

Sémantický web - úvod

Doc. Ing. Vojtěch Svátek, Dr.

Letní semestr 2011

<http://nb.vse.cz/~svatek/rzzw.html>

Struktura přednášky

- Současné webové aplikace vs. sémantický web
- Specifika sémantického webu pro zpracování dat
- Úskalí a omezení sémantického webu

Webové aplikace dnes

- Vysoce specializované
- Často sdružují velký počet funkcionalit
 - využívání externích API a **datových zdrojů**
 - možnost jejich snadného (?) kombinování formou mashupů
 - někdy i sofistikované „inteligentní“ aplikace využívající **odvozování** (pravidla, taxonomie, fuzzy logika, ...)
- Ale
 - Propojení více zdrojů probíhá v aplikaci samotné (procedury nad RDB; transformace XSLT; ...)
 - „Inteligentní“ chování závisí na aplikaci samotné

Co nabízí sémantický web (1)

- Aplikace mohou pracovat s daty **propojenými** pomocí URI již v rámci webové **infrastruktury**
 - Úspora práce, předem vyřešené nekonzistence, automaticky propagované změny
- Datový formát je **standardizovaný**, **flexibilní**, a odvolává se na respektované sdílené **slovníky** (vocabularies)



DOAP



- Data jsou (do jisté míry) zpracovatelná i při nepředpokládané struktuře

Co nabízí sémantický web (2)

- Aplikace mohou využívat strukturované **znalosti** ve **standardních** formátech – **ontologie** jako sofistikovanější variantu slovníků
 - Lze **odvozovat** pomocí běžně dostupných nástrojů (lokálně nebo přes webová API)
 - Samotné znalosti je možné znovupoužít bez dodatečných nákladů

Konkrétní technologie

- Transparentní propojování dat zajišťují principy **Linked Data** (viz 4.týden)
- Flexibilní (síťový) datový formát je **RDF** (viz 2. a 3. týden)
- Bohaté odvozovací možnosti poskytuje ontologický jazyk **OWL** (Web Ontology Language, viz od cca 7. týdne dále)

Specifika (sémantického) webu pro tvorbu a zpracování dat

- „Kdokoli může říci cokoli o čemkoli“ (AAA)
- Síťový efekt
- Předpoklad otevřeného světa (OWA)
- Neplatí předpoklad jednoznačného pojmenování entit (non-UNA)

Princip AAA

- „Anyone can say Anything about Any topic“
 - Distribuovaná tvorba a zpracování dat
 - Absence kontroly nad přidávaným obsahem
- Vede k dostupnosti informací reflektujících
 - Rozporný postoj („Pluto pro astronomy a astrology“)
 - Úmyslný podvod
 - Nevědomý omyl
 - Starší, již neplatný stav věcí
- Lze řešit pouze v prostředí, které netrvá na jedinečných identifikátorech pro každý zdroj...

Síťový efekt

- S nárůstem počtu datových zdrojů v síti sémantického webu roste počet aplikací, které tato data zpracovávají
- ... a naopak

Předpoklad otevřeného světa - OWA

- Musíme vždy předpokládat existenci nových faktů (a entit), o kterých zatím nevíme
 - Na rozdíl od databázových systémů, které vesměs pracují s předpokladem uzavřeného světa – CWA
- Např. výčty entit můžeme málokdy chápat jako konečné

Non-UNA

- **Stejný zdroj** může na sémantickém webu existovat pod **různými identifikátory**
- Dodatečně lze zdroje prohlásit za totožné
 - Jedna z nejdůležitějších operací v současných aplikacích sémantického webu (resp. Linked Data) !
- Nutno brát do úvahy při odvozování

Úskalí sémantického webu

- Daní za flexibilitu datového formátu je
 - Horší škálovatelnost aplikací při nárůstu velikosti dat
 - Náročnější návrh uživatelských rozhraní pro zobrazování/editaci (vs. RDB, XML nebo fulltext)
- Odvozování často není pro běžného uživatele intuitivní
 - Může vést k neočekávanému chování aplikací
 - V generické podobě často neúnosně pomalé
- Proto i do budoucna zřejmě pro komerční sféru spíš jako „niche“ technologie, ale respektovaná