

Sémantický web

Ing. Vilém Sklenák, CSc.

Vysoká škola ekonomická, fakulta informatiky a statistiky,

katedra informačního a znalostního inženýrství

sklenak@vse.cz

1 Úvodem

Jednou ze světově proslulých knihoven je Babylónská knihovna. Tato knihovna je smyšlená a je ústředním motivem stejnojmenné povídky J. L. Borgese [3]. Babylónská knihovna je nesmírná, obsahuje všechny znalosti světa, ale přitom v ní působí jen málo knihovníků a obsah není nijak uspořádán. Pro nalezení čehokoli musí jedinec procházet mezi regály, přičemž ví, že hledané pravděpodobně někde existuje, ale současně neví, ve které z nekonečně mnoha šestiúhelníkových galerií je hledaná znalost uložena. Babylónská knihovna obsahuje vše, co kdy bylo nebo bude napsáno, ale ke svému bohatství nenabízí žádné přístupové metody (použijeme-li současnou terminologii). Povídka byla napsána v roce 1941, ale v mnohém lze spatřovat podobnost se současným webem. Borges říká: „*Když se rozhlásilo, že Knihovna obsahuje všechny knihy, vyvolalo to nejprve pocit podivného štěstí. Všem připadalo, jako by byli pány nedotčeného a tajného pokladu ...*“ [3, str. 68], ale později dodává „*nehoráznou naději přirozeně vystřídala nesmírná sklíčenost. Jistota, že některý regál v některém šestiúhelníku skrývá drahocenné knihy a že ty knihy jsou nedostupné, se zdála nesnesitelná*“ [3, str. 69].

World Wide Web (WWW, web) dramaticky zvýšil dostupnost informací. To ale současně neznamená, že by bylo jednodušší jejich vyhledávání, zpracování nebo vyhodnocování. Pomocí služby WWW je v současnosti dostupných více než 3 miliardy statických dokumentů, kromě nich je nutno brát do úvahy minimálně stejný objem dokumentů uložených v databázích, které mají WWW rozhraní (tzv. neviditelný web).

2 Krátce z historie

Vznik myšlenky a rozvoj základních principů sémantického webu není záležitostí posledních dvou let, jak by se mohlo na první pohled znát vzhledem k rostoucímu počtu publikací, konferencí, workshopů apod. Je však pravda, že k širší popularizaci sémantického webu došlo především zásluhou článku, který „otec“ webu T. Berners-Lee společně s dalšími spoluautory vydali v prestižním časopise Scientific American [2]. V té době však práce na sémantickém webu trvaly již několik let. Od počátku T. Berners-Lee zdůrazňoval, že „*sémantický web je rozšířením současného webu, jež datům přiřazuje přesný význam, díky kterému bude možná kooperace jak lidí, tak softwaru*“.

3 Základní jazyky

Nezbytným předpokladem pro strojové zpracování informací je zachycení *struktury* dat. Toho lze dosáhnout více způsoby. Například v databázích relačního typu je vyjádřena struktura dat tabulkami s atributy. V prostředí webu, v němž převažují dokumenty textového charakteru, se uplatňuje především *značkování dokumentů*. Tím se rozumí, že určité znakové sekvence obsahují informaci, která přisuzuje obsahu dokumentu určitou roli. Značky mají nejčastěji podobu slov uzavřených do lomených závorek, toto spojení se obvykle označuje jako *tagy*, např. `<title>` nebo `<h1>`.

Značkovací jazyk specifikuje sadu tagů, které mohou být používány, dále jak mohou být navzájem kombinovány, a konečně také jejich význam. Nejznámějším a nejrozšířenějším jazykem tohoto druhu v prostředí webu je jazyk HTML. Ten nabízí tagy pro abstraktní vzhled textových dokumentů. Například text vymezený `<h1>` a `</h1>` je považován za nadpis první úrovně, zatímco text určený tagy `` a `` je chápán jako položka seznamu. Tato informace slouží prohlížečům, aby byl text zformátován odpovídajícím způsobem.

Jazyk HTML dobře slouží účelu pro zprostředkování informací z hlediska výsledného vzhledu webových dokumentů. To ovšem nepostačuje potřebám pokročilých služeb, které jsou předpokládány v rámci sémantického webu. Tyto aplikace budou vyžadovat specializované značky, které budou vymezovat roli částí obsahu a to způsobem, který bude srozumitelný dané aplikaci. Například pro aplikaci, která bude porovnávat ceny zboží, by bylo užitečné, aby všechny ceny byly zapsány pomocí tagu typu `<cena>`. Proto je nezbytné specifikovat *význam* tagů.

