

# DBpedia internationalization - a graphical tool for I18n infobox-to-ontology mappings

Charalampos Bratsas<sup>1</sup>, Lazaros Ioannidis<sup>1</sup>, Dimitris Kontokostas<sup>1</sup>, Sören Auer<sup>2</sup>, Christian Bizer<sup>3</sup>, Sebastian Hellmann<sup>2</sup>, and Ioannis Antoniou<sup>1</sup>

<sup>1</sup> Web Science, Mathematical Department, Aristotle University of Thessaloniki  
cbratsas/jimkont/lioannid/ioannis.antoniou@math.auth.gr

<sup>2</sup> University Leipzig, Institut für Informatik  
{lastname}@informatik.uni-leipzig.de

<sup>3</sup> Freie Universität Berlin, Web-based Systems Group  
chris@bizer.de

**Abstract.** During the past two decades, the use of the Web has spread across multiple countries and cultures. While the Semantic Web is already served in many languages, we are still facing challenges concerning its internationalization. The DBpedia project, a community effort to extract structured information from Wikipedia, is already supporting multiple languages. This paper presents a graphical tool for creating internationalized mappings for DBpedia.

**Keywords:** DBpedia, mapping-tool, internationalization

## 1 Introduction

Open knowledge sources, like Wikipedia, are already localized to match the cultural diversity of their users. Extracting knowledge from non-English Wikipedia editions should also not give rise to compatibility problems with the internationally accepted scheme. The DBpedia project leverages the source of knowledge present in Wikipedia, converting the structured content into a rich, multi-domain knowledge base [1]. In an effort to extend the *DBpedia Information Extraction Framework* (DIEF)<sup>4</sup> with local Wikipedia resources the I18n functionality has been developed [2] and the *DBpedia Internationalization Committee* (IC)<sup>5</sup> has been established. The purpose of the IC is to support the extension of the DIEF and to tailor language specific aspects of local Wikipedia language editions. In this direction the DIEF was extended with *pluggable filters* [2], in order to deal with resource namespace and naming strategies and the adoption of IRI (rfc 3987) [3] for non-Latin languages.

In this article we give an overview of the *DBpedia Mapping System*<sup>6</sup> and introduce the new DBpedia Mapping Tool (DMT). Finally, a DMT language-agnostic, as part of the wider DBpedia internationalization effort is presented.

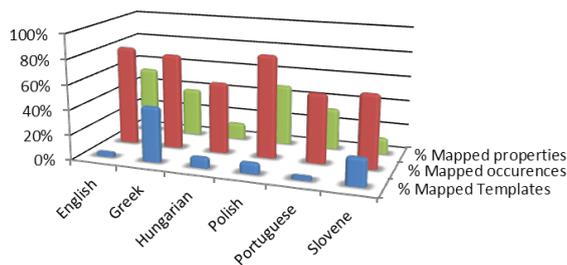
<sup>4</sup> <http://dbpedia.org/documentation>

<sup>5</sup> <http://dbpedia.org/internationalization>

<sup>6</sup> <http://mappings.dbpedia.org>

## 2 Infobox-to-Ontology mappings

The infobox templates<sup>7</sup>, created by Wikipedia users are significant knowledge sources in Wikipedia - extracted by DBpedia. Since there is no central coordination of the naming and usage, infobox template and parameter naming relies solely at the discretion of their authors. This decentralized and self-organizing approach demands the definition of unique identifiers for the same concept. This issue becomes even more important within the DBpedia internationalization effort, where different local Wikipedia editions define their own infobox templates that may not comply with similar English infobox template definitions. To address this ambiguity, the DBpedia Ontology was created [1]. Using the *Mapping system*, different infobox templates and infobox parameters can be collaboratively mapped by the DBpedia community to their corresponding DBpedia Ontology classes and properties. The current mapping status of the most active languages in the DBpedia mapping System is presented in Figure 1. Detailed statistics about the percentages of mapped infoboxes and infobox properties for the 16 languages that are currently covered are available in the DBpedia Mapping Wiki<sup>8</sup>.



**Fig. 1.** Mapping completion per language in terms of whole templates, their properties and the properties occurrences into Wikipedia articles.

### 2.1 DBpedia Mapping System through wiki and the MappingTool

The DBpedia Mapping System is a customized MediaWiki installation, where contributors use special wiki syntax to manually create and edit the mappings and the DBpedia ontology (cf. Figure 2). A mapping is defined by providing the keywords for the DBpedia Ontology classes, their properties and the original Wikipedia template definition, thus being a time-consuming and error-prone operation. In January 2011, the *DBpedia Mapping Tool*<sup>9</sup> (DMT) was introduced

<sup>7</sup> <http://en.wikipedia.org/wiki/Help:Infobox>

<sup>8</sup> [http://mappings.dbpedia.org/index.php/Mapping\\_Statistics](http://mappings.dbpedia.org/index.php/Mapping_Statistics)

