

# Ontologie a sémantický web

## ***Motivace pro realizaci předmětu***

Ontologie se v průběhu posledních desetiletí ukazují jako jedinečný nástroj pro integraci expresivních a heterogenních datových zdrojů v komplexních doménách, např. biologie, lékařství, kultura, atp. Jejich hlavním aplikačním polem však v dnešní době jsou zejména data publikovaná na webu. S využitím ontologií je v posledních letech vytvářen tzv. sémantický web, tedy web obsahující data strojově čitelná a s přesně definovaným sdíleným významem.

Jednou z nejznámějších realizací sémantického webu jsou v dnešní době tzv. Linked Data, která zahrnují stovky vzájemně propojených datových sad publikovaných buď přímo sémantickými anotacemi začleněnými do existujících webových stránek, nebo pomocí specializovaných metod přístupu (např. dotazovací rozhraní jazyku SPARQL).

Předmět “Ontologie a sémantický web” provede zájemce existujícími metodami, technikami a technologiemi sémantického webu. Důraz bude kladen na zvládnutí návrhu ontologií, jejich formalizaci, reprezentaci, efektivní dotazování a aplikační přístup, s ohledem na Linked Data. V dalším bloku budou představeny metody automatického dokazování využívané na sémantického webu. Dále budou představeny některá speciální témata sémantického webu, zejména sémantických GIS, či sémantické anotování webových zdrojů s využitím RDF-A a mikroformátů.

## ***Profil absolventa předmětu***

Předmět poskytne absolventům přehled v oblasti sémantického webu, který stojí na pomezí umělé inteligence (deskripční logiky, automatická dedukce) a webového inženýrství (publikace sémantického webu). Absolvent předmětu získá praktické zkušenosti s formálními jazyky pro návrh sémantického webu, metodikami návrhu formálních ontologií, integrací dat, a tvorbou aplikací nad sémantickým webem. Na jejich základě se bude schopen v praxi podílet na tvorbě složitých formálních taxonomií, tezaurů a ontologií a vytvářet aplikace přistupující k ontologiím.

## ***Rozsah předmětu***

- 2+2 c – pravidelně každý týden přednáška a s ní související cvičení,
- 4 kredity ,
- Kapacita předmětu se očekává mezi 20-40 studenty.

## ***Požadavky na studenty***

Předmět je doporučován ve třetím semestru magisterského studia. Pro tento kurz je nutné, aby se studenti orientovali v databázových technologiích (A4B33DS), matematické logice a automatickém dokazování (např. kurz A4M33RZN). Předmět bude přednášen a cvičen v anglickém jazyce.

## ***Plán přednášek***

1. Historický vývoj sémantického webu, Ontologie
2. Ontologické inženýrství, návrh a modelování ontologií
3. Jazyky pro sémantický web – RDF(S), OWL (2), SWRL
4. Tezaury, slovníky, SKOS
5. Dotazování do ontologií – SPARQL
6. Perzistence sémantických dat – triple/quad stores (SDB, Sesame, Virtuoso), graph stores,

- aplikační přístup k ontologickým datům
7. Linked Data – aktuální stav, rozsah, principy
  8. Linked Data – nástroje (LD consumers – crawlers, LD producers)
  9. Automatické uvažování na sémantickém webu – pravidlové (RDF, OWL-RL), tablové (OWL-DL), hybridní (OWL2-QL)
  10. Mapování ontologií
  11. Sémantické GIS, GeoSPARQL
  12. Sémantické anotace webového obsahu – mikroformáty, RDF-A
  13. Vybrané aplikace sémantického webu
  14. Přehledová přednáška – další trendy data science (MapReduce, NoSQL, Pig)

## **Plán cvičení**

Cvičení budou probíhat v počítačové učebně. Během cvičení budou studenti získávat praktické zkušenosti s existujícími nástroji sémantického webu. Na jejich základě budou zpracovávat semestrální práci (SP), která bude sestávat z dílčího úkolu integrace dostupných webových dat, včetně případné sémantizace existujících datových sad.

1. Úvod, organizace predmetu, zadání SP
2. Návrh ontologií – příklad v známé doméně a porovnání s relačními databázemi
3. Nástroje pro podporu ontologického návrhu - Protege, NeON
4. Návrh SKOS tezauru, test 1
5. Praktické příklady SPARQL
6. Představení triple stores, Sesame/Virtuoso
7. Nástroje Linked Data
8. Práce na SP, konzultace
9. Automatické uvažování na sémantickém webu, práce na SP, konzultace
10. Ontology Matching, test 2
11. Ukázka práce s GeoSPARQL, práce na SP, konzultace
12. Tvorba sémanticky anotované webové stránky, práce na SP, konzultace
13. Práce na SP, konzultace
14. Odevzdání SP

## **Zakončení předmětu**

Předmět je zakončen klasifikovaným zápočtem (KZ). Ten se uděluje při splnění těchto podmínek:

1. účast na cvičeních
2. nadpoloviční hodnocení každého z testů (max 25 bodů každá, min. 13)
3. nadpoloviční hodnocení ze semestrální práce (max 50 bodů, min. 25). Semestrální práce bude sestávat z návrhu a implementace ontologické integrační vrstvy pro zvolenou doménu, publikaci integrovaných dat a vytvoření jednoduché aplikace nad těmito daty.