3.1 XML

Extensible Markup Language (XML) je metajazyk, který nabízí možnost definovat nové značkovací tagy přesně podle specifických potřeb konkrétní aplikace. XML tak představuje mechanismus pro standardizovanou reprezentaci jiných jazyků. Jinak řečeno – XML poskytuje datový formát pro strukturované dokumenty, aniž by musel být specifikován platný slovník. Pro definici povolených značek a jejich možných kombinací se však obvykle takový slovník definuje a to buď jako *Document Type Definition (DTD)*, nebo nověji jako *XML schéma*. Jazyk, který je definován pomocí XML, se nazývá *XML aplikace*.

Jazyk XML neimplikuje specifickou interpretaci dat. Řekněme, že máme část XML kódu s tímto obsahem: `<jmeno>Jan</jmeno>`. Člověku je význam srozumitelný, a to i bez jakékoli formální specifikace. Za jedinou legitimní interpretací XML kódu lze považovat, že obsahuje pojmenované entity s případnými subentitami a jejich hodnoty. Každý XML dokument si lze představit jako uspořádaný strom. Tato obecnost, resp. neurčitost, je současně předností i slabinou jazyka XML. Zatímco je možné kódovat rozmanité typy datových struktur pomocí jednoznačné syntaxe, XML se nezabývá užitím dat a sémantikou. Pro použití jazyka XML jako formátu pro výměny dat se musejí obě strany předem domluvit na slovníku, použití a významech.

3.2 RDF a RDF Schema

Jazyk XML je standardní mechanismus pro *strukturu* dat. *Resource Description Framework (RDF)* představuje mechanismus pro *sdělení* něčeho o datech, tj. pro určení významu. Jak název napovídá, nejde o jazyk, ale o model pro reprezentaci dat o „věcech na webu“. Tento typ dat o datech se označuje jako *metadata*. „Věci“ jsou *zdroje* podle RDF terminologie. Za zdroje lze považovat vše identifikovatelné na webu, přičemž rozmezí sahá od specifických, identifikovaných části dokumentů až po osoby nebo firmy.

RDF datový model je poměrně jednoduchý. Obsahuje tvrzení o zdroji, které je vyjádřeno trojicí objekt-atribut-hodnota. Například tvrzení, že „Jan“ je autorem článku na určité URL lokaci, by mohlo být reprezentováno trojicí: <http://www...-autor-„Jan“>. Atributy, jako je např. „autor“, se nazývají *vlastnosti*.

RDF schéma (RDFS) představuje nadstavbu, která umožňuje vytvářet RDF struktury. RDFS nabízí mechanismus pro definování doménově specifických vlastností a tříd zdrojů, pro něž lze dané vlastnosti použít. Základní modelovací konstrukce v RDFS se skládají z: 1. definice *tříd* a *podtříd*, je možné vytvářet hierarchie tříd; 2. definice *vlastností* a *podvlastností*, opět s možností vytváření hierarchie; 3. *definiční obor* a *obor hodnot* z důvodu vyloučení nevhodných kombinací vlastností a tříd; 4. atribut *type*, který přiřazuje zdroj jako instanci specifické třídy.

RDF a RDFS společně nabízejí jednoduchý mechanismus reprezentace znalostí pro webové zdroje. Na druhou stranu je nutno uvést, že RDFS je velmi jednoduchý v porovnání s plnohodnotnými jazyky pro reprezentaci znalostí. Obsahuje pouze několik modelovacích možností a neumožňuje precizněji specifikovat podmínky příslušnosti ke třídám, rovněž postrádá datové typy aj.

4 Ontologie

Ontologie představují klíčovou technologii sémantického webu. Umožňují, aby použité symboly byly jednak srozumitelné člověku, jednak strojově zpracovatelné. Ontologie byly předmětem zkoumání v rámci umělé inteligence s cílem sdílení a opakovaného použití znalostí. Počátkem devadesátých let se ontologiím věnovaly různé výzkumné komunity, a to včetně znalostního inženýrství, zpracování přirozeného jazyka nebo reprezentace znalostí. Ještě dříve pronikly ontologie do takových oblastí, jako jsou inteligentní vyhledávání informací, integrace informací, kooperativní informační systémy, elektronické obchodování a knowledge management. Příčina popularity ontologií v tak rozmanitých oblastech tkví v tom, co nabízejí: sdílení a obecné porozumění předmětné oblasti ve formě vhodné pro komunikaci mezi člověkem a aplikacemi.

