DBnary: Wiktionary as a Lemon-Based Multilingual Lexical Resource in RDF

Tento abstrakt pojednává o článku zveřejněném v časopise Semantic Web Journal v roce 2014, který představuje dceřiný projekt Wikipedie s názvem Wiktionary[1], jako multijazyčný lexikální zdroj pro vytvoření multijazyčného překladového slovníku v RDF a to za použití modelu LEMON[2].

Autorem tohoto článku je **Gilles Sérasset**[3], mezi editory článku pak patří:

- *Sebastian Hellmann, AKSW, University of Leipzig, Germany*
- *Steven Moran, LMU Munich, Germany*
- *Martin Brümmer, AKSW, University of Leipzig, Germany*
- *John McCrae, CITEC, University of Bielefeld, Germany*

Článek nejprve čtenáře uvádí do dané problematiky a uvádí svou základní motivaci k vytvoření multijazyčného lexikálního zdroje, který by obsahoval správné překlady slov či frází v různých jazycích a také obsahoval jejich definice. Autor se rozhodl využít pro takový slovník interpretaci pomocí propojených dat (linked data), jako zdroj dat pro následnou extrakci do LD pak zvolil Wikislovník (Wiktionary). Jako motivaci pak uvádí množství nekonzistentních dat ve Wikislovníku a zmiňuje fakt, že pro *„vkládání dat“* je často použito různých *„šablon“* respektive přístupu k zanesení informace, což přináší právě zmíněnou nekonzistenci, ale zvyšuje nároky na uživatele těchto dat při jejich zpracování. S těmito problémy se pak musel vypořádat jejich DBnary extractor[4], napsaný v jazyce Java.

Autor připomíná, že jejich snahou nebylo převést obsah Wikislovníku, ale vytvořit lexikální zdroj, který je strukturován jako soubor jednojazyčných slovníků, doplněn o informaci o dvojjazyčných překladech. Během extrakce dat byla pozornost zaměřována pouze na slovníková data, nikoliv pak na odkazování na jiné ontologie atp.

Pro interpretaci pomocí RDF modelu byl zvolen model LEMON, který následně musel být rozšířen o další třídy[5] a vlastnosti. DBnary tak pak tvoří rozšíření modelu LEMON a článek následné třídy a vlastnosti demonstruje na konkrétních modelech (grafech). V rámci rozšíření se neopomínají i sekundární přínosy tohoto slovníku a dochází tak k využití dalších dostupných dat z Wikislovníku k definici objektů *„Nyms“*[6], pro tyto objekty se pak zavádí odpovídající vlastnosti ve jmenném prostoru dbnary[7].

V závěru článku jsou uvedena data aktuální k době psaní tohoto článku (Červen 2013), která ilustrují počet překladů, definic, překladů a jazyků, která již byla extrahována. Zmíněna je také kvalita extraktu a to pomocí porovnání s jiným RDF slovníkem. V samotném závěru pak autor stanovuje cíle do budoucna a sumarizuje možné využití DBnary.

---

[1] Wiktionary – www.wiktionary.org
[2] LEMON – Lexicon model for Ontologies (lemon-model.net)
[3] Gilles Sérasset – e-mail: Gilles.Serraset@imag.fr
[4] DBnary extractor – dbnary.forge.imag.fr
[5] LexicalEntity, Vocable, Nyms, Translations
[6] Nyms – Synonyma, antonyma, hypernyma atp.
[7] Dbnary namespace – kaiko.getalp.org/dbnary#

Reference

SERASSET, Gilles. DBnary: Wiktionary as a Lemon-Based Multilingual Lexical Resource in RDF. *Semantic Web Journal* [online]. **2014**, 7 [cit. 2016-11-25]. Dostupné z: http://www.semantic-web-journal.net/content/dbnary-wiktionary-lemon-based-multilingual-lexical-resource-rdf