

Propojování dat a mapování schémat

Doc. Ing. Vojtěch Svátek, Dr.

Zimní semestr 2011

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

Smysl propojování dat

- Samotné vystavení dat ve formátu RDF není postačující pro efektivní vývoj aplikací nad takovými daty, pokud by aplikace musela sama vyřešit problém heterogenity dat pocházejících z různých zdrojů
- Není cílem mít „datová sila“ s „otevřeným poklopem“ a transformovaným obsahem...
- ...ale jeden velký zdroj – „datový web“!
- Zásadní jsou proto **externí odkazy**

Externí odkazy

- Z datového zdroje mohou vést externí odkazy trojího typu
 - Vztahové odkazy
 - Vyjadřují vztahy mezi entitami obdobně jako kdyby šlo o lokální data, ale URI objektu je z externího datového zdroje
 - Typicky využití predikátů lokálního schématu
 - Identitní odkazy
 - *owl:sameAs* pro „reálné“ objekty
 - *skos:exactMatch* nebo „opatrnější“ *skos:closeMatch* pro entity typu „téma“, „oblast“ npod.
 - („Slovníkové“) odkazy na externí schéma
 - *rdf:type*

Důsledek propojení

- Vztahový odkaz umožní doplnit ke zkoumané entitě **jeden** další údaj
- Identitní odkaz zpravidla umožní doplnit ke zkoumané entitě **větší počet** dalších údajů
- Odkaz na externí schéma umožní doplnit ke zkoumané entitě údaje vyplývající z definičních vlastností daného typu, za předpokladu **využití odvozování**
- Zpravidla důraz na vytváření identitních odkazů
 - Často relativně snadná identifikace, funguje přímo (bez odvozování), velký efekt

Nástroje pro linkování

- Existuje několik, momentálně nejpoužívanější **Silk Workbench**
 - Vyvíjen na FU Berlin
 - Součást LOD Stacku
 - Viz <http://www4.wiwiss.fu-berlin.de/bizer/silk/>
- Budeme s ním pracovat na cvičeních

Mapování schémat

- V případě, že se datové zdroje odvolávají na odlišná schémata, je potřeba vzájemně namapovat i schémata (slovníky)
- Mapování se opět vyjadřuje nativně
 - *owl:equivalentClass*, *owl:equivalentProperty*
 - *rdfs:subClassOf* , *rdfs:subPropertyOf*
- Lze uchovávat a vyhledávat stejně jako data a schémata samotná
- Mapování lze provádět částečně i **automaticky** – viz samostatná prezentace
 - V kontextu LD jsou významné metody založené na **instancích** tříd a vlastností