

Linked Data on the Web in order to Improve Recommendation

Luboš DEMOVIČ*

*Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova, 842 16 Bratislava, Slovakia
demovic@gmail.com*

Currently, the Web provides a large amount of knowledge, and also has the potential to become the largest source of information in the world. Data published on the Web is largely unstructured, intended for people, without a clear definition of entities, their meaning or relationships between them. Presently, a number of researchers are dedicated to processing unstructured data in the form of facts about entities from selected domains from which knowledge bases arose, such as DBpedia, EntityCube, ReadTheWeb, YAGO-NAGA and others [3].

Linked Data principles describe a method for publishing structured data on the Web so that they are connected to each other in a way that makes them useful. In addition, Linked Data contains various variants of links between entities that make it possible to create graphs describing the selected domain. Promoting the importance of data represents the next stage of Web development referred to as Web 3.0.

One of the biggest challenges in the field of intelligent information processing is using Web as a platform for integrating data and information [1]. By intelligent processing and linking of data on the Web, we could create useful new datasets, such as: quick search, translation, personalization, recommendation and user navigation.

In our work, we focus on machine-automated identification of new entities on the Web and discovery of relationships between them. A key technology to achieve this is known as extracting information [2]. This provides us with models, algorithms and tools for transferring web content, text and unstructured data sources into a comprehensible form. Commonly used methods are based on rules and patterns, natural language processing and statistical machine learning. However, existing methods do not guarantee the accuracy and relevance of the results. In addition, in our research work we want to focus on processing data in Slovak language, which brings a lot of obstacles and requires a different approach.

* Supervisor: Michal Holub, Institute of Informatics and Software Engineering

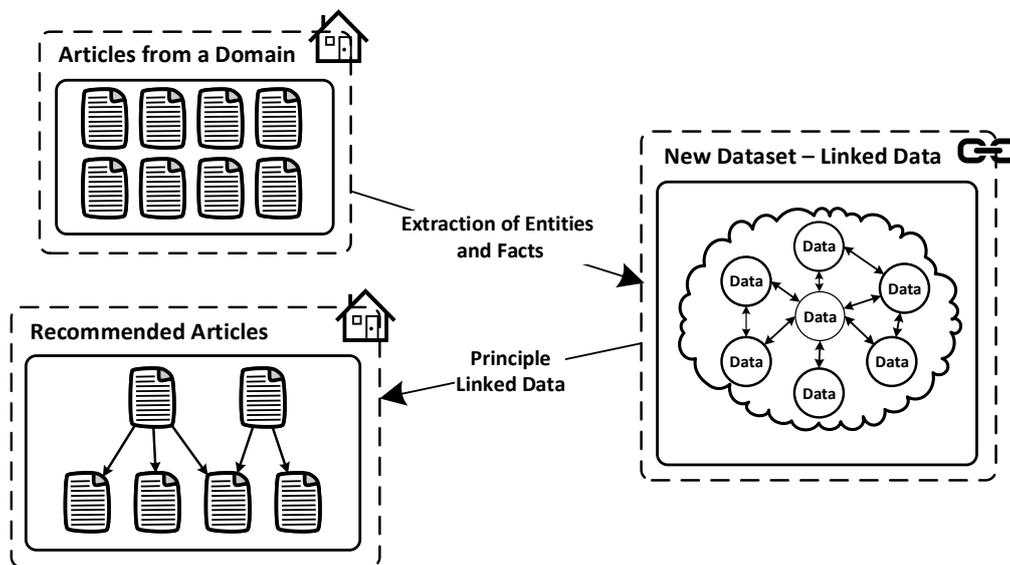


Figure 1 Using Linked Data on the Web in order to improve recommendation

We deal with the analysis of automated machine processing of data on the Web in order to identify and extract entities and facts from Web content. We also deal with exploring the possibility of creating automated datasets obtained from the extracted entities and facts, using the principles of Linked Data.

The aim of our work is to propose a method that allows automated identification and extraction of entities and facts about them using lightweight semantics. Obtained facts will be used to create a dataset describing Linked Data from knowledge in the selected domain. We will verify the proposed method experimentally, by implementing a software tool that will exploit the knowledge base for recommendations. Figure 1 shows how a recommendation using the principles of Linked Data.

Acknowledgement. This work was partially supported by the Slovak Research and Development Agency under the contract No. APVV-0208-10.

References

- [1] Sören Auer, Jens Lehmann, Axel-Cyrille Ngonga Ngom: Introduction to Linked Data and Its Lifecycle on the Web. *Reasoning Web: Semantic Technologies for the Web of Data*, LNCS 6848 (2011). Springer, pp. 1-75.
- [2] Arnaud Sahuguet, Fabien Azavant: Building Intelligent Web Applications Using Lightweight Wrappers. *Data & Knowledge Engineering*, Vol. 36, No. 3 (2001). Elsevier Science, pp. 283-316.
- [3] Gerhard Weikum, Martin Theobald: From Information to Knowledge: Harvesting Entities and Relationships from Web Sources. In: *Proc. of the 29th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (2010)*, Indianapolis, Indiana, USA. ACM Press, pp. 65-76.