# Vojtěch Svátek – researcher carrier summary
*As of Summer 2015*

## Professional affiliation/role

Associate Professor, Deputy Head of Department
Department of Information and Knowledge Engineering
(Previously called Department of Scientific and Technical Information)
University of Economics, Prague, Czech Republic

## Degrees

- MSc in Information Science and Librarianship, in 1991
- PhD in Informatics, in 1998
- Habilitation (tenured) in Informatics, in 2007

All at University of Economics, Prague (UEP).

## Publicly available records

- DBLP record: http://dblp.uni-trier.de/pers/hd/s/Sv=aacute=tek:Vojtech.html
- Google Scholar profile: https://scholar.google.cz/citations?user=E6iKStUAAAAJ
- UEP publications database: http://eso.vse.cz/~sklenak/pcvse/pcvse-search.php?au=Sv%E1tek,%20Vojt%ECch

## Research areas/communities

While the formal part of my MSc study was in Information Science, I already became involved in activities of the department in the field of artificial intelligence, especially expert systems and *machine learning*. My MSc thesis was on machine learning using the rough set theory. My primary field of research during most of 1990s – as of PhD student and early post-doc – had been machine learning and *data mining* (with ECML and later PKDD as principal European conferences, which I repeatedly attended). My PhD thesis was on machine learning enhanced with prior knowledge. At that time I however also kept an eye on manually built knowledge models such as *ontologies* (Ontolingua, OCML) and abstract problem solving methods (PSMs, e.g., in CommonKADS), as well as on *web mining*, in particular on automatic analysis of web content and structure.

After 2000 I followed my previously ignited interest in both knowledge models (ontologies) and mining from text/web, while not completely abandoning the topic of prior knowledge in data mining. I joined the newly forming *semantic web* community via participating at the meetings of the EU OntoWeb project since its start in 2001. Furthermore, I had been involved as researcher in the EU FP4 MGT project (1998-2001), whose topic was *computerized clinical guidelines*. I therefore expanded the scope of my communities of interest to those of knowledge engineering (EKAW conference), semantic web (ISWC/ESWC conferences) and medical informatics (MIE conference).

In the period 2000-2005, and to lesser extend in the next few years, most of my publications dealt with web content/structure mining. Within the RAINBOW project (http://rainbow.vse.cz/) I led a group of undergraduate and PhD students developing tools for analyzing various modalities of web data (such as HTML tags, free text, URIs, site link graph or metadata), and I provided models (PSMs and ontologies) for dynamically combining them into complex tools; this involved collaboration with VU Amsterdam (A. ten Teije). This research was essential for my habilitation thesis (2007).

Since about 2004, until present time, the core of my research has been revolving around creation, management and exploitation of semantic web *ontologies*, including:

- *Ontology matching*: I co-developed, together with Ondřej Šváb-Zamazal, the *OntoFarm* collection used (with numerous enhancements) as the benchmark set of the 'conference' track of the Ontology Alignment Evaluation Initiative (OAEI) annually since 2006, see http://oaei.ontologymatching.org/.
- *Ontology learning from text*: With M. Kavalec (and once with A. Mädche) I co-authored three papers on this topic that together received over 150 citations by foreign researchers.
- *Ontology-based information extraction*: I collaborated with M. Labský and M. Nekvasil on the *Ex* information extraction tool, which exploited ontologies as prior knowledge, see http://eso.vse.cz/~labsky/ex/. The tool has been used in two EU projects.
- *Ontology-enhanced data mining*: I collaborated with J. Rauch, T. Kliegr and others on exploiting prior ontological knowledge in association mining and publishing the discovered knowledge in semantic format: *Sewebar* project, see http://www.easyminer.eu/research.
- *Ontology patterns*: I co-authored one of the first papers on this topic (at BIS'04 conference) and then several dozen others. I have been engaged in the *ontologydesignpatterns.org* portal (as Quality Committee member), co-chaired the WOP workshop at ISWC 2009 and co-presented an Ontology Design Patterns tutorial at ESWC 2009.
- *Ontology debugging and transformation*: Within the PatOMat project I coordinated (http://patomat.vse.cz, 2010-2012), a suite of ontology analysis and (OWL2OWL) transformation tools has been developed (by O. Zamazal and M. Dudáš) and maintained till present. Both the logical and naming structures of OWL ontologies have been addressed.
- *Ontological background models (OBM)*: This is the most active thread at the moment. I initiated (in 2012) and still coordinate the work on the PURO OBM formalism serving as interlingua between different modeling styles usable within the OWL ontological language. The OBM2OWL transformation is an alternative way of 'making ontologies metamorphic' in terms of modeling style, in contrast to the OWL2OWL patterns from PatOMat. PURO is already supported by several prototype implementations for design of new OWL ontology structures (starting from background models) and annotation of existing OWL ontologies, and related papers have been accepted, e.g., for K-CAP, OWLED or WOP.

While most of this research aims at providing support for the semantic web community itself, I engage, in parallel, in activities aiming to bring semantic approaches (esp. linked data) to widespread practical use. This includes both EU projects (LOD2, OpenBudgets.eu) and national initiatives such as *OpenData.cz*, see http://opendata.cz/en. I participated, e.g., in the design of the *Public Contracts Ontology* (http://lov.okfn.org/dataset/lov/vocabs/pc), which is in experimental use in several countries. I also initiated a project on *adopter readiness models* for semantic technology, which was then elaborated by my PhD student M. Nekvasil and resulted in a paper in *J. Web Semantics* (2013).