4.1 Co jsou ontologie?

Termín *ontologie* pochází původně z filozofie, ve které se používá jako označení pro *nauku o bytí*, tedy o tom „co je“. V souvislosti s vývojem databází a zejména rozsáhlýchází znalostí uložených v nesčetných souborech na webu došlo k posunu významu a především k jeho rozšíření. V novém významu se ontologií myslí zhruba označení pro všechny možné metody získávání znalostí „o tom, co je v datech“ v nejširším slova smyslu. Protože tyto metody jsou vždy spojeny s použitím jistého (dotazovacího) jazyka a s problémy spojenými s významem výrazů, pomocí nichž jsou data popisována a získávána, hovoří se proto o tzv. *konceptualizaci* jednotlivých oblastí, neboli o *dílčích ontologiích*.

V posledních deseti letech vznikla řada definic ontologie. Snad nejlépe vystihuje podstatu ontologií Gruberova definice: *Ontologie je formální, explicitní specifikace sdílené konceptualizace*. *Konceptualizace* říká, že se jedná o takový abstraktní model výseku reálného světa, který identifikuje relevantní koncepty daného výseku. *Explicitní* znamená, že je jednoznačně definován typ konceptu i podmínky jeho použití. *Formální* poukazuje na to, že ontologie by měla být strojově zpracovatelná. Je možná různá míra formality. Rozsáhlé ontologie, jako je třeba WordNet, obsahují více než 100 000 termínů, které jsou vyloženy pomocí přirozeného jazyka. Na druhé straně existují ontologie, které nabízejí formální axiomatické teorie. *Sdílený* odráží představu, že ontologie zachycuje konsensuální znalosti, tj. že není omezena na jedince, ale je akceptována šířeji.

4.2 Tezaurus vs. ontologie

Na první pohled nemusí být úplně zřejmý rozdíl mezi ontologií a tezaurem. Ztotožnění tezauru a ontologie je však mylné. Lze najít rysy společné, ale i odlišující. Hlavní podobnost spočívá v uspořádání terminologie a v použití stromové struktury. Jak ontologie, tak tezaurus se soustřeďují na terminologické pokrytí určité předmětné oblasti a na vymezení vztahů mezi jednotlivými termíny. Hierarchická struktura slouží k seskupení termínů do kategorií a podkategorií. Ontologie i tezaurus se používají pro popis a organizaci informací. Podstatný rozdíl spočívá v nejednoznačnostech vztahů v tezauru. Tezaurových vztahů (BT, NT, RT, UF apod.) není příliš mnoho a nejsou příliš formálně vymezeny, a proto jsou používány různě. Například vztah BT/NT (broader then/narrower then) může být použit pro vyjádření vztahu jednak, že určitý koncept je zvláštním případem jiného, jednak že koncept je částí jiného. Vztah RT (related to) zahrnuje všechny ostatní vztahy, a to bez rozlišení, zda se jedná o asociaci, nebo o jiný vágní vztah. Dobrá ontologie zavádí množství strukturních i konceptuálních vztahů včetně vztahů nadtřída/podtřída/instance, vztahů k času a dalších podle typu jazyka použitého pro reprezentaci. Obecně obsahuje ontologie mnohem více vztahů, které jsou formálně definovány a jsou jednoznačné.

Rozdílné jsou rovněž cíle, s nimiž jsou tezaury a ontologie vytvářeny. Tezaurus se snaží zachytit vztahy mezi *termíny*, zatímco u ontologie jde o definování *konceptů* a vztahů mezi nimi. V nejjistší podobě nejde u ontologie o termíny, ale pouze o koncepty, které jsou

reprezentovány ve formě nezávislé na termínech libovolného přirozeného jazyka. Tezaurus se nesnaží definovat, ale pouze formálně reprezentovat význam konceptů, soustřeďuje se na vztahy mezi termíny v určitém přirozeném jazyce (nebo více jazycích, jde-li o vícejazyčný tezaurus). Proto mechanismus reprezentace konceptů v ontologii musí být mnohem mohutnější. Nicméně i u ontologií je nutno použít termíny některého přirozeného jazyka, a proto musí ontologie zahrnovat mapování mezi termíny a koncepty, analogie s tezaurem již není.

