

Sémantický web jako základ pro vývoj aplikací

Doc. Ing. Vojtěch Svátek, Dr.

Zimní semestr 2015

<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 kombinování zejména 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 – zejména **ontologie** jako sofistikovanější variantu slovníků
 - Lze **odvozovat** pomocí běžně dostupných nástrojů (lokálně nebo přes webová API)
 - Samotné znalosti je možné přepoužít bez dodatečných nákladů

Konkrétní technologie

- Flexibilní (síťový) datový formát je **RDF**
- Srozumitelné publikování a propojování dat zajišťují principy **Linked Data**
- Základní odvozovací postupy nabízí **RDFS**
 - Bohatší odvozovací možnosti poskytuje **OWL** (Web Ontology Language)

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...

Sít'ový efekt

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

Předpoklad otevřeného světa

- „Open-world assumption“ (OWA)
- Lze vyjádřit zhruba takto:
v situaci kdy nevíme, jestli je nějaké tvrzení pravdivé, s ním nemůžeme pracovat tak, jako kdyby bylo nepravdivé
 - Opakem je předpoklad uzavřeného světa (CWA), běžně používaný v databázových systémech
- Na webu musíme téměř vždy předpokládat existenci relevantních faktů (a objektů), o kterých jen zatím nevíme

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
 - Např.: „Jan má syna Karla. Jan má syna Hynka. Kolik má Jan synů?“

Úskalí dnešního sém. webu

- Data jsou chybová a neúplná, je málo externích propojení, není dořešena otázka dynamické aktualizace
- Daní za flexibilitu formátu RDF je horší škálovatelnost aplikací a náročnější návrh uživatelských rozhraní
- Pokud se aplikuje strojové odvozování, může vést k neočekávanému chování aplikací, a v generické podobě je navíc na větších datech neúnosně pomalé
- Proto je technologie RDF v současnosti komerční sférou
 - v úplnosti aplikovaná jen v jednotlivých případech, které pro firmu nejsou kritické
 - ostře sledovaná, s tím, že si firmy ze standardů a akademických prototypů „vyzobávají“ nápady pro svá vlastní ad hoc řešení (a nabírají z univerzit lidi s tímto profilem)