Abstrakt článku FOOD – FOod in Open Data

Jiří Netušil, VŠE, ZS 2016

4IZ440

Článek popisuje vznik, průběh a přínosy projektu italeského projektu FOOD, zabývající se vytvářením otevřených dat pro zemědělský a potravinářský sektor v Itálii.

Projekt vznikl v reakci na vzestup popularity a poptávky po otevřených datech (Open Data), za předpokladu, že výstup by mohl vést k inovacím a zlepšení ve veřejném sektoru a vědeckých disciplínách.

Projekt má 3 hlavní potenciální dopady:

1. Vytvořená otevřená propojená data (linked open data) je možné opětovně použít v jiných projektech ,
2. vytváří návrh pro vzniknutí standardizačních procesů pro definici vlastností produktů,
3. používání běžných modelů k popisování dat může zlepšit vnitřní fungování ve výměně dat mezi ministerstvem zemědělství a ostatními veřejnými správami.

Nejprve bylo potřeba vytvořit ontologie. Jako zdroj dat sloužilo nad 800 dokumentů ve formátech PDF nebo DOCX. Při vytváření ontologií se kladl důraz na využívání už existujících modelů (DBpedia, AGRO-VOC, atd.). Takto vznikly 3 nejdůležitější třídy tohoto projektu, a to ProtectedName, Type a Product ve vzájemném vztahu. Další ontologie byly postupně dále vytvářeny, avšak především pouze pro další nezbytné vlastnosti produktů, tak aby pokryly různé druhy produktů.

Ontologie byly vytvořeny jako diagramy v nástroji Graffoo (Graphical Framework for OWL Ontologies)[1], nejprve jedna obecná, a poté dalších 20 pro konkrétný typy produktů (chléb, obiloviny, sýr, esenciální olej, ryba, ovoce, med, lékořice, maso, měkkýš, olej, těstoviny, ricotta, šafrán, sůl, salám, sladidlo, zelenina, ocet a víno). Pro všechny tyto diagramy byly vytvořeny soubory typu OWL, ke kterým byly vytvořeny další, popisující vytvořené ontologie pomocí DBpedie, Dolce, WordNet a použitých ontologických návrhových vzorů.

Dalším krokem bylo vytváření samotných otevřených propojených dat z dokumentů poskytnutých ministerstvem zemědělství podle vytvořených ontologií. Dokumenty pocházely z různých zdrojů, takže se často vzájemně lišily, a nebylo možné proces vytváření otevřených dat plně automatizovat. Vytváření datasetů probíhalo ve 3 krocích, extrakci, validaci a konverzi. Všechny záznamy byly definovány permanentními IRI podle vzoru http://w3id.org/food/data/[[class]]/[[product]] (např.: http://w3id.org/food/data/prodotto/vino-abruzzo-rosso).

I když výstupy tohoto projektu nebyly ještě plně integrovány do italského veřejného sektoru, využívání těchto ontologií pro publikování dat je veřejnou administrací povzbuzováno, a vytvořené ontologie a výstupy už našli několik uplatnění. Prvním příkladem je vnitrozemní kraj Itálie Umbrie,

[1] http://www.essepuntato.it/graffoo/

která zveřejnila na svém portálu svá otevřená data využívající již zmíněné hlavní 3 třídy vytvořené v rámci tohoto projektu. Posledním zmínka o využití byla aplikace s názvem "*eat it*", která využívala z projektu vzniklé datasety.

Zdroj:

1. PERONI, Silvio, Giorgia LODI, Luigi ASPRINO, Aldo GANGEMI a Valentina PRESUTTI. FOOD: FOod in Open Data. *The Semantic Web – ISWC 2016: Part 2*. Springer International Publishing AG 2016, 2016, **2016**, 168. - 176. ISSN 1611-3349.