Přepoččet získaných bodů proběhne s využitím stupnice ECTS.

## **Odpovědné osoby**

**Ing. Petr Křemen, Ph.D. (garant)** obhájil disertační práci Building Ontology-Based Information Systems v červnu 2012 v oboru Umělá inteligence a biokybernetika. Zabývá se technologiemi

sémantického webu, automatickým usuzováním nad ontologiemi, vývojem ontologií, zodpovídáním složitých dotazů nad ontologiemi a návrhem informačních systémů založených na ontologiích. V minulosti se podílel na výzkumu v rámci pěti výzkumných projektů financovaných některým z výzkumných programů Evropské komise: CIPHER, SevenPro, Netcarity, IRIS, CHEF a Tech-it-easy. V současnosti je technickým vedoucím projektu MONDIS financovaného Ministerstvem kultury ČR, zaměřeného na vývoj znalostního systému pro evidenci strukturálních poruch nemovitých památek.

V současnosti se podílí na přednáškách a cvičeních předmětů A4M33RZN Pokročilé metody reprezentace znalostí a Webové a podnikové aplikace A7B39WPA. V minulosti působil jako vedoucí cvičení předmětů X33SPR – Systémy na podporu rozhodování, X33MIS – Manažerské informační systémy, vedl cvičení a podílel se na přednáškách předmětů Y33ZSY – Znalostní systémy, A7B33SUI – Systémy s umělou inteligencí a X33EJA – Enterprise Java.

Dr. Křemen byl na dvou tříměsíčních zahraničních pobytech – v roce 2007 byl zaměstnán u firmy Clark & Parsia, LLC a podílel se na vývoji systému Pellet, open-source inferenčního stroje pro jazyk OWL 2 a v roce 2011 byl výzkumným pracovníkem na Knowledge Media Institute, Open University a podílel se na řešení evropského projektu Tech-it-easy.

Dr. Křemen obdržel v roce 2010 certifikát SUN Certified Java Programmer. V témže roce absolvoval čtyřdenní prezenční kurz firmy Oracle a další dva návazné on-line kurzy s problematikou Java EE 6.

### **Vybrané publikace:**

#### **Časopisecké články**

- Křemen, P. - Kostov, B.: Expressive OWL Queries: Design, Evaluation, Visualization. *International Journal on Semantic Web and Information Systems*. 2013, ISSN 1552-6283, IF=2.308.
- Křemen, P. - Kouba, Z.: Ontology-Driven Information System Design. *IEEE Transactions on Systems, Man, and Cybernetics: Part C*. 2012, vol. 42, no. 3, p. 334-344. ISSN 1094-6977, IF=2.105.
- Žáková, M. - Křemen, P. - Železný, F. - Lavrač, N.: Automating Knowledge Discovery Workflow Composition Through Ontology-Based Planning. *IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING*. 2011, vol. 8, no. 2, p. 253-264. ISSN 1545-5955, IF=1.396.
- Bierhoff, I. - Sproll, S. - Mueller, S. - Avatangelou, E. - Delaney, S. - et al.: Making User-Centred Design a Reality: Experiences from the NETCARITY and SOPRANO Project. *Gerontechnology*. 2010, vol. 9, no. 2, p. 134. ISSN 1569-1101.

#### **Knížní kapitoly**

- Křemen, P. - Kouba, Z. - Blaško, M.: Semantic Annotation of Objects. In *Handbook of Research on Social Dimensions of Semantic Technology and Web Services*. Hershey, Pennsylvania: IGI Global, 2009, p. 223-238. ISBN 978-1-60566-650-1.
- Křemen, P. - Kouba, Z.: Incremental Approach to Error Explanations in Ontologies. In *Networked Knowledge - Networked Media*. Heidelberg: Springer, 2009, p. 171-185. ISBN 978-3-642-02183-1.
- Drdácký, M. - Valach, J. - Křemen, P. - Abrahamčík, J.: Damage database. In *Cultural Heritage Protection Against Flooding*. Prague: Institute of Theoretical and Applied Mechanics Academy of Sciences of the Czech Republic, 2011, p. 185-195. ISBN 978-80-86246-37-6.

