

VÝVOJ ONTOLOGIÍ PRO VEŘEJNOU SPRÁVU

Otakar ČERBA¹, Václav ČADA², Šimon ZEMAN³

¹ Katedra geomatiky, Fakulta aplikovaných věd, Západočeská univerzita v Plzni, Technická 8, 30614 Plzeň, Česká republika
cerba@kgm.zcu.cz

² Katedra geomatiky, Fakulta aplikovaných věd, Západočeská univerzita v Plzni, Technická 8, 30614 Plzeň, Česká republika
cada@kgm.zcu.cz

³ Národní agentura pro komunikační a informační technologie s.p., Kodaňská 1441/46, 101 00 Praha 10-Vršovice, Česká republika
simon.zeman@nakit.cz

Abstrakt

Katedra geomatiky Západočeské univerzity se podílí (či v minulosti podílela) na tvorbě několika sémantických nástrojů určených pro veřejnou správu. Jedná se například o budování Ontologického systému pro výměnný formát Digitální technické mapy, který byl na konferenci Geoinformace ve veřejné správě prezentovaný v letech 2016 a 2017.

V tomto příspěvku je představená tvorba ontologie, která vzniká ve spolupráci se státním podnikem Národní agentura pro komunikační a informační technologie (NAKIT). Ontologie je zaměřená na popis prvků síťové infrastruktury (například servery, routery) a jejich vztahu k prostorovým entitám, jako jsou budovy nebo areály. Z hlediska typu ontologie se tedy jedná o tzv. geo-ontologii, v níž klíčovou roli hrají především topologické, hierarchické a mereologické vazby. Jako základní formát ontologie je používán standard Web Ontology Language (OWL) s rozšířeními, které vychází z dalších standardizovaných formátů a slovníků (například Resource Description Framework nebo Simple Knowledge Organization System).

Ontologický systém bude využíván nejen jako moderní sémantický prostředek (slovník, tezaurus) pro definování a popis prvků infrastruktury, ale především jako nástroj umožňující efektivní kontrolu získaných dat pomocí podmínek deskripční logiky. V neposlední řadě je potřeba zmínit i implementaci identických a podobnostních vazeb, které umožní propojit ontologii s jinými externími sémantickými zdroji, které obsahují totožné nebo podobné typy objektů.

Vývoj ontologie vychází z existujících metodik (například Simplified Agile Methodology for Ontology Development). Součástí procesu tvorby ontologie tedy není pouze návrh modelu ontologie a jejího naplnění, ale mnoho dalších aktivit, které souvisí se sběrem uživatelských požadavků, tvorbou metodik, vzděláváním, testováním nebo konzultacemi. Vzhledem k charakteru vývoje ontologie pro potřeby NAKIT, orientaci na otevřený přístup (Open Access) a vědecké vzdělávání (Science Education), lze takový aplikovaný výzkum zařadit do kategorie responzivní výzkum a inovace (Responsible Research and Innovation, RRI), které jsou v Česku podporovány například mezinárodním projektem Rosie, na kterém spolupracuje České centrum pro vědu společnost, jenž je jedním z klíčových partnerů Katedry geomatiky Západočeské univerzity na poli mezinárodních projektů.

Abstract

The Department of Geomatics of the University of West Bohemia in Plzeň (Czechia) participates (or has participated) on the development of some semantic tools designed for the public administration. They include for example the Ontological system for the Exchange format of the Digital Technical Map published in the conference Geoinformace ve veřejné správě in 2016 and 2017.

This presentation introduces a building of the ontology, which is developed in the cooperation with the state enterprise Národní agentura pro komunikační a informační technologie (NAKIT). The ontology is focused on a description of elements of network infrastructure (for example servers or routers) and their relations to spatial entities such as buildings or areas. Such ontology could be taken as so-called geo-ontology with crucial role of topological, hierarchical and mereological relations. The Web Ontology Language is used as

the basic format of the ontology, but there are implemented elements from other standards (such as Resource Description Framework or Simple Knowledge Organization System) as well.

The designed system will be used not only as modern semantic tool (vocabulary, thesaurus) for definition and description of parts of infrastructures. It will verify efficiently the acquired data on infrastructures with use of description logic. Last but not least it is necessary to mention an implementation of identical and similar relations making possible to interconnect the ontology to other external semantic tools with the same or similar types of objects.

The development of the ontology is based on existing methodologies (such as Simplified Agile Methodology for Ontology Development). The development has several phases, including design of the ontological model, collection and evaluation of user requirements, education, testing or consulting. Considering characteristics of the development of the ontology for NAKIT, open access and sciences education the development process can be taken as the Responsible Research and Innovation (RRI). This way of research and innovation is supported in the Czech Republic by the Rosie project. This project is in Czechia represented by the Czech Centre for Science and Society, which is very important partner of the University of West Bohemia in the field of international projects.