Rozdíl mezi ontologií a tezaurem vynikne budeme-li uvažovat využití při vyhledávání. Protože ontologie obsahuje strojově interpretovatelnou definici konceptů, může podporovat terminologické odvozování. Tím se myslí, že dotaz uživatele může být analyzován z hlediska významu použitých termínů, a proto může být přesněji mapován vzhledem k informačním zdrojům. Pomocí ontologie lze např. odvodit, že jeden koncept je speciálním případem jiného, a to díky porovnání jejich logických definic. Je-li koncept B speciálním případem konceptu A, pak může být automaticky klasifikován jako podřízený konceptu A. To dává širší možnosti pro zpracování dotazu a vyhledávání.

4.3 Typy ontologií

Ontologie lze klasifikovat z různých hledisek, nejčastěji se používá rozdělení podle zdroje konceptualizace:

- **Generické ontologie** nebo **ontologie vyššího řádu** usilují o zachycení obecných zákonitostí, které platí napříč věcnými oblastmi.
- **Doménové ontologie** jsou nejčastější typ, předmětem je vždy určitá specifická věcná oblast.
- **Úlohové ontologie (reprezentační, metaontologie)** označují generické modely znalostních úloh a metod jejich řešení. Oproti ostatním ontologiím, které zachycují znalosti o světě, se zaměřují na procesy odvozování.
- **Aplikační ontologie** jsou nejspeciřtější a jde o konglomerát modelů adaptovaných na konkrétní aplikaci, zahrnující zpravidla doménovou i úlohovou část.

WordNet [15] je příkladem terminologické ontologie, která původně vznikla pro angličtinu, ale nyní již existují i vícejazyčné varianty, jako je např. *EuroWordNet* [8]. Mezi slovy je definována řada vztahů. Obecně známé jsou vztahy synonymie a antonymie. Synonyma jsou považována za ekvivalentní, a proto základní jednotkou *WordNetu* jsou množiny synonym, které se označují jako *synsety*. Výsledek vyhledávání slova „search“ ukazuje obr. 1.

Bylo nalezeno celkem 9 významů, z nichž u pěti se jedná o podstatná jména a čtyři jsou slovesa.

WordNet sleduje řadu vztahů pro vyjádření „blízkosti“. Například lze uvést *hyperonyma* (viz obr. 2) a *hyponyma* (viz obr. 3) pro podstatná jména, nebo *troponyma* pro

The noun "search" has 5 senses in WordNet.

1. search, hunt, hunting – (the activity of looking thoroughly in order to find something or someone)
2. search – (an investigation seeking answers; "a thorough search of the ledgers revealed nothing"; "the outcome justified the search")
3. search, lookup – (an operation that determines whether one or more of a set of items has a specified property; "they wrote a program to do a table lookup")
4. search – (the examination of alternative hypotheses; "his search for a move that would avoid checkmate was unsuccessful")
5. search – (boarding and inspecting a ship on the high seas; "right of search")

The verb "search" has 4 senses in WordNet.

1. search, seek, look for – (try to locate or discover, or try to establish the existence of; "The police are searching for clues"; "They are searching for the missing man in the entire county")
2. search, look – (search or seek; "We looked all day and finally found the child in the forest"; "Look elsewhere for the perfect gift!")
3. research, search, explore – (inquire into)
4. search – (subject to a search; "The police searched the suspect"; "We searched the whole house for the missing keys")

Obrázek 1: Výsledek vyhledávání „search“ ve WordNetu

slovesa. Jiný sledovaný typ vztahů je vztah mezi částí a celkem, čili jde o *meronyma* a *holonyma*.

search, lookup – (an operation that determines whether one or more of a set of items has a specified property; "they wrote a program to do a table lookup")

=> operation – ((computer science) data processing in which the result is completely specified by a rule (especially the processing that results from a single instruction); "it can perform millions of operations per second")

=> data processing – ((computer science) the temporary retention of data in a processor while calculations or logical operations are performed on it)

=> processing – (preparing or putting through a prescribed procedure; "the processing of newly arrived immigrants"; "the processing of ore to obtain minerals")

=> process – (a sustained phenomenon or one marked by gradual changes through a series of states; "events now in process"; "the process of calcification begins later for boys than for girls")

=> phenomenon – (any state or process known through the senses rather than by intuition or reasoning)

Obrázek 2: Hyperonyma pro „search“ ve třetím významu podstatného jména ve WordNetu

search, hunt, hunting – (the activity of looking thoroughly in order to find something or someone)

