

Towards Building a Link Set Backed by Domain Experts using the Alignment Tool

Ondřej Zamazal¹, Sotirios Karampatakis^{2,3}, and Charalampos Bratsas^{2,3}

¹ University of Economics, Prague, Dept. Information and Knowledge Engineering
ondrej.zamazal@vse.cz

² Aristotle University of Thessaloniki, School of Mathematics
{sokaramp@auth.gr|cbratsas@math.auth.gr}

³ Open Knowledge Greece

1 Introduction

Discovering semantic relations between entities (entity linking) is one of the most important activity for both semantic web and linked data areas. Either we need link sets of instances or concepts we can rely on automatic systems only to a certain extent. As a result, an automatic linking is accompanied with a user interaction which enables to increase the quality of resulted link sets. Often, in order to reach as much quality of link set as possible the user should be a domain expert for an area of linking task [1]. This user specifics should be considered by designers of interactive entity linking tools. This work presents an experience from an experiment of building a link set for two fiscal code lists where domain experts have been involved. The experiment has been done using the Alignment tool.^{4,5}

2 The Alignment Tool

While the Alignment tool is now a general linking tool, it has originally been developed in order to facilitate linking heterogeneous fiscal code lists in the OpenBudgets project. It is a web application for online, collaborative, system aided manual entity linking. The tool can be used to manually create link sets between two knowledge graphs or to validate already existing link sets. It further offers a number of utilities to aid the linking such as a graph visualization as a tree, a search bar, an entity description and finally suggestions based on linking algorithms provided by Silk [2] or by other automated linking tools. Multiple users can work on the same linking project simultaneously, thus enabling crowdsourcing of a link set creation and reducing required time and effort.

The user can select a semantic meaning of the link by selecting from a number of predefined link types (e.g *skos:related*,⁶ *skos:broadMatch*, *owl:sameAs* etc.) or provide a custom one. The tool can also be used to crowdsource a link validation using a voting system. You can upload links produced by an automated procedure or the tool itself and create polls to check eligibility. Finally, link sets can be exported in various RDF formats and CSV.

⁴ <http://alignment.okfn.gr/>

⁵ <https://github.com/okgreece/Alignment/>

⁶ Skos prefix refers to <http://www.w3.org/2004/02/skos/core# namespace>.

3 Building a Link Set by Involving Domain Experts

European union countries often apply their own different categorization systems for funded projects. As a consequence, this hinders straightforward fiscal analyses. Since there is already integrated European categorization system for funded projects, one of possible solutions to enhance fiscal analyses is to interlink categorization systems of individual EU countries to the European one. For improving this situation we started with building the one link set, the Czech code list (44 items) to the European one (142 items).⁷ In order to ensure the quality of the link set we involved two domain experts and we used the Alignment tool. Thus this work enabled us testing the Alignment tool in action and examining the task of interlinking code lists with domain experts.

Our two domain experts worked separately. They followed detail guidelines⁸ where they were informed about the goal of correctly interlinking as many source items to target items as possible. The guidelines also includes a brief manual how to use the Alignment tool and the instruction that experts should prefer certain types of links more, i.e. there was the following preference *skos:exactMatch*, then *skos:narrowMatch* and *skos:broadMatch* and then the others.

Both experts interlinked 32 same items where expert 1 linked 84% (37) items from the source code list and expert 2 linked 82% (36) items from the source code list. While the expert 1 employed all skos link types (out of all 53 links) more or less uniformly (21 *skos:narrowMatch*, 11 *skos:closeMatch*, 9 *skos:exactMatch*, 8 *skos:relatedMatch*, 4 *skos:broadMatch*), the expert 2 created mainly *skos:narrowMatch* links (116), additionally 8 *skos:exactMatch* and 1 *skos:broadMatch*, out of all 125 links. Both experts managed 32 times to linked the same two entities in one link and, more importantly, they managed to create the very same link 23 times where there were 7 *skos:exactMatch*, 1 *skos:broadMatch* and 15 *skos:narrowMatch*.⁹ The resulted link set of 23 links represents the nucleus of the reference link set. Since there are many links created by only one expert (57% in the case of expert 1 and 82% in the case of expert 2) we further plan to let experts discuss those not agreed links to extend the current reference link set.

During the interlinking by experts we continually received a feedback in terms of bugs and improvement suggestions for the Alignment tool as also reflected via GitHub.

Acknowledgments

We thank to experts. The work has been supported by the H2020 project no. 645833.

References

1. Dragisic Z, Ivanova V, Lambrix P, Faria D, Jimnez-Ruiz E, Pesquita C. User validation in ontology alignment. In: International Semantic Web Conference 2016. Springer.
2. Volz J, Bizer C, Gaedke M, Kobilarov G. Silk-A Link Discovery Framework for the Web of Data. LDOW. 2009.

⁷ Both code lists are extracted from existing data sets.

⁸ The English translation is available at <https://goo.gl/vRYc5r>

⁹ The further information is available at <http://owl.vse.cz:8080/OM2017/>