Klíčová slova: veřejná správa; ontologie; geo-ontologie; konceptualizace; Web Ontology Language.

Keywords: public administration; ontology; geo-ontology; conceptualization; Web Ontology Language.

ÚVOD

V tomto příspěvku je představená tvorba ontologie, která vzniká ve spolupráci státního podniku Národní agentura pro komunikační a informační technologie (NAKIT) a Katedry geomatiky (KGM), Fakulty aplikovaných věd (FAV), Západočeské univerzity v Plzni (ZČU). Ontologie je zaměřená na popis prvků síťové infrastruktury (například servery, routery, antény) a jejich vztahu k prostorovým entitám, jako jsou místnosti, budovy nebo areály. Z hlediska typu ontologie se tedy jedná o tzv. geo-ontologii, v níž klíčovou roli hrají především lokalizované typy objektů a topologické, hierarchické a mereologické vazby (vazby typu část-celek). Jako základní formát ontologie je používán standard Web Ontology Language 2.0 (OWL) s rozšířeními, které vychází z dalších standardizovaných formátů a slovníků, například Resource Description Framework (RDF), RDF Schema (RDFS) nebo Simple Knowledge Organization System (SKOS).

Článek postupně popisuje klíčové části ontologie a její tvorby – volba metodiky, sestavování uživatelských požadavků, technologie, návrh slovníkové části ontologie, realizace vazeb, testování a edukační aktivity.

Účelem článku je představení nově plánovaného projektu, který je zatím ve fázi sběru uživatelských požadavků (informace a závěry publikované v tomto textu vychází z dosud získaných a akceptovaných požadavků a také z obecných pravidel pro tvorbu ontologií), vyvolání diskuze o praktické potřebě ontologií a také získání zpětné vazby pro aktivity spojené s ontologiemi realizované na obou výše zmíněných pracovištích.

VÝVOJ ONTOLOGIE

V roce 2004 parafrázoval John Mylopoulos výrok Stefana Ceriho z roku 1998 – „třemi největšími problémy databází jsou sémantika, sémantika a sémantika“ (Mylopoulos, 2004). Pod toto tvrzení se lze podepsat i v současnosti. Důvodů je mnoho, ale jako nejdůležitější lze označit nekontrolovaný nárůst objemu dat (viz Gantz & Reinsel, 2011, Dragland, 2013).

Sémantické problémy dat, včetně dat prostorových, a především jejich důsledky (nízká interoperabilita, omezená komunikace) se dají řešit pomocí celé řady nástrojů, technologií a standardů, které jsou primárně určené pro podporu sémantiky. Jedná se například o slovníky, registry, metadata, datové modely, schémata, jazyky pro popis dokumentu, tezaury a také ontologie.

V případě sémantické podpory datovýchází v organizaci NAKIT byla zvolena ontologie. Důvodů pro tento krok bylo několik (viz následující seznam):

1. Přání zákazníka / uživatele
2. Velké množství konceptů z různých domén (elektrotechnika, správa sítí, katastr nemovitostí, informační modely budov – BIM...)
3. Nástroj pro sémantickou podporu s velkou pravděpodobností nebude obsahovat data.
4. NAKIT potřebuje modelovat vzájemné vztahy mezi typy objektů (koncepty), které budou v ideálním případě využity pro automatickou kontrolu dat v různých databázích.

Vzhledem k tomu, že ontologie bude obsahovat lokalizovatelné geografické koncepty (teoreticky popsané například v publikaci Smith & Mark, 1998), může být označena jako geo-ontologie – ontologie geografických typů objektů (Smith & Mark, 1998). Podle autorů článku zmíněného v předchozí větě výhody geo-ontologií spočívají především v

1. zlepšení porozumění struktury geografického světa mezi různými skupinami lidí za účelem výměny (nebo naopak zamezení výměny) geografických informací;
2. pochopení (porozumění) různých způsobů vnímání geografických konceptů napříč kulturami, státy, náboženstvími nebo sociálními skupinami;
3. podpoře geografických informačních systémů týkající se manipulace s geografickými entitami;
4. možnosti vytvoření základu výměnného formátu (standardu).

Pro NAKIT jsou důležité především body 2 a 3, přičemž heterogenními skupinami s různým chápáním významu konceptů jsou především dodavatelé technologií a správci různých typů zařízení.