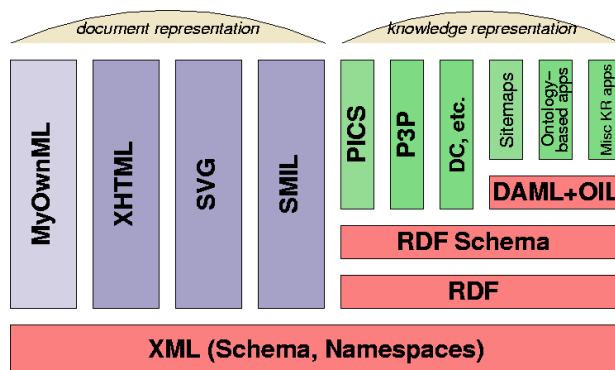
- => exploration – (a careful systematic search)
- => foraging, forage – (the act of searching for food and rovisions)
- => frisk, frisking – (the act of searching someone for concealed weapons or illegal drugs)
- => looking, looking for – (the act of searching visually)
- => manhunt – (an organized search (by police) for a person (charged with a crime))
- => quest, seeking – (the act of searching for something; ”a quest for diamonds”)
- => ransacking, rummage – (a thorough search for something (often causing disorder or confusion); ”he gave the attic a good rummage but couldn’t find his skis”)
- => scouring – (moving over territory to search for something; ”scouring the entire area revealed nothing”)

Obrázek 3: Hyponyma pro „search“ v prvním významu podstatného jména ve WordNetu

Při striktním posouzení není WordNet „čistou“ ontologií, ale dobře slouží jako prostředek pro zkoumání přirozeného jazyka.

4.4 Jazyky pro zachycení ontologií

Jazyk RDFS byl úmyslně navržen jako minimalistický. Z pohledu sémantického webu ne nabízí dostatečný potenciál k vytváření přesných ontologií. Tento nedostatek se snaží napravit další vrstva (viz obr. 4) – jazyk *DAML+OIL*, který vznikl spojením jazyků *DAML* (*DARPA Agent Markup Language*) a *OIL* (*Ontology Inference Layer*).



Obrázek 4: Jazyky sémantického webu

Jazyk DAM+OIL není jediný, který se používá pro reprezentaci ontologií. Měl historické předchůdce: *Cyc*, *Ontolingua*, *OCML*, *OKBC*, *XOL*, *SHOE*, *Ontobroker* [14]. Na druhou

stranu po zkušenostech s jazykem DAML+OIL vzniká jazyk nový: *OWL (Ontology Web Language)*, který je vyvíjen pod záštitou konsorcia W3C.

5 Sémantický web a vyhledávání informací

Současné služby pro vyhledávání informací ve webovém prostředí jsou v zásadě dvojího typu:

- **Vyhledávací stroje s roboty** umožňují zpracovat velké množství dokumentů (až miliardy), které jsou nalézány robotem. Z dokumentů se vytváří fulltextový index. Služby tohoto typu zvyšují úplnost vyhledávání na úkor přesnosti.
- **Vyhledávací služby s asistencí člověka** jsou založeny na hierarchické klasifikaci dokumentů/zdrojů. Úzkým místem je způsob přiřazování dokumentů k jednotlivým kategoriím. Na druhou stranu jejich výhodou je vysoká přesnost, nevýhodou je pak nízká úplnost.

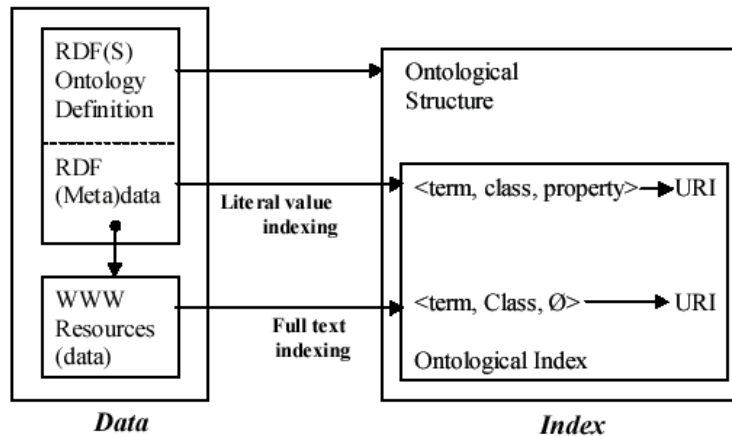
Lze předpokládat, že jednou z velmi žádaných služeb pro prostředí sémantického webu budou právě služby vyhledávání informací. Ty budou samozřejmě vycházet z toho, že webové stránky budou obsahovat text společně se sémantickými značkami.

