

Architektura aplikací sémantického webu

Doc. Ing. Vojtěch Svátek, Dr.

Zimní semestr 2013

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

Výzkum architektur SW

- Studie Heitmann et al., SWESE@ISWC'09
- Analyzováno 98 aplikací, které se zúčastnily soutěží „Semantic Web Challenge“ a „Scripting for the Semantic Web“
- Výsledkem je schéma architektury o 7 komponentách (+ další poznatky)
- Souvisí s životním cyklem LD

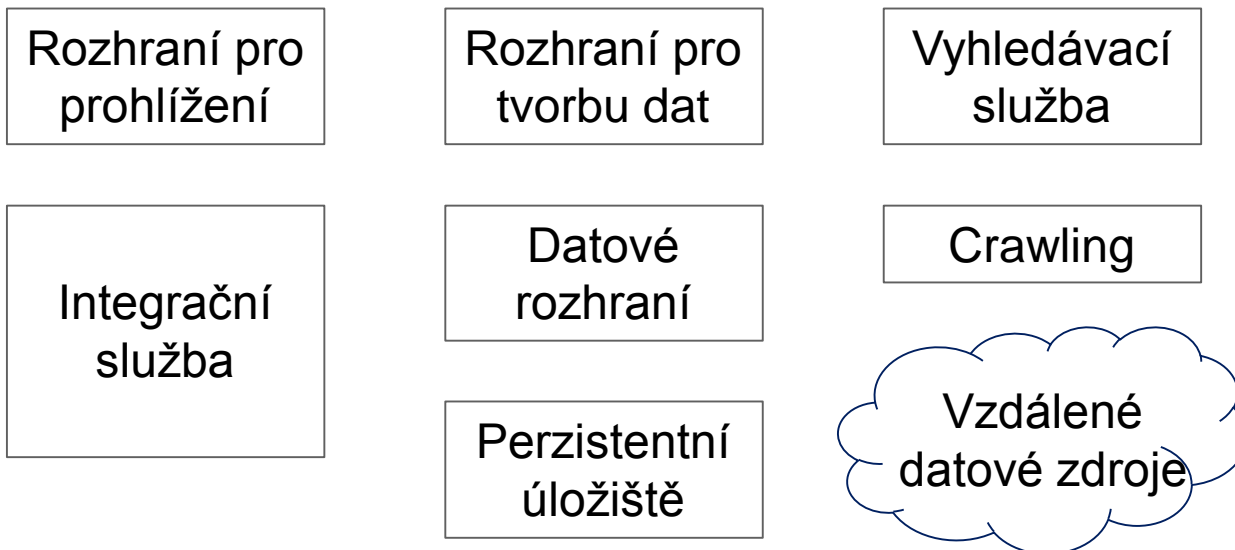
Komponenty

- Datové rozhraní
 - K lokálním a/nebo vzdáleným datům
 - SPARQL, případně programový přístup
- Persistentní úložiště
 - Triple store, případě RDB
 - Datové rozhraní často přímo nad úložištěm
- Uživatelské rozhraní pro prohlížení dat
 - Generické (např. tabulky) nebo doménově specifické (např. mapy) pohledy na data

Komponenty

- Integrovaná služba
 - Heterogenní a/nebo distribuovaná data
 - Někdy i integrace schémat – převážně ruční
 - Někdy (taxonomické) automatické odvozování
- Vyhledávací/dotazovací služba
 - Strukturovaně i plnotextově
- Crawling (pro specifická data)
- Rozhraní pro ruční tvorbu dat
 - Zpravidla rozšíření rozhraní pro prohlížení

Referenční architektura



Problémy aplikací LD

- Chybovost a heterogenita integrovaných dat
 - Vyžaduje ruční předzpracování / následné zpracování (nebo aspoň prohlížení pro účel vzniku skriptů)
 - To snižuje efektivitu integrace
 - Příčiny
 - Používání entit, které nejsou ve slovnících; nebo používání existujících entit špatným způsobem (foaf:img)
 - Proliferace URI pro stejnou entitu; využití inverzně funkčních vlastností částečně řeší, ale může zanášet chyby, např. hodnotou „none“
 - Příklady viz <http://bugs.semanticweb.org/>

Problémy aplikací LD

- Odlišnosti datových modelů a API
 - Různé komponenty používají různé modely: objektový, relační, síťový (RDF)
- Objektově-síťová mapování: zatím spíš jen pokusy
 - <http://ActiveRDF.org> (pro Ruby)
 - <http://packages.python.org/SuRF>

Problémy aplikací LD

- Pomalá standardizace přístupu typu “CRUD“
- Směřuje ke kombinaci REST + SPARQL
 - SPARQL Update: od března 2012 pro verzi 1.1
 - <http://www.w3.org/TR/sparql11-update/>
 - HTTP přístup: Linked Data Platform
 - <http://www.w3.org/TR/ldp/>
- Zkoumání problematiky dynamiky datasetů
 - Dataset Dynamics interest group
 - <http://www.w3.org/wiki/DatasetDynamics>
 - Aktualizace linků, synchronizace a caching dat

Problémy aplikací LD

- Distribuovanost aplikační logiky
 - Rozložena do různých komponent
 - Kód aplikace
 - Dotazy ve SPARQL
 - Generické principy odvozování
 - Doménově specifické axiomy (formule) konkrétních ontologií a slovníků
 - Doménově specifická pravidla a integritní omezení
 - Centralizace aplikační logiky (skrže programový přístup k dalším komponentám) dosud pro aplikace LD není rozvinuta

Předpokládané trendy

- Delegování generických komponent externím poskytovatelům
 - Typicky crawling, vyhledávání, integrace
 - Umožňuje úsporu z rozsahu
 - Poskytuje např. Sindice, přístup k výsledkům přes SPARQL (eliminuje neshodu API)
- Výstavba aplikací z předpřipravených komponent
 - Vznik frameworků analogických k Ruby on Rails, PHPCake, Django...
 - LMF (viz dále)
 - Graphity (<http://graphityhq.com/>)

W3C LDP WG

- Linked Data Platform Working Group
 - http://www.w3.org/2012/ldp/wiki/Main_Page
- Založena v květnu 2012
- Experti z akademické sféry i z praxe (IBM, Oracle, Apache)
- Vznikající standard definuje způsob využití HTTP pro přístup a aktualizaci zdrojů LD
 - Working draft z července 2013:
<http://www.w3.org/TR/ldp/>

Linked Media Framework

- Vyvinul SalzburgResearch
- <https://code.google.com/p/lmf/>
- Serverová aplikace pro kombinaci různých komponent pro LD
- Podporované typy aplikací
 - Publikování dat jako LD
 - Sémantické vyhledávání nad daty
 - Využití tezauru ve SKOS pro extrakci informací

Další integrované platformy

- DataLift platform

<http://datalift.org/en/node/12>

- OntoWiki

<http://ontowiki.net/Projects/OntoWiki>

Použitá literatura

- B. Heitmann, S. Kinsella, C. Hayes, S. Decker, Implementing Semantic Web applications: reference architecture and challenges, in: SWESE 2009: 5th International Workshop on Semantic Web Enabled Software Engineering, <http://ceur-ws.org/Vol-524/>