Paralelně se sběrem uživatelských požadavků probíhá volba vhodné metodiky pro tvorbu ontologie. Z hlediska metodiky byl vznesený jediný klíčový požadavek ze strany uživatele – postupné doplňování obsahu (nových konceptů a logických podmínek). Během diskuze mezi oběma zainteresovanými subjekty bylo dále dosaženo souladu v dalším důležitém parametru ontologie – maximální znovu využívání existujících řešení (slovníky, ontologie, sady propojených dat), které umožní propojení ontologie s dalšími sémantickými systémy, a tak umožnit široké sdílení informací (viz definice ontologie v Borst, 1997).

Při tvorbě ontologií je vhodné nepostupovat živelně (ačkoli článek Cardoso, 2007 tvrdí, že zhruba 60% ontologií vzniká bez jakékoli metodiky), ale vycházet z nějaké existující metodiky. Detailní srovnání různých metodik pro konstrukci ontologií je publikováno například v článcích autorů Fernández-López (1999), Corcho et al. (2003) nebo Öhgren & Sandkuhl (2005). Pro účely tvorby popisované ontologie byla zvolená kombinace metodik, které kladou velký důraz na interakci doménových expertů a odborníků na ontologie, potažmo geo-ontologie. Z tohoto důvodu byly zvoleny metodiky On-To-Knowledge (Sure et al., 2004), Gruninger & Fox (Gruninger & Fox, 1995) a Uschold & King (Uschold & King, 1995). Zohledněny byly také metodiky METHONTOLOGY (Fernandez et al., 1997) a SAMOD – Simplified Agile Methodology for Ontology Development (Peroni, 2016), které jsou poměrně obecné, ale i ony poukazují na nutnost kooperace uživatele a faktického tvůrce modelu ontologie.

Tvorba ontologie pro NAKIT se skládá ze čtyř základních iterativních kroků, které vznikly kombinací elementů výše uvedených metodik:

1. Identifikace (upřesnění) cílů a sběr požadavků – metodiky Gruninger & Fox a On-To-Knowledge doporučují metodu kompetenčních dotazů, které byly zařazeny například do úvodního školení.
2. Vyhledání existujících ontologií a jiných vhodných sémantických nástrojů.
3. Konstrukce ontologie, která podle (Uschold & King, 1995) spočívá ve vytvoření jednotlivých prvků ontologie (v první řadě se jedná o slovník ve formě seznamu, který bude dále strukturován s využitím objektových a datotypových vlastností), jejich formalizaci a integraci existujících externích komponent. Slovník bude sloužit především pro stabilizaci pojmového aparátu (zjednodušení

komunikace), zatímco účelem aplikace objektových vlastností a s nimi spojenými logickými podmínkami je především možnost kontroly integrity ontologie a dat navázaných na ontologie, ale také vazba na externí prvky publikované jako propojená data (Bauer & Kaltenböck, 2011, Jain et al., 2010).

4. Evaluace, testování a dokumentace. Tato fáze byla doplněna též o edukační a konzultační aktivity, které jsou klíčové z hlediska používání, údržby a dalšího rozvoje ontologie).

Z technologického hlediska došlo mezi oběma partnery k dohodě, že základním klíčovým kamenem bude standard Web Ontology Language (Motik et al., 2009), verze 2.0, doplněná o vybrané prvky z dalších respektovaných formátů a slovníků, například SKOS (Miles et al., 2005).

ZÁVĚR

Součástí procesu tvorby ontologie není pouze samotný návrh modelu ontologie a jejího naplnění, ale také mnoho dalších aktivit, které souvisí se sběrem uživatelských požadavků, tvorbou metodik, vzděláváním, testováním nebo konzultacemi.

Spolupráce na tvorbě ontologie je daná charakterem obou zapojených organizací. NAKIT je správcem domény (doména v tomto případě zahrnuje především technickou infrastrukturu komunikačních sítí) a jeho zaměstnance lze označit jako doménové experty. Katedra geomatiky Západočeské univerzity se podílí (či v minulosti podílela) na tvorbě Ontologického systému pro výměnný formát Digitální technické mapy (Čada & Čerba, 2016a, Čada & Čerba, 2016b, Čada & Čerba, 2017) nebo ontologie umožňující transformaci kategorií různých klasifikačních systémů a taxonomií v oblasti „land use“ a „land cover“ (Čerba, 2011). Role Západočeské univerzity je důležitá i z hlediska expertních aktivit na doméně prostorových dat a geografických konceptů.

Projekt ontologie pro NAKIT se v současné době nachází v první fázi vývoje (identifikace cílů a sběr primárních požadavků). Další rozvoj závisí především na budoucím provozovateli ontologie. Přesto je možné v současné době vidět první výsledky, které přináší nové poznatky do světa tvorby geo-ontologií. V první řadě se jedná o návrh inovativní metodiky pro tvorbu ontologie, která kombinuje prvky z existujících řešení, které doplňuje vlastními kroky. Druhým faktickým přínosem je vytvoření dvoudenního školení zaměřeného na geo-ontologie, které může být dále realizováno s potenciálními zájemci.

