wnportal/index.php?akce=uvod>

<sup>††</sup> <http://www.foaf-project.org/>

Skončíme mottem R. Reddyho z univerzity Carnegie Mellon: „*Universal Access to All Human Knowledge*“. Zdá se, že Sémantický web podporuje tuto vizi. Na druhé straně teprve DK, využívající možnosti Sémantického webu, jí dávají reálnější obrysy.

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

## Použitá literatura a WWW odkazy

1. DUBLINCORE INITIATIVE. *Dublin Core Metadata Element Set, Version 1.1: Reference Description*. Dostupné na World Wide Web: <http://dublincore.org/documents/dces/>.
2. BRATKOVÁ, Eva. Metadata jako nový nástroj pro komunikaci webovských informačních zdrojů. *Národní knihovna. Knihovnická revue*. 1999, č. 4, s. 178-195. Dostupné na World Wide Web: <http://full.nkp.cz/nkkr/Nkkr9904/9904178.html>.
3. BRICKLEY, D., GUHA, R.V. *Resource Description Framework (RDF) Schema Specification 1.0. TR W3C, 2000. W3C Recommendation*. Dostupné na World Wide Web: <http://www.w3.org/TR/2000/CR-rdf-schema-2000372/>, 2000.
4. FREW, J., FREESTON, M., FREITAS, N., HILL, L. L., JANEY, G., LOVETTE, K., NIDIFFER, R., SMITH, T. R., ZHENG Q. The alexandria digital library architecture. In *Proceedings of the Second European Conference on Research and Advanced Technology for Digital Libraries*, Springer-Verlag, 1998, s. 61–73.
5. KOIVUNEN, M.-R., MILLER, E. W3C Semantic Web Activity. In HYVÖNNEN, E. (ed.). *Semantic web Kick-off in Finland*, 2002.
6. KRUK, S. R. Marcont initiative. Technical report, DERI. Galway, Ireland, Bibliographic description and related tools utilising Semantic Web technologies. Dostupné na World Wide Web: <http://www.marcont.org/>, 10 2004.
7. LASSILA, O., SWICK, R. R. Resource description framework (RDF) model a syntax specification. TR W3C, 1999. W3C Recommendation. Dostupné na World Wide Web: <http://www.w3.org/TR/REC-rdf-syntax>.
8. MENA, E., KASHYAP, V., ILLARRAMENDI, A., 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*.
9. POKORNÝ, J. Digitální knihovny: více problémů než očekávání? In RAMAJZLOVÁ, Barbora, TKAČÍKOVÁ, Daniela (sest.). *Automatizace knihovnických procesů – 9 : sborník z 9. ročníku semináře pořádaného ve dnech 15.–16. května 2003 v Liberci*. Praha : ČVUT, 2003, s. s. 51-57. ISBN 80-0102-738-4.
10. POKORNÝ, J. Web Searching and Information Retrieval, *Computing in Science & Engineering*, 2004, vol. 6, no. 4, s. 43-48.
11. SHAH, U., FININ, T., JOSHI, A., COST, R.S., MAYFIELD, J. Information Retrieval on Semantic Web. In *10th International Conf. on Information a Knowledge Management*. ACM Press, 2003.
12. WEINSTEIN, P. C., BIRMINGHAM, W. P. Creating ontological metadata for digital library content and services. *International Journal on Digital Libraries*, 2(1):20–37, October 1998. s. 1432-1300.
13. W3C. Semantic web. Dostupné na World Wide Web: <http://www.w3.org/2001/sw/>.
14. W3C. DAML+OIL web Ontology Language. Dostupné na World Wide Web: <http://www.w3.org/Submission/2001/12/>.
15. W3C. Extensible Markup Language (XML) 1.0 (Third Edition). Dostupné na World Wide Web: <http://www.w3.org/TR/REC-xml/>.