

4IZ440 Propojená data na webu

Cvičení na modelování dat s využitím datových slovníků

Vyučující: prof. Ing. Vojtěch Svátek, Dr.

Zimní semestr 2018

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

Cíl cvičení

- Ohledat si problémy spojené s
 - datovou interpretací slovního popisu určité situace
 - výběrem vhodných slovníků pro RDF data
 - např. jak vyvážit popularitu slovníků a přesnost vyjádření pojmu
 - možným kombinováním více slovníků
 - detailním namodelováním dat s využitím zavedených vzorů a nejlepší praxe, včetně odkazování do existujících datasetů

Zdroje

- Slovníky:
 - <https://lov.linkeddata.es/dataset/lov/>
- Datasets: (např.)
 - DBpedia, textové vyhledávání entit přes <http://dbpedia.org/fct/>
- Vzory a nejlepší praxe pro linked data:
 - <http://patterns.dataincubator.org/book/>
(relevantní: *identifier + modeling patterns*)

Obecné zásady práce

- Zadání dostanete vylosované v týmu o 3-4 lidech
- Ujasněte si, že v týmu chápete zadání stejně
 - výhodné může být načrtnout si hrubý diagram, který často nemusí mít stejnou strukturu jako finální graf RDF dat
- Společně vytvořte model v Turtle
- Pracujte nezávisle na jiných skupinách
- Až budete přesvědčeni, že jste hotovi, ohlaste se

Tvorba modelu

- Vytipujte relevantní slovníky a jednotlivé entity (třídy a vlastnosti) v LOV
 - zpravidla vyhledáváním „Terms“ nebo „Vocabs“
- Pokud nenaleznete vhodnou entitu, navrhňte (pro snadné odlišení) vlastní ve fiktivním jmenném prostoru <http://vocab.example.org>
 - Tento jmenný prostor použijte pouze pro slovníkové entity, ne pro konkrétní zdroje ze zadaného příkladu
- Navrhňte principy pro tvorbu IRI vašich dat
- Při tvorbě dat se snažte přepoužívat i všeobecně známé zdroje z existujících datasetů (např. DBpedia)

Hodnocení

- Lze získat 1 bod na místě za každé ze 3 zadání, a další 1 bod za krátkou zprávu (plain text) o zpracování 1 až 2 zadání, vloženou do odevzdáárny nejpozději 3.10. večer
 - Zprávu může za tým odevzdat kterýkoliv jeho člen
 - Musí obsahovat jak validní RDF (Turtle) kód pro obě cvičení, tak i vysvětlení, proč a jaké byly využity vzory a obecné principy