Důležitou vlastností ontologií obecně je možnost integrace sémantických nástrojů, které mohou být i jiné úrovně podrobnosti (granularity). Proto je navrhovaná ontologie pro NAKIT konstruována způsobem, aby byla maximálně interoperabilní, což umožní některé její části využít například pro popis inženýrských sítí v projektu DTM (Digitální technická mapa).

Vzhledem k charakteru vývoje ontologie pro potřeby NAKIT, orientaci na otevřený přístup (Open Access), aktivní spolupráci s uživatelem a vědecké vzdělávání (Science Education), lze takový aplikovaný výzkum zařadit do kategorie odpovědný výzkum a inovace (Responsible Research and Innovation, RRI), které jsou v Česku na poli geoinformačních technologií podporovány například mezinárodním projektem ROSIE, na kterém spolupracuje České centrum pro vědu společnost (CCSS), jenž je jedním z klíčových partnerů Katedra geomatiky Západočeské univerzity na poli mezinárodních projektů.

LITERATURA

Bauer, F., & Kaltenböck, M. (2011). *Linked open data: The essentials*. Edition mono/monochrom, Vienna.

Borst, W. N. (1997). *Construction of engineering ontologies for knowledge sharing and reuse*. University of Twente.

Cardoso, J. (2007). The semantic web vision: Where are we?. *IEEE Intelligent systems*, 22(5).

Corcho, O., Fernández-López, M., & Gómez-Pérez, A. (2003). Methodologies, tools and languages for building ontologies. Where is their meeting point?. *Data & knowledge engineering*, 46(1), 41-64.

Čada, V. & Čerba, O. (2016a). Koncepce ontologie pro základní geodata. Geoinformace ve veřejné správě, 2.-3. května 2016, Praha.

Čada, V. & Čerba, O. (2016b). Ontologie VF XML. ISAF & Geomatics in Projects & Plan4all Joint conference 5.-6. 10. 2016, Kozel (Česká republika).

Čada, V. & Čerba, O. (2017). Koncepce ontologie pro základní geodata, rok poté. Geoinformace ve veřejné správě, 4.-5. května 2017, Praha.

Čerba, O. (2011). Ontologie jako nástroj pro návrhy datových modelů vybraných témat příloh směrnice INSPIRE. Disertační práce, Univerzita Karlova v Praze.

Dragland, A. (2013). Big Data—for better or worse. SINTEF. no. 22 May 2013. Web. 27 Oct.

Fernández-López, M. (1999). Overview of methodologies for building ontologies.

Fernández-López, M., Gómez-Pérez, A., & Juristo, N. (1997). Methontology: from ontological art towards ontological engineering.

Gantz, J., & Reinsel, D. (2011). Extracting value from chaos. IDC iView, 1142(2011), 1-12.

Grüniger, M., & Fox, M. S. (1995). Methodology for the design and evaluation of ontologies.

Jain, P., Hitzler, P., Sheth, A. P., Verma, K., & Yeh, P. Z. (2010). Ontology alignment for linked open data. In International semantic web conference (pp. 402-417). Springer, Berlin, Heidelberg.

Miles, A., Matthews, B., Wilson, M., & Brickley, D. (2005). SKOS core: simple knowledge organisation for the web. In International Conference on Dublin Core and Metadata Applications (pp. 3-10).

Motik, B., Grau, B. C., Horrocks, I., Wu, Z., Fokoue, A., & Lutz, C. (2009). OWL 2 web ontology language profiles. W3C recommendation, 27, 61.

Mylopoulos, J. (2004). Data Semantics Revisited: Databases and the Semantic Web. In Proceedings of the DASFAA (Vol. 4, pp. 17-19).

Öhgren, A., & Sandkuhl, K. (2005). Towards a methodology for ontology development in small and medium-sized enterprises. In IADIS AC (pp. 369-376).

Peroni, S. (2016). A simplified agile methodology for ontology development. In OWL: Experiences and Directions—Reasoner Evaluation (pp. 55-69). Springer, Cham.

Smith, B., & Mark, D. M. (1998). Ontology and geographic kinds.

Sure, Y., Staab, S., & Studer, R. (2004). On-to-knowledge methodology (OTKM). In Handbook on ontologies (pp. 117-132). Springer, Berlin, Heidelberg.

Uschold, M., & King, M. (1995). Towards a methodology for building ontologies.