

4IZ440 Reprezentace a zpracování znalostí na WWW

Motivační úvod

Doc. Ing. Vojtěch Svátek, Dr.

Zimní semestr 2012

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

Hlavní téma

- „Datový web“, „Entitní web“
 - web vytvořený ze strukturovaných dat (o reálných objektech - entitách) místo z textových dokumentů
- „Linked data“ („Propojená data“)
 - přes web se propojí dosud uzavřená „datová sila“ (vně i uvnitř) různých organizací
- „Sémantický web“, „Web se sémantikou“
 - termín používaný hlavně v akademické sféře, kde navazuje na výzkum umělé inteligence

Hlavní řešené úlohy

- Konverze dat do „jednotného“ formátu
 - Standardizovaný datový formát (RDF)
 - Různá datová schémata (slovníky, ontologie)
- Propojení („linking“) na jiná data
- Vystavení (publikování) dat
 - Pro lidi (HTML) i aplikace
- Vizualizace dat a tvorba aplikací nad nimi
 - Využití standardizovaného dotazovacího jazyka (SPARQL)
 - Data mining

Aktuální scénář 1

- Anotované nabídky v e-commerce
 - Umožní zobrazení strukturovaných dat v rámci výsledků vyhledávání a sociálních sítí (+kvalitnější vyhledávání)
 - Nejpoužívanější schémata
 - Schema.org (Google, Yahoo!, MS) – formát mikrodata (HTML5)
 - OpenGraph (FaceBook) – formát RDF
 - GoodRelations - formát RDF
- <http://www.rachaelraystore.com/Product/detail/Rachael-Ray-Stoneware-2-pc-Bubble-Brown-Baker-Set-Eggplant/316398>
- <http://www.google.com/webmasters/tools/richsnippets?url=http%3A%2F%2Fwww.rachaelraystore.com%2FProduct%2Fdetail%2FRachael-Ray-Stoneware-2-pc-Bubble-Brown-Baker-Set-Eggplant%2F316398&view=>
- Pluginy do e-shopových aplikací
 - Perspektivně i inteligentní komparativní nakupování a doporučování produktů

Aktuální scénář 2

- Vyhledávání a navigace v oblasti zábavy a zájmové činnosti
- Data vznikají z velké části na dobrovolné bázi (wiki model)
- DBpedia <http://dbpedia.org>
 - Wikipedia pravidelně převáděná do strukturované podoby
 - Aplikace např. DBpedia Mobile
<http://mes-semantics.com/DBpediaMobile/?location=Berlin>
- BBC WildLifeFinder
 - <http://www.bbc.co.uk/nature/wildlife>
 - Využívá Wikipedii/DBpedii jako svůj „CMS“

Aktuální scénář 3

- Otevřená data veřejné správy
 - USA (<http://data.gov>), Velká Británie (<http://data.gov.uk>), EU (<http://publicdata.eu>), aj.
- Rychlý vývoj aplikací pro „hlídací psy demokracie“ (neziskovky, novináři apod.)
- Příklad: vizualizace vybraných oblastí výzkumu ve Velké Británii: <http://bis.clients.talis.com>

Aktuální scénář 4

- Vnitropodniková integrace
- Výhodou využití konceptu „datového intranetu“ na bázi RDF je možnost obohacení o data z externích (zejména veřejnosprávních) zdrojů
- Prozatím zpravidla interní integrace řešena konvenčně, a externí data poskytuje specializovaná firma ve formě „software jako služby“
 - Jednoduchá API, komplexita (model RDF atd.) je umístěna mimo dohled klienta

Linked Data Cloud

- Přidaná hodnota vzniká především tehdy, když se datové zdroje propojí
- <http://linkeddata.org/>
- Tím vzniká analogie běžného webu
 - Indexování – nástroj Sindice
 - Browsing – nástroj Sig.ma <http://sig.ma/>

Vybrané aktivity v ČR

- Účast VŠE v projektu EU LOD2 <http://lod2.eu>
 - 2010-2014, 15 partnerů, rozpočet cca 9,5 ME
 - VŠE koordinuje pracovní balíček pro oblast veřejných zakázek
- Iniciativa <http://opendata.cz/>
 - Zaměření do praxe, zejména na veřejnou správu
 - Úzká spolupráce s MFF UK (za VŠE: KIZI a KIT)
- Iniciativa <http://semanti-cs.org>
 - Zaměření na akademickou sféru
- Další související výzkumné projekty KIZI
 - EU LinkedTV, 2011-2015 <http://www.linkedtv.eu/>
 - GAČR PatOMat, 2010-2012 <http://patomat.vse.cz>
 - GAČR Sémantizace webu (s MFF)