Na základě sémantických značek bude možné realizovat různé typy vyhledávání [9]:

- **Vyhledávání informací (information retrieval, IR)** – identifikace relevantních dokumentů a jejich řazení podle míry vhodnosti.
- **Jednoduché odpovědi na otázky (simple question answering, Q&A)** – např. „Kdo je prezidentem České republiky?“.
- **Komplexní odpovědi na otázky (complex question answering)** – např. „Jaká je současná situace v Egyptě?“.

U všech typů lze očekávat zvýšenou efektivnost vyhledávání. Současně platí, že u všech typů vyhledávání budou používány různé techniky usuzování a odvozování, u komplexního typu Q&A jsou navíc zapotřebí techniky extrakce a sumarizace informací.

Současné webové vyhledávací služby, jako je třeba Google, vycházejí z teoretických poznatků i praktických zkušeností, jež jsou výsledkem desítky let trvajícího rozvoje v oblasti Information Retrieval. Zahrnutí sémantiky do vyhledávání, práce s ontologiemi si však žádá řadu modifikací i nových přístupů, a to ve všech fázích zpracování informací. Počínaje popisem, přes shromažďování a indexování až k samotnému procesu vyhledávání. Například obr. 5 ukazuje modifikovanou strukturu indexu. Jiné problémy se naskýtají s použitím ontologií [12]. Například jde o posouzení podobnosti dokumentů, které používají trochu odlišné ontologie, nebo problém dotazovacích jazyků, aby bylo možné zadávat



Obrázek 5: Struktura ontologického indexu

dotazy strukturovaně i nestrukturovaně, ... Existuje ještě mnoho otevřených otázek, než abychom mohli očekávat masový nástup služeb inteligentního vyhledávání informací. Na druhou stranu se však v některých projektech ověřují dílčí řešení, např. lze uvést *QuizRDF* popsaný v [5], nebo *HOWLIR* [9].

6 Nástup sémantického webu

V posledních letech bylo dosaženo značného pokroku ohledně různých technologií nebo jazyků nutných pro fungování sémantického webu. Na druhou stranu se však odborníci nedokáží shodnout o způsobu jeho zavádění – tedy zda se vydat revoluční nebo evoluční cestou.

WWW vděčí za své rozšíření do značné míry jednoduchosti a průhlednosti jazyka HTML, která umožňovala i začátečníkovi vytvořit během velmi krátké doby funkční webovou stránku (např. cestou znovupoužití existujícího zdrojového kódu). Sémantický web naproti tomu předpokládá u autorů stránek systematické vkládání metadat RDF, a tudíž aspoň základní znalost konstruktů tohoto jazyka, který není úplně jednoduchý.

Jeden názor doporučuje netrvat pro první fázi na sémantické interoperabilitě (globálně závazných ontologiích) (viz třeba [7]). Tím se uvolní prostor pro *lokální* (tj. proprietární) sémantické značkování, realizované pomocí elementů XML se „samovysvětlujícími“ názvy, vnořenými do běžného HTML (viz obr. 6). To povede nejprve ke vzniku „místních sémantických webů“ univerzit či firem a teprve později se na sebe příslušné lokální ontologie budou mapovat, a tím dojde k propojení do většího celku.

I druhá skupina názorů uznává nutnost udržování informací ve formátu HTML. Prostě přidávání metadat RDF by podle těchto názorů však vedlo k duplicitě informací a riziku

```

<html xmlns:uw="http://wash.edu/schema/example">
<uw:course about="http://wash.edu/courses/692">
<h1><uw:name>Networking Seminar</uw:name></h1>
<p>Office hours for additional assistance:
<uw:instructor><uw:name><b>Prof.</b> John Fitz</uw:name>
<uw:workPhone>543-6158</uw:workPhone></uw:instructor>
<uw:instructor><uw:name><b>Prof.</b> Helen Long</uw:name>
<uw:workPhone>543-8312</uw:workPhone></uw:instructor>
<table> <tr><th>2003 Schedule</tr>
<uw:reglist=<tr><uw:event><td><uw:date>*</uw:date><td><uw:topic>*</uw:topic>
</uw:event></tr>>
<tr> <td>Jan 11 <td>Packet loss</tr>
<tr> <td>Jan 18 <td>TCP theory</tr>
</uw:reglist>
</table>
</uw:course></html>

```

Obrázek 6: Sémantické značkování v rámci HTML kódu (převzato z [7])

nekonzistence. Jako částečné řešení je proto navrhována tvorba informačního obsahu pomocí jazyka XML s využitím jednoduchých ontologií a šablon. Podstatné je, že viditelným výstupem celého procesu jsou HTML stránky (s vysokou mírou sjednocení stylu) a vedle toho automaticky vznikají RDF metadata, ovšem jako vedlejší produkt, čili to nestojí uživatele žádné úsilí.

