Abstrakt článku: Lexvo.org: Language-Related Information for the Linguistic Linked Data Cloud

Tomáš Kiezler

25. listopadu 2016

Autorem článku Lexvo.org: Language-Related Information for the Linguistic Linked Data Cloud je Gerard de Melo, který je zároveň i tvůrcem projektu Lexvo.org. Článek byl publikován v roce 2015 v žurnálu Semantic Web. Autor v něm například popisuje, k čemu služba slouží, jak je v ní řešen problém spojení jazykových výrazů a RDF formátu či odkud jsou použitá data čerpána. Obsahu článku je shrnut v několika následujících odstavcích.

Lexvo.org je služba, která poskytuje popisy lidských jazyků, a to jak současných, ale i historických, uměle vytvořených a jiných. Mimo to zahrnuje i dialekty a jazykové rodiny. K identifikování jazyků používá třímístný kód dle normy ISO 639-3, která čítá více než 7.000 kódů. Jazykové identifikátory služby Lexvo.org jsou využívány např. Britskou knihovnou, Španělskou národní knihovnou nebo univerzitním dokumentačním systémem SUDoc. Jako alternativu lze pro identifikaci použít projekt Glottolog.

Jednotlivé popisy jazyků, jazykových rodin a dialektů jsou extrahovány ze zdrojů, jakými jsou Wikipedia či Unicode CLDR. Predikáty a třídy použité v datech se opírají o vlastní ontologii (`www.lexvo.org/ontology#`), která zahrnuje název jazyka v jiných jazycích, geografické údaje spojené s jeho rozšířením, identifikační kódy, vztahy mezi jazyky apod. Součástí služby je i popis znaků a písem, ve kterých se nacházejí. Označení těchto písem vychází ze standardu ISO 15924.

Mimo to nabízí Lexvo.org i identifikaci výrazů či slov a jejich vlastností. Vzhledem k syntaxi jazyka RDF, kde nemůže literál figurovat jako subjekt, poskytuje služba Lexvo.org jednotné schéma pro tvorbu URI jakéhokoliv výrazu v jakémkoliv jazyce a způsobu zápisu. Výrazy použité v URI adresách jsou zaznamenány bez sémantického či historického hlediska, jejich význam se nerozlišuje. Slovo „dáma“ tudíž bude mít stejnou URI jak ve smyslu ženy, tak ve smyslu deskové hry. Je tomu tak i proto, aby bylo možné strojové

zpracování, neboť stroj vidí zpočátku pouze slovo bez jakýchkoliv dodatečných znalostí o něm. Situace je obdobná u lidí, kteří chtějí vyhledat slovo bez předchozí znalosti jeho významu. Lexvo.org ale rozlišuje, v jakém jazyce se daný výraz vyskytuje, URI proto bude pro slovo „let“ jiné v angličtině a jiné v češtině. Pro ulehčení tvorby URI nabízí Lexvo.org svou vlastní aplikaci napsanou v Javě, kam stačí jen zadat výraz a jeho jazyk.

Služba Lexvo.org využila jako první ve svém projektu UWN lexikální databázi WordNet vytvářenou na Princetonské univerzitě ve formě propojených dat. Databáze obsahuje významy výrazů primárně pro angličtinu, ale mohou být použity i pro jiné jazyky. Lexvo.org tudíž nabízí i vyhledávání výrazů podle jejich významů v různých jazycích.

Webová aplikace služby Lexvo.org poskytuje jak uživatelsky přívětivé prostředí pro vyhledávání informací vztahující se k jazykům, tak i strojově čitelná propojená data ve formátu RDF.

Zdroj: MELO, Gerard de. Lexvo.org: Language-Related Information for the Linguistic Linked Data Cloud. Semantic Web [online]. 2015, **6**(4), 393–400. IOS Press [cit. 2016-11-20]. ISSN 1570-0844. Dostupné z: `http://semantic-web-journal.net/system/files/swj521.pdf`