<sup>9</sup> <http://mappings.dbpedia.org/mappingtool/web/>

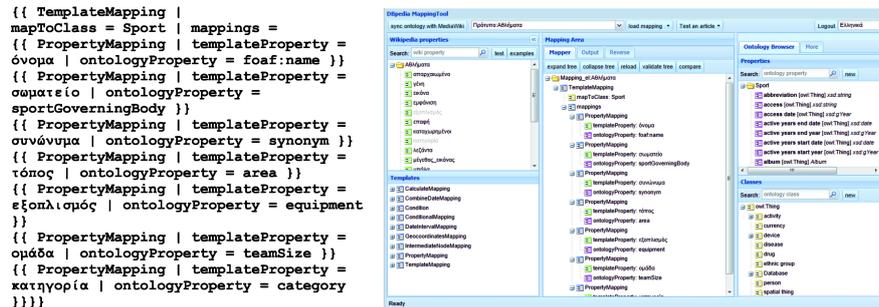


Fig. 2. Mapping using the wiki and the mapping language (left) or the MappingTool (right).

to the DBpedia community<sup>10</sup> but supported only English mapping definitions. The DMT has a purely graphical user interface so that the user avoids the manual cross - DBpedia Ontology and Wikipedia template - keyword checking.

The DMT is built employing a client-server architecture (cf. Figure 3). The server component (Mapping System server) exposes an authenticated HTTP API that retrieves, stores and validates mapping and ontology definitions. The client component is a Javascript application that uses the server API to load mapping and ontology definitions, along with the Wikipedia’s MediaWiki API to retrieve infobox template definitions.

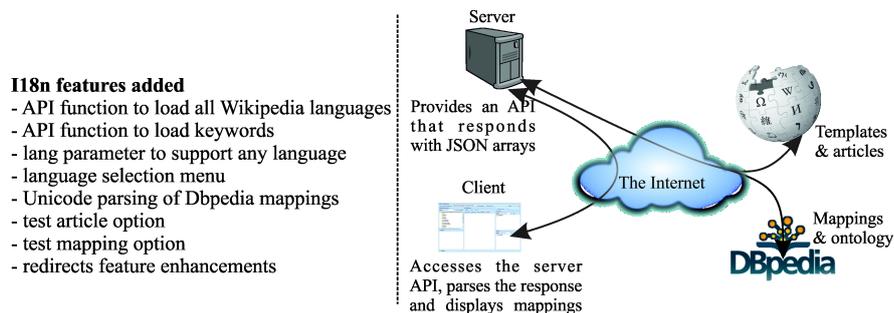
The DMT interface consists of four main panels (cf. Figure 3). The top panel contains the infobox template selector, where the user selects a Wikipedia infobox template, through autocomplete. The left panel consists of two sub-panels: a list with all the available template parameters for the selected infobox template and a second list with all the possible template mappings. The right panel contains two lists with all the defined DBpedia Ontology classes and their properties. In the middle panel the user can view and edit the mappings by dragging and dropping mapping types, parameters and ontology properties from the left and right panels and then validate and store them back to the Mapping System.

## 2.2 I18n DBpedia Mapping Tool Functionality.

The main objective of the *DBpedia IC* for the DBpedia MappingTool was to support the non-Latin characters that appear in IRIs and namespaces, in article names and among the content in any international Wikipedia article, as faced in I18n DIF.

Our main contribution is to make the MappingTool capable of manipulating mappings for resources extracted from any non-English Wikipedia edition. First, the server’s API was extended to include functions for listing all DBpedia supported languages and returning special words for each language, such

<sup>10</sup> [http://sourceforge.net/mailarchive/message.php?msg\\_id=26911324](http://sourceforge.net/mailarchive/message.php?msg_id=26911324)



**Fig. 3.** Technical architecture overview of the DBpedia MappingTool: The changes implemented are shown in the left part.

as “Template”. The server can distinguish the language requested by the *lang* parameter attached to the request URL. Thus, the API becomes language agnostic. The languages are loaded dynamically from the DBpedia mappings wiki, requiring no additional configuration for different languages.

At the client side, the user selects a language available in DBpedia, and proceeds to the main form. At any time, the user can select a new language from a menu at the top right corner. The client parses the mappings correctly due to changes made in the regular expressions of the DMT. Furthermore, two new buttons were added for testing the mapping, against the DBpedia Extraction Results: one showing the articles found to contain the mapped template along with the mapped properties and another one to list the mapped properties found in an arbitrary article.

### 3 Conclusions and future work

The DBpedia Mapping Tool provides a simpler way to create mappings for the DBpedia project based on user contribution. By making the DBpedia Mapping Tool language agnostic, it can now be used for different DBpedia language editions. Thus DBpedia can now further enrich its multilingual resources.

### References

1. Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., Hellmann, S.: DBpedia - A Crystallization Point for the Web of Data, Journal of Web Semantics: Science, Services and Agents on the World Wide Web, Issue 7, Pages 154-165, 2009
2. Kontokostas, D., Bratsas, C., Auer, S., Hellman, S., Antoniou, I., Metakides, G.: Towards Linked Data Internationalization - Realizing the Greek DBpedia, In: Proceedings of the ACM WebSci’11, June 14-17 2011, Koblenz, pp. 1-4, 2011
3. Duerst M. and Suignard M. :Internationalized Resource Identifiers (IRIs) . Network Working Group, 2005.