DESKMatcher

Michael Monych[2], Jan Portisch[1,2], Michael Hladik[2], and Heiko Paulheim[1]

[1] Data and Web Science Group, University of Mannheim, Germany
{jan, heiko}@informatik.uni-mannheim.de
[2] SAP SE Product Engineering Financial Services, Walldorf, Germany
{michael.monych, jan.portisch, michael.hladik}@sap.com

Abstract. This paper describes *DESKMatcher*, a label-based ontology matcher. It utilizes background knowledge from the financial services and enterprise domain to better find matches in these domains. The background knowledge utilized for the enterprise domain was in the form of documentation of terms used in SAP software (textual). Therefore, *Word2Vec* and *GloVe* were used for these corpora. The *Financial Industries Business Ontology* (*FIBO*) was used as more specific background knowledge for the Financial Services domain. Vector Space Embeddings for this corpus were trained using *RDF2Vec* and *KGloVe*. Individual matchers utilizing one set of embeddings (generated from a combination of method and corpus) are pipelined together after a string-based matchers, searching only for matches between entities that have not been assigned to a match in a previous step. Results on the *OAEI* tracks are expected to be sub-par, because low overlap between corpus and task vocabulary is expected.

Keywords: Ontology Matching · Ontology Alignment · Domain Specific Background Knowledge

1 Presentation of the System

1.1 State, Purpose, General Statement

DESKMatcher utilizes vector space embeddings trained using various techniques on three background knowledge datasets specific to the enterprise and financial services domain. Due to the focussed nature of its background knowledge source, it should be best suited for matching tasks in the domains the background knowledge is from. While no good performance on any of the tracks is expected, it would be interesting to see whether any vocabulary overlap existed, that led to an increased performance.

1.2 Specific Techniques Used

The matcher is implemented as a pipeