

Linking Slovak Entities from Educational Materials with English DBpedia

Luboš Demovič*

*Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
demovic@gmail.com*

Currently, the Web provides a large amount of information, important facts and services to access them. It also has the potential to become the largest source of information and it is, therefore, desirable to be able to automatically gather information about entities and their relationships.

Data published on the Web is largely unstructured, without clear marking of entities and their relationships [1]. Linked Data describe a set of principles for publishing interlinked structured data.

Availability of vast amounts of interconnected data opens up great opportunities to create new generations of applications capable of using these Linked Data structures. With intelligent processing of the available data we can create useful methods aimed at: quick search, translation, personalization, recommendation, and context enrichment or user navigation [2].

In our research work we propose a method that allows the identification and extraction of entities and facts in the Slovak language using search queries. Subsequently, we use the acquired facts for correct and automated linking with the English DBpedia dataset. As DBpedia is the core of the Linked Open Data cloud, such interconnection of entities enables us to use additional information from various datasets. The results can be used for various tasks of personalized Web, e.g. for enriching the information presented to the user with additional facts.

For the purpose of linking Slovak entities, we use the Wikipedia API. One of its services is the OpenSearch service that returns the users' demand for the most popular Wikipedia articles. It is like the autocomplete for Wikipedia, thus for a specific term it returns several articles sorted by popularity. Next, we choose the best result from the offered choices. Now we have linked a Slovak entity with the Slovak Wikipedia.

We use another service from the Wikipedia API to link the Slovak Wikipedia article with all its available linguistic variations throughout Wikipedia. This is convenient, because most articles on Slovak Wikipedia have its English counterpart. At

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

this point, we have a Slovak entity link with the English Wikipedia, which means that we also have a link to the English DBpedia.

The algorithm for linking of entities in the Slovak language consists of five steps:

1. Search for entities via the Wikipedia OpenSearch API.
2. Find the English version of the entity for the selected result via Wikipedia API.
3. Disambiguation check of entities.
4. Choose the correct DBpedia URI
5. Link with the DBpedia dataset.

Figure 1 schematically shows how the algorithm works. We extract the entities using popular search queries from categorized study materials and subsequently we link these entities with DBpedia. As a result, we get DBpedia's entity together with all of its connections.

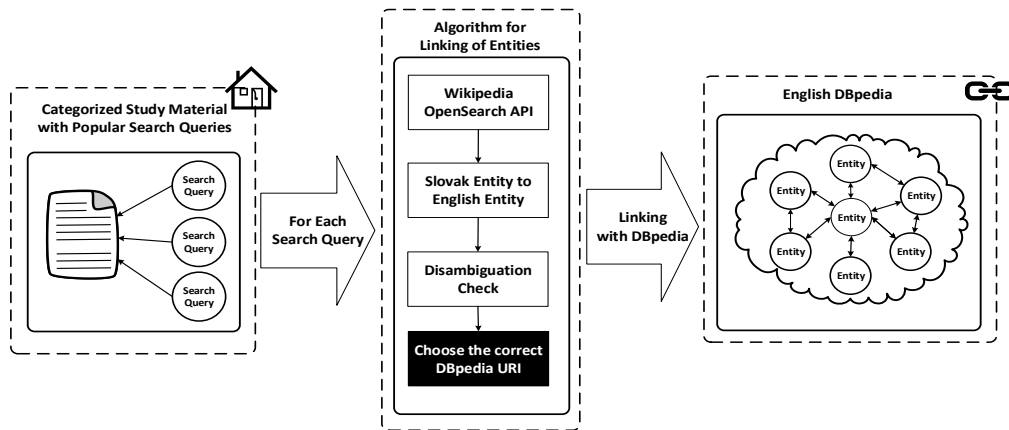


Figure 1. Diagram of the algorithm for linking of entities in Slovak language with DBpedia

Extended version was published in *Proc. of the 10th Student Research Conference in Informatics and Information Technologies (IIT.SRC 2014), STU Bratislava, 137-142.*

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

References

- [1] DeRose, Shen, Chen, Doan, Ramakrishnan. Building Structured Web Community Portals: A Top-down, Compositional, and Incremental Approach. *Proc. of the 33rd Int. Conf. on Very Large Data Bases*. Vienna, Austria: VLDB Endowment, 2007.
- [2] Weikum, Theobald. From Information to Knowledge: Harvesting Entities and Relationships from Web Sources. *Proc. of the 29th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*. Indianapolis, Indiana, USA: ACM Press, 2010.