7 Závěr

Rozvoj sémantického webu od počátku vzbuzoval řadu nadějí. Například již v roce 1999 se Ch. Sherman ve svých úvahách o budoucnosti vyhledávání na webu zmiňuje o sémantickém webu jako o možném řešení některých problémů nejednoznačného a neúplného vyhledávání [13] ... A jeho tvrzení platí i dnes.

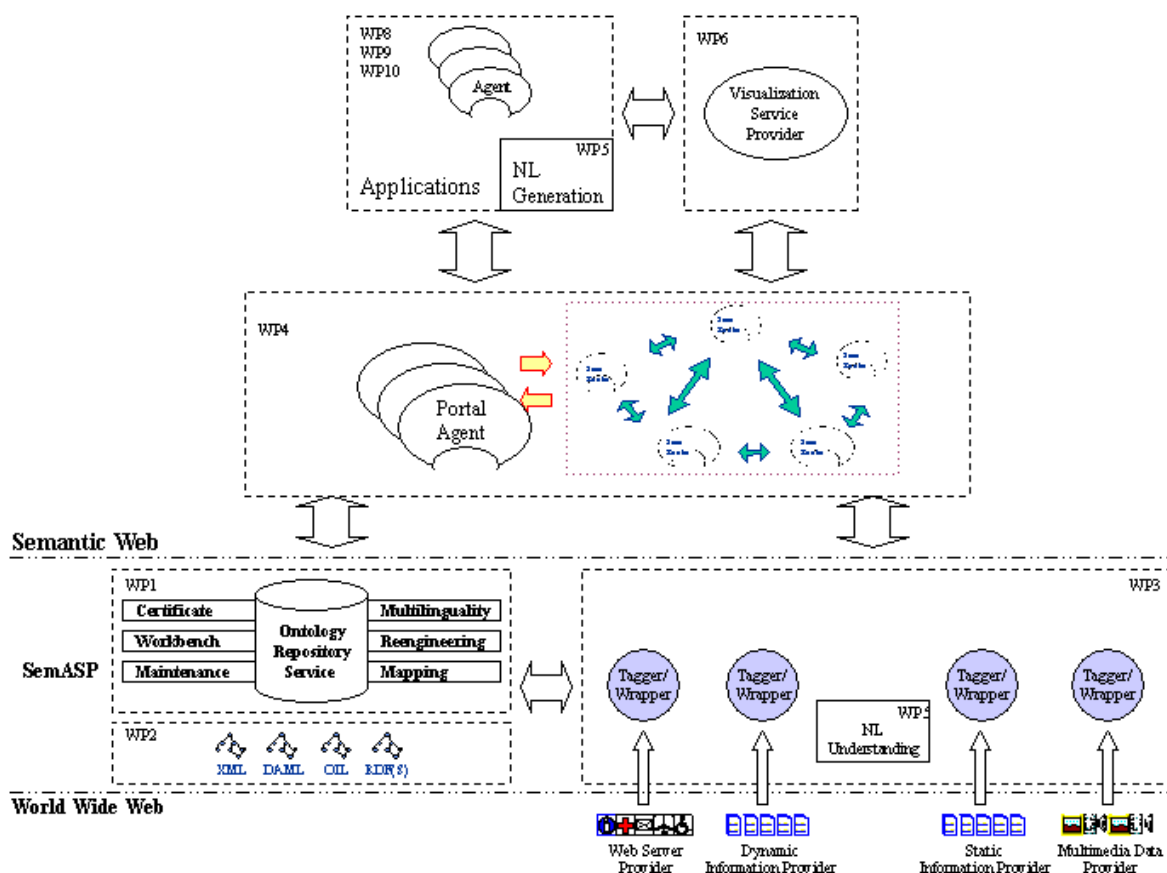
Sémantický web má za sebou již několikaletý výzkum, ale přesto se nachází v počátečním údobí svého vývoje. Mnoho bylo uděláno v oblasti jazyků, reprezentace znalostí, ontologií. Další krok, který bude následovat, je vývoj aktivních komponentů, které budou nabízet uživatelům inteligentní služby. V první řadě půjde o webové služby pro přístup k informacím a e-business. Podstatný bude rozvoj *UDDI (Universal Description, Discovery and Integration of Web Services)*¹ pro adresáře poskytovatelů, produktů a služeb. Pro jejich popis se používá jazyk *WSDL (Web Services Description Language)*² a protokol *SOAP (Simple Object Access Protocol)*³ pak pro přístup k nim. Existuje řada projektů, které se