In the last decade I gained a certain degree of recognition by the semantic web community, reflected in membership in the Editorial Board of Elsevier's *Journal of Web Semantics* and PC membership at relevant conferences (ISWC, ESWC, EKAW, SEMANTiCS). In 2006 I co-chaired EKAW (where I have been consequently serving as SC member, too) and in 2014 I co-chaired the LOD track of ESWC. I also co-chaired 12 topical workshops at major conferences (ECML/PKDD, ESWC, EKAW, ISWC and MIE), three editions of the *Linked Data Mining Challenge* contest (2013-2015), and co-edited a special issue of the *Semantic Web* journal on "Inductive Reasoning and Machine Learning for the Semantic Web". At EKAW and ESWC I also repeatedly served as PhD Symposium reviewer and mentor.

What can perhaps be seen as 3 red lines in my research in long term, running across different fields:

- Finding the *deeper ontological background* behind + *shared patterns* between various data structures (text, data tables, ontologies themselves, clinical guidelines)
- Finding balance between *automatic* and *manual* (expertise-driven) analysis and design of various models
- Finding what may be usable and useful in practice and not just "theory for theory".

## Major publications and citation response

Publications:

- 14 journal papers, including top journals such as *J. Web Semantics*, *J. Biomed. Semantics* or *I. J. Medical Informatics*
- About 200 conference/workshop papers and book chapters, of which over 40 were published by major computer science publishers such as Springer, IOS Press or ACM; this includes major 'semantics' conferences such as EKAW, ESWC or K-CAP.

Citations:

- Over 1800 in Google Scholar, H-index 20
- Over 250 in Web of Science, by Cited Reference Search, w/o self-cites
- Over 140 in Scopus, w/o self-cites.

## International and national projects

Participant coordinator in EU projects (funded partnership, total budget for UEP > 1.5 MEuro):

- H2020 OpenBudgets.eu, 2015-2017
- FP7 LinkedTV, 2011-2015
- FP7 LOD2, 2010-2014
- FP6 IST „Knowledge Space of Semantic Inference for Automatic Annotation and Retrieval of Multimedia Content" (K-Space), 2006-2008
- DG SANCO Public Health „Quality Labeling of Medical Web Content using Multilingual Information Extraction" (MedIEQ), 2006-2008.

Principal investigator or co-investigator in Czech Science Foundation projects:

- „PatOMat – Automation of Ontology Pattern Detection and Exploitation", 2010-2012
- „Web Semantization", 2010-2012 (co-investigator, with P. Vojtáš, Charles Univesity)
- „Intelligent analysis of WWW content and structure", 2003-2005
- "Knowledge model construction based on text documents", 2000-2003 (post-doc project).

I was also involved as collaborator in several other EU and national projects, and as participant contact in several EU projects where UEP was a non-funded / marginal partner only (e.g., FP5 OntoWeb, FP5 Knowledge Web, FP6 Peta Media).

## Principal activities at national/institutional level

Within Czech Republic:

- Co-founder of the national (precisely, Czecho-Slovak) knowledge technology conference *Znalosti* ("Knowedge"), held annually between 2001-2014; I was part of the SC and chaired the 2003 and 2014 editions. I was also engaged in the merger of Znalosti with the national database conference *Datakon*, yielding the *Data&Znalosti* conference starting in 2015 (again, serving as SC member there).
- Popularization of semantic web technology via *tutorials* (web ontologies, semantic web, linked data) at Datakon conference (2002, 2007, 2012) and some others.
- Founder (in 2002) and regular organizer of a *knowledge technology seminar* (labeled as KEG, for "Knowledge Engineering Group"), http://keg.vse.cz/seminars.php, held at UEP, but with most speakers and large part of audience from outside, including recognized speakers from abroad (e.g., from the semantic web community, Richard Cyganiak, Martin Hepp, Robert Stevens or Steffen Staab). Until now there have been 207 seminars with 260 individual talks.
- Since 2014 member of the preparatory committee of the *Prague Computer Science Seminar*, http://praguecomputerscience.cz/index.php?l=en.

- I reviewed 19 PhD theses and two habilitation theses (at seven different Czech universities and one Slovak one), and a number of grant projects at various levels. I also served in three habilitation committees and numerous PhD defence committees.
- I supervise two PhD students at Biomedical Informatics at Charles University, Prague (both working in biomedical text mining).

Within UEP:

- Since 2011 I have been guarantor of the PhD program on *Applied Informatics*, with over 60 PhD students from three departments. In this position I have been responsible for the overall direction of the program, preparation of formal accreditation materials (for the Ministry of Education) and internal regulations, nomination of various committees, organization of discussion meetings, as well as helping the regular PhD advisors in mentoring their students.
- At UEP I currently supervise four PhD students in Applied Informatics: Marek Dudáš (ontological engineering and ontology visualization), Jindřich Mynarz (linked data matchmaking), Václav Zeman (linked data mining) and David Chudán (combination of data mining with OLAP).
- Four of my PhD students graduated up to know: Ondřej Šváb-Zamazal (now continuing as Assistant professor / postdoc), Jan Nemrava, Marek Nekvasil and Petr Nálevka.
- Within my department I engineered (as Deputy Head, responsible for research) its restructuring into four *research groups* in 2014, see http://kizi.vse.cz/research. Currently I am (jointly) leading the *Semantic Web and Ontological Engineering* group (SWOE), see http://kizi.vse.cz/swoe.
- I created and currently lecture two courses (both mandatory within the *Knowledge and web technologies* MSc specialty): *Linked data on the web* and *Mining knowledge from the WWW*. I also lectured and conducted labs in several other (MSc and Bc level) courses. I have been teaching at the department without interruption since 1992.