#### **Konferenční příspěvky**

- Křemen, P. - Mička, P. - Blaško, M. - Šmíd, M.: Ontology-Driven Mindmapping. In *Proceedings of the 8th International Conference on Semantic Systems [CD-ROM]*. New York: ACM, 2012, p. 125-132. ISBN 978-1-4503-1112-0.

- Křemen, P. - Sirin, E.: SPARQL-DL Implementation Experience. In *OWL: Experiences and Directions 2008* [CD-ROM]. Washington: OWL Working Group, 2008, vol. 1, p. 39-48. ISSN 1613-0073.
- Blaško, M. - Cacciotti, R. - Křemen, P. - Kouba, Z.: Monument Damage Ontology. In *Progress in Cultural Heritage Preservation*. Heidelberg: Springer, 2012, p. 221-230. ISBN 978-3-642-34233-2.
- Bierhoff, I. - Goosen, B. - Wintjens, K. - Huijnen, C. - Křemen, P. - et al.: PARTICIPATORY DESIGN OF NETCARITY SERVICES USING DIFFERENT PERSPECTIVES. In *Proceedings of the first International AEGIS Conference* [CD-ROM]. Seville: AEGIS project, 2010, p. 202-209.
- Žáková, M. - Křemen, P. - Železný, F. - Lavrac, N.: Using Ontological Reasoning and Planning for Data Mining Workflow Composition. In *SoKD: ECML/PKDD 2008 workshop on Third Generation Data Mining: Towards Service oriented Knowledge Discovery* [CD-ROM]. Antwerp: University of Antwerp, 2008,
- Žáková, M. - Křemen, P. - Železný, F. - Lavrač, N.: Planning to Learn with a Knowledge Discovery Ontology. In *Planning to Learn Workshop at ICML 2008* [CD-ROM]. Madison: Omnipress, 2008, p. 29-34.
- Aubrecht, P. - Kejkula, M. - Křemen, P. - Nováková, L. - Rauch, J. - et al.: Mining in Hepatitis Data by LISp-Miner and SumatraTT. In *Proceedings of the Workshop W10- Discovery Challenge*. Porto: Universidade de Porto, 2005, p. 131-138.
- Blaško, M. - Křemen, P. - Kouba, Z.: Privacy Protection Using Semantic Technologies. In *Proceedings of the Twentieth European Meeting on Cybernetics and Systems Research*. Vienna: Austrian Society for Cybernetics Studies, 2010, p. 585-590. ISBN 978-3-85206-178-8.
- Křemen, P. - Kouba, Z.: Conjunctive Query Optimization in OWL2-DL. In *Database and Expert Systems Applications*. Berlin: Springer-Verlag, 2011, vol. 2, p. 188-202. ISBN 978-3-642-23090-5.
- Křemen, P. - Šmíd, M. - Kouba, Z.: OWLDiff: A Practical Tool for Comparison and Merge of OWL Ontologies. In *Twenty-Second International Workshop on Database and Expert System Applications*. Los Alamitos: IEEE Computer Society, 2011, p. 229-233. ISBN 978-0-7695-4486-1.
- Matoušek, K. - Křemen, P. - Küng, J. - Stumptner, R. - Anderlik, S. - et al.: On Transforming a Knowledge Base from Topic Maps to OWL. In *Computer Aided Systems Theory - EUROCAST 2011*. London: Springer, 2011, p. 208-215. ISBN 978-3-642-27548-7.
- Žáková, M. - Železný, F. - Garcia-Sedano, J. - Masia Tissot, C. - Lavrac, N. - et al.: Relational Data Mining Applied to Virtual Engineering of Product Designs. In *16th International Conference on Inductive Logic Programming*. Corunna: University of Corunna, 2006, p. 231-233. ISBN 84-9749-206-4.
- Žáková, M. - Železný, F. - Garcia-Sedano, J. - Massia Tisot, C. - Lavrač, N. - et al.: Relational Data Mining Applied to Virtual Engineering of Product Designs. In *Inductive Logic Programming - 16th International Conference, ILP 2006, Santiago de Compostela, Spain, August 24-27, 2006, Revised Selected Papers*. Heidelberg: Springer, 2007, p. 439-453. ISBN 978-3-540-73846-6.
- Žáková, M. - Železný, F. - Křemen, P. - Tissot, C.M. - Lavrač, N.: Relational Data Mining through Propositionalization and Subsequent Propositional Search for Semantic Virtual Engineering. In *KCAP 2007 Workshop on Knowledge Management and Semantic Web for Engineering Design (KW4ED'07)*. New York: ACM, 2007, p. 49-58. ISBN 978-1-59593-643-1.