

Utilization of Linked Data in Domain Modeling Tasks

Michal HOLUB*

*Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
holub@fiit.stuba.sk*

The idea of Semantic Web has found many followers among the web researchers. A lot of datasets publicly available use the Linked Data principles. These may be a perfect source for additional metadata utilizable in various tasks of web personalization, recommendation, information retrieval, and data processing. However, only few works actually pursue the idea of wider adoption of such datasets. The aim of our work is to use Linked Data for various tasks of web personalization.

We propose a method for relationship discovery among various concepts forming a concept map. The concept map is intended as a basis of domain models. For this purpose we use unstructured data from the Web, which we transform to concepts and discover links between them. Primarily, we aim at creating concept maps usable for recording the technical knowledge and skills of software engineers, and research fields of interest of scientists (especially in the domain of information technologies).

We examine the utilization of such concept maps and Linked Data utilization in order to 1) improve the navigation in a digital library based on what the user has already visited [7], 2) find similarities between scientists and authors of research papers and recommend them to the readers browsing a digital library [2], 3) analyze Linked Data graphs and find identities and similarities between various entities [5], 4) enhance the displayed articles based on linking entities to DBpedia [4] and recommend additional interesting information to the reader within a digital library, 5) enable users to search for information using natural language queries in English [3].

In all of the problems mentioned above we were able to improve the results of current research methods. We also proved that using Linked Data and concept maps in such problems has potential to improve results in various Web personalization and adaptation tasks.

Linked Data are being used in various datasets forming a Linked Data Cloud. In the center of this cloud there are two main datasets: DBpedia [1] and YAGO [6]. Both

* Supervisor: Mária Bielíková, Institute of Informatics and Software Engineering

use Wikipedia as their primary source of information. The goal of these datasets is to extract and define as many entities as possible, so that others can link to them.

For the purpose of describing the knowledge of software developers we propose the creation of a concept map. The map is composed of a set of concepts representing various technologies and principles the developers are familiar with.

We use a similar approach to describe the professional interests of IT researchers. However, in this case the concept map includes concepts representing various research areas, problems, principles, methods and models studied by the researchers.

Concepts are linked together using relationships like *is part of*, *is a*, *is written in*, or *uses*. We can later utilize the relationships for reasoning, e.g. deduce that when a programmer knows jUnit (a testing framework) he also has to know a bit of Java, because there is a relationship stating “jUnit uses Java”.

Using this process we not only populate the domain model with particular technology, we also find all terms which can describe a technology used when developing software.

We evaluate the models and methods of their creation directly by comparing them to existing ones or by evaluating facts from them using domain experts. Moreover, we evaluate the models indirectly by incorporating them in adaptive personalized web-based systems and measure the improvement in the experience of users (i.e. they get better recommendations, search results, etc.).

Acknowledgement. This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG1/0675/11.

References

- [1] Auer S., Lehmann J.: What Have Innsbruck and Leipzig in Common? Extracting Semantics from Wiki Content. In: *The Semantic Web: Research and Applications*, LNCS Vol. 4519. Springer, (2007), pp. 503-517.
- [2] Chen H., Gou L., Zhang X., Giles C.L.: CollabSeer: A Search Engine for Collaboration Discovery. In: *Proc. of the 11th Annual Int. ACM/IEEE Joint Conf. on Digital Libraries*, ACM Press, (2011), pp. 231-240.
- [3] Chong W., Xiong M., Zhou Q., Yu Y.: PANTO: A Portable Natural Language Interface to Ontologies. In: *The Semantic Web: Research and Applications*, LNCS Vol. 4519. Springer (2007), pp. 473-487.
- [4] Exner P., Nugues P.: Entity Extraction: From Unstructured Text to DBpedia RDF Triples. In: *The Web of Linked Entities Workshop*, Boston, USA, (2012).
- [5] Halpin H., Hayes P.J., McCusker J.P., McGuinness D.L., Thompson H.S.: When owl:sameAs isn't the Same: An Analysis of Identity in Linked Data. In: *Proc. of the 9th Int. Semantic Web Conf. on The Semantic Web – Volume Part I*. Springer (2010), pp. 305-320.
- [6] Suchanek F.M., Kasneci G., Weikum G. YAGO: A Core of Semantic Knowledge Unifying WordNet and Wikipedia. In: *Proc. of the 16th int. Conf. on World Wide Web*, ACM Press, (2007), pp. 697-706.
- [7] Waitelonis J., Harald S.: Augmenting Video Search with Linked Open Data. In: *Proc. of Int. Conf. on Semantic Systems*, Verlag der TU Graz, Austria, (2009).