¹<http://www.uddi.org>

²<http://www.wSDL.org>

³<http://www.soap.org>

snaží sémantické principy uplatňovat na úrovni systému. Architektura takových systémů je poměrně složitá, příklad možného řešení ukazuje obr. 7, jde o návrh v rámci evropského projektu *Esperanto*⁴.



Obrázek 7: Architektura sémantického webu podle projektu Esperanto

Sémantický web je vizí, která je postupně naplňována a která je stále nadějí.

Reference

- [1] Beck, Howard, Pinto, Helen Sofia. *Overview of Approach, Methodologies, Standards, and Tools for Ontologies*. UN FAO 2002. Dostupné na WWW: <http://www.fao.org/agris/aos/Documents/BackgroundPaper.pdf>.

⁴<http://www.esperanto.net>

- [2] Berners-Lee, Tim, Hendler, James, Lassilla, Ora. The Semantic Web. *Scientific American*, 2001. vol. 284, no. May. str. 35–43. Dostupné také na WWW: <http://www.sciam.com/2001/0501issue/0501berners-lee.html>.
- [3] Borges, Jose Luis. Babylónská knihovna. In *Zrcadlo a maska*. Odeon : Praha 1989, str. 64–72. Dostupné také na WWW: http://www.analitica.com/bitblioteca/jjborges/library_babel.asp.
- [4] Colomb, Robert M. *Information Spaces: The Architecture of Cyberspace*. Springer : London 2002.
- [5] Davies, John, Fensel, Dieter, van Harmelen, Frank. *Towards the Semantic Web: Ontology Driven Knowledge Management*. John Wiley : Chichester 2003.
- [6] Ding, Ying, Fensel, Dieter, Klein, Michel, Omelayenko. The Semantic Web: Yet Another Hip? *Data and Knowledge Engineering*, 2002. vol. 41, no. 3. str. 205–227. URL: <http://citeseer.nj.nec.com/ding02semantic.html>.
- [7] Etzioni, Oren et al. *Evolving the Semantic Web with Mangrove*. UW Tech Report February 2003. URL: <http://www.cs.washington.edu/research/semweb/pubs/mangroveTechFeb2003.pdf>.
- [8] EuroWordNet. URL: <http://www.illc.uva.nl/EuroWordNet/>.
- [9] Finin, Tim, Shah, Urvi, Mayfield, James. Information Retrieval on the Semantic Web. In *0th International Conference on Information and Knowledge Management*. 2002. URL: <http://www.cs.umbc.edu/~finin/papers/cikm02/cikm02.pdf>.
- [10] Greenberg, Jane, Sutton, Stuart, Campbell, D. Grant. Metadata: A Fundamental Component of the Semantic Web. *Bulletin of the American Society for Information Science and Technology*, 2003, vol. 29, no. 4., str. 16–18. URL: <http://www.asis.org/Bulletin/Apr-03/BulletinAprMay03.pdf>.
- [11] Jacob, Elin K. Ontologies and the Semantic Web. *Bulletin of the American Society for Information Science and Technology*, 2003, vol. 29, no. 4., str. 19–22. URL: <http://www.asis.org/Bulletin/Apr-03/BulletinAprMay03.pdf>.
- [12] Mayfield, James. Ontologies and text retrieval. *The Knowledge Engineering Review*, 2002, vol. 17, no. 1., str. 71–75.
- [13] Sherman, Chris. The Future of Web Search. *Online*, 1999, vol. 23, no. 3. str. 54–61. Dostupné také na WWW: http://www.findarticles.com/cf_0/m1388/3_23/54474833/print.jhtml.
- [14] Svátek, Vojtěch. *Ontologie a WWW*. Datakon 2002. URL: <http://nb.vse.cz/~svatek/onto-www.doc>.

- [15] WordNet: A Lexical Database for the English Language. URL: <http://www.cogsci.princeton.edu/~wn/>.