

4IZ440 Propojená data na webu

Organizační úvod

Vyučující: Doc. Ing. Vojtěch Svátek, Dr.

Zimní semestr 2014

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

Náplň předmětu

- Technologie **propojených dat** (linked data) na („sémantickém“) webu
- Cílem je naučit se data na sémantickém webu
 - **vystavovat** (včetně návrhu schémat / slovníků)
 - **zpracovávat** (především propojovat)
 - **využívat** (v aplikacích)
- Důraz na aktivní přístup a interakci, není „nalejvárna“

Co předmět je a není?

- Nejde o ustálenou disciplínu, pro kterou by existovaly „tradiční“ osnovy a učebnice (i když dobré knihy už vyšly!)
- Obsahem výuky není jedna přesně vymezená technologie (jazyk, metodika, software...)

ALE

- Jde o relativně volné propojení výzkumných iniciativ, softwarových nástrojů, jazykových standardů...
- „Správné odpovědi“ na mnohé klíčové otázky v oboru zatím nikdo na 100% nezná
- Vzhledem k tomu je u studentů velmi vítána vlastní iniciativa, zvědavost, zkoumání toho, co se nově objevuje na webu, na konferencích apod.!

Zdroje pro studium

- Základní literatura:
 - **knih**a: *Linked Data: Evolving the Web into a Global Data Space* od T. Heatha a C. Bizera, online <http://linkeddatabook.com>
 - Tutoriál L. Feigenbaum, E. Prud'hommeaux: *SPARQL by Example*, online <http://www.cambridgesemantics.com/semantic-university/sparql-by-example>
 - Slidy k přednáškám, budou postupně vystavovány na webu
 - učební text *Aktuální problémy a perspektivy sémantického webu* přístupný z webu předmětu
- Doplnková literatura:
 - Viz web předmětu
- Definice používaných **jazyků** jsou dostupné na WWW
- V případě zájmu mnoho další literatury u vyučujícího, články na WWW a v digitálních knihovnách

Kontrolní testy

- Teoretické znalosti budou ověřeny třemi písemnými **testy (12+14+14 bodů)**
 - zaškrtačky, psaní „kódu“, grafy, odpovědi volným textem
- Bodové minimum z každého testu je 5; v jednom testu lze minimum nesplnit bez dalších sankcí
- V případě nedosažení minima ve dvou testech lze o absolvování předmětu uvažovat jen pokud budou splněny všechny ostatní povinnosti a součet bodů dosáhne 60
 - následuje ústní přezkoušení, a pokud je úspěšné, student získá známku „vyhověl“ bez ohledu na celkový počet bodů
- *Testy budou zřejmě ve 3., 7. a 12. týdnu semestru*

Praktické úkoly

- Zpracovat samostatně **semestrální projekt** (max. **40 bodů**)
 - *Detaily a termíny odevzdání budou upřesněny*
- Zpracovat samostatně ústní a písemný **minireferát** (max. **15 bodů**) - v průběhu semestru (prezentace 1.12.)
 - První možnost: z odborného článku z hlavní vědecké konference v oboru - ISWC 2014, přístup k článkům bude zajištěn
 - Druhá možnost: zpracovat s využitím širokého okruhu materiálů podrobnou odpověď na otázku položenou na portálu <http://answers.semanticweb.com/>
- Průběžné úkoly na cvičeních (celkem až **15 bodů**), předběžně
 - Tvorba dotazů SPARQL (na různých úrovních náročnosti)
 - Tvorba anotace v RDFa jako součásti stránky HTML
 - Tvorba datového modelu komodity pro e-commerce
 - Převod dat do RDF
 - Tvorba propojovacího pravidla v nástroji Silk
 - Tvorba jednoduché aplikace (mesh-up) s využitím RDF API
 - Průběžná prezentace semestrálního projektu

