Facilitating Scientometrics in Learning Analytics and Educational Data Mining - the LAK Dataset

LAK dataset je datová sada, poskytující přístup ke strukturovaným textům a metadatům z klíčových výzkumných publikací v oblasti LA (Learning Analytics) a EDM (Educational Data Mining). Jedná se o pravidelně aktualizovaná data, například ze série LAK (Learning Analytics and Knowledge) a EDM konferencí.

Článek pokrývá více než pět let vědeckých publikací z nejrelevantnějších konferencí a vědeckých časopisů. LAK Dataset poskytuje propojená data o bibliografických metadatech, stejně jako celé texty. To vše bylo umožněno díky speciální licenční dohodě s ACM (Association for Computing Machinery), protože tato data zatím nejsou k dispozici otevřeně.

Dataset byl navržen na základě zavedeného vzorku propojených dat a znovu použitelných datových slovníků a poskytuje odkazy na zavedená schémata a entitní koreference na příbuzná data. Díky časovému a tematickému pokrytí datasetu, kterým je téměř kompletní soubor výzkumných publikací konkrétního oboru, usnadňuje scientometrické vyšetřování – například o vývoji vědní oblasti v čase nebo vzájemných vztahů s ostatními obory, což je doloženo jeho použitím napříč širokým spektrem vědeckých studií a aplikaci.

V další části článek rozebírá detaily LAK datasetu, jehož faktová tabulka vypadá následovně:

Name	LAK Dataset
Dataset Home	http://lak.linkededucation.org
Schema	http://lak.linkededucation.org/schema/lak.rdf
Example resource	http://data.linkededucation.org/resource/lak/ conference/lak2013/paper/93
SPARQL endpoint	http://data.linkededucation.org/request/lak-conference/sparql
Dump (RDF/XML)	http://lak.linkededucation.org/lak/LAK-DATASET-DUMP.rdf.zip
Dump (R)	http://crunch.kmi.open.ac.uk/people/~acoope r/data/ LAK-Dataset.RData
Dump (N-Triples)	http://lak.linkededucation.org/lak/LAK-DATASET-DUMP.nt.zip
Publication date	12/12/2012
Last update	30/09/2014
Licence	Creative Commons Attribution (cc-by) for metadata and Open Access graph, special terms for full text of ACM publications (as described in Section 3.1)

Ve čtvrté kapitole se pak věnuje schématu LAK datasetu, mapování extrahovaných dat a jejich propojování. Součástí jsou také přehledové tabulky, ve kterých lze dohledat seznam jmenných prostorů či počty akademických publikací v datasetu obsažené.

V páté kapitole autoři článku ukazují příklady dotazů prostřednictvím jazyka SPARQL.

V šesté se pak článek věnuje praktickému využití LAK datasetu. Součástí kapitoly je také pár odstavců o LAK Data Challenge. Jedná se o každoroční soutěž, která je organizována autory a týmy výzkumníků spojenými se SoLAR, LinkedUp a dalšími přidruženými organizacemi.

Zdroj:

DIETZE, Stefan, Davide TAIBI a Mathieu D'AQUIN. Facilitating Scientometrics in Learning Analytics and Educational Data Mining - the LAK Dataset. *Semantic Web Journal* [online]. 2016, , 9 [cit. 2016-11-25]. Dostupné z: http://www.semantic-web-journal.net/content/facilitating-scientometrics-learning-analytics-and-educational-data-mining-lak-dataset-3