

4IZ440 Propojená data na webu

Organizační úvod

Prof. Ing. Vojtěch Svátek, Dr. (přednášející a cvičící)
Ing. Marek Dudáš, Ph.D. (cvičící)

Zimní semestr 2019

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

Zdroje pro studium

- Základní literatura:
 - **knih**a: *Linked Data: Evolving the Web into a Global Data Space* od T. Heath a C. Bizer, online <http://linkeddatabook.com>
 - Slidy k přednáškám, budou postupně vystavovány na webu
 - Tutoriál L. Feigenbaum, E. Prud'hommeaux: *SPARQL by Example*, online <http://www.cambridgesemantics.com/semantic-university/sparql-by-example>
 - Kapitoly z knihy **Umělá inteligence 6**, Academia (knihovna/studovna/zakoupit?)
 - učební text *Aktuální problémy a perspektivy sémantického webu* přístupný z webu předmětu
- Doplnková literatura:
 - Viz web; zvl. B. DuCharme: *Learning SPARQL* (knihovna/studovna)
 - Specifikace **jazyků a formátů** (RDF, Turtle, SPARQL), dostupné na webu
 - Tutoriály k nástroji LP-ETL přístupné z <https://etl.linkedpipes.com/>
 - Další tutoriály ze Semantic University na <https://www.cambridgesemantics.com/blog/semantic-university/>
 - V případě zájmu mnoho další literatury u vyučujících, články na WWW a v digitálních knihovnách

Kontrolní testy

- Teoretické znalosti budou ověřeny třemi písemnými **testy** (předběžně **13+14+13 = 40 bodů**)
 - psaní „kódu“, grafy, odpovědi volným textem, zaškrtačky
- 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 v 5., 8. a 13. týdnu semestru, a to na přednáškách

Praktické úkoly

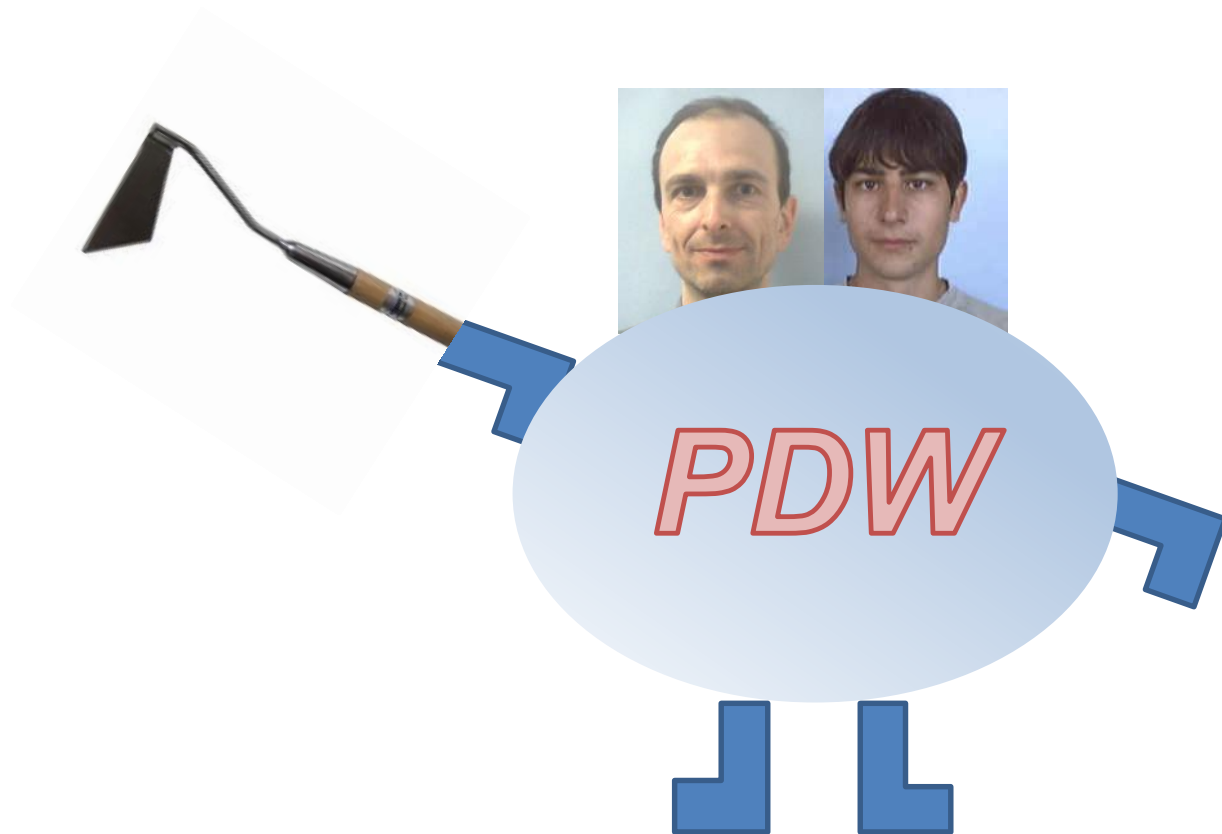
- Zpracovat samostatně **semestrální projekt** (max. **42 bodů**)
 - Průběžná prezentace na posledním cvičení – ústní, plus odevzdání jednostránkového konceptu
 - Finální odevzdání ve zkuškovém období (termín bude upřesněn)
- Zpracovat samostatně ústní a písemný **minireferát** (max. **12 bodů**)
 - Ústní prezentace na přednášce (konec listopadu / začátek prosince)
 - Z odborného článku z hlavní vědecké konference v oboru - ISWC 2019 (vysoce doporučeno, po dohodě lze i z jiné tematicky vhodné konference, případně časopisu), přístup k článkům bude zajištěn
- Průběžné úkoly na cvičeních (celkem až **18 bodů**), předběžně
 - Převody RDF mezi různými způsoby vyjádření
 - Tvorba dotazů SPARQL na různých úrovních náročnosti (dotazování, kontrola a transformace dat, propojování...) a pro různé typy dat
 - Návrh jednoduchého datového slovníku
 - **Týmová** případová studie zahrnující extrakci, čištění, propojení a využití (v primitivní aplikaci) **fiskálních** dat

Upozornění

- Předmět je z kapacitních důvodů vyučován jen **v zimním semestru**. Zvažte proto omluvy a předčasné vzdávání předmětu: na další možnost absolvování byste museli čekat až do dalšího ZS!

Obecné zásady

- Není (vůbec) nutné být programátor, ale pokud ano, lze to v předmětu efektivně využít
 - Alespoň jeden technicky zdatnější člen by měl být v týmu případové studie
 - Pestré spektrum možností pro semestrálky
- Předpokládá se solidní schopnost číst v angličtině
- Při výuce je důraz na aktivní přístup a interakci
 - Když vám něco není jasné nebo připadá divné, ptejte se i zcela sami!
 - Vyučující občas nebudou vědět odpověď, ale mají „přátele na telefonu“ 😊



Těšíme se na spolupráci!