Aktuální harmonogram cvičení

1. RDF, Turtle
2. SPARQL
3. SPARQL (SELECT)
4. SPARQL (CONSTRUCT, práce s grafy)
5. Slovníky, RDFa
6. *(odpadá)*
7. Extrakce z tabulek
8. SPARQL (UPDATE)
9. *(odpadá)*
10. Linkování dat
11. Programový přístup
12. Extrakce z Wikipedie
13. Prezentace projektů

Aktuální harmonogram cvičení

Základní stavební kameny

1. RDF, Turtle
2. SPARQL
3. SPARQL (SELECT)
4. SPARQL (CONSTRUCT, práce s grafy)
5. Slovníky, RDFa
6. *(odpadá)*
7. Extrakce z tabulek
8. SPARQL (UPDATE)
9. *(odpadá)*
10. Linkování dat
11. Programový přístup
12. Extrakce z Wikipedie
13. Prezentace projektů

Aktuální harmonogram cvičení

Získávání linked data (z ne-LD zdrojů)

1. RDF, Turtle
2. SPARQL
3. SPARQL (SELECT)
4. SPARQL (CONSTRUCT, práce s grafy)
5. Slovníky, RDFa
6. *(odpadá)*
7. Extrakce z tabulek
8. SPARQL (UPDATE)
9. *(odpadá)*
10. Linkování dat
11. Programový přístup
12. Extrakce z Wikipedie
13. Presentace projektů

Aktuální harmonogram cvičení

Zpracování a publikování linked data

1. RDF, Turtle
2. SPARQL
3. SPARQL (SELECT)
4. SPARQL (CONSTRUCT, práce s grafy)
5. Slovníky, RDFa
6. *(odpadá)*
7. Extrakce z tabulek
8. SPARQL (UPDATE)
9. *(odpadá)*
10. Linkování dat
11. Programový přístup
12. Extrakce z Wikipedie
13. Presentace projektů

Aktuální harmonogram cvičení

Využívání linked data (v koncových aplikacích)

1. RDF, Turtle
2. SPARQL
3. SPARQL (SELECT)
4. SPARQL (CONSTRUCT, práce s grafy)
5. Slovníky, RDFa
6. *(odpadá)*
7. Extrakce z tabulek
8. SPARQL (UPDATE)
9. *(odpadá)*
10. Linkování dat
11. Programový přístup
12. Extrakce z Wikipedie
13. Prezentace projektů

A k čemu Vám to dále může být?

- Pro studenty bezprostředně směřující do praxe
 - Nový pohled na techniky, které se v praxi běžně využívají (integrace datových zdrojů, textová a webová analytika, architektury orientované na služby, metamodelování IS, groupware, ...)
 - Přípravenost na novinky, které do praxe možná proniknou v příštích 5-10 letech
- Pro studenty se zájmem o prozkoumávání neprozkoumaného
 - Přehled významné části výzkumných témat, která jsou na KIZI řešena a mohou být předmětem doktorských disertací, ale i diplomových prací

Pro hlubší zájemce

- Pracovní skupina KIZI **SWOE**
 - „semantic web and ontological engineering“, viz <http://kizi.vse.cz/swoe>
 - účastní se i studenti
- Neformální výzkumný seminář **KEG**
„Knowledge Engineering Group“
 - některé čtvrtky od 10.30 do cca 12.00 zpravidla na 473NB (zasedačka FIS)
 - možno přijít bez předchozího přihlášení
 - viz <http://keg.vse.cz/seminars>
 - kdo chcete dostávat oznámení, napište!
- Aktivity mezi-institucionální iniciativy **OpenData.cz**: <http://opendata.cz>



Pro hlubší zájemce

- Možnost zapojení do **vědeckých projektů** KIZI a fakulty s touto problematikou – na evropské, národní i školní úrovni
 - Vývojářské (Java, PHP, Python aj.), ev. výzkumné činnosti vč. možné spoluúčasti na publikacích a cest na konference
 - Honorováno přes mimořádná stipendia, ev. DPP
 - Možnost zpracovat DP na aktuální témata řešená světovým výzkumem



- Konference **Znalosti**

- Viz <http://znalosti.eu>, 2015 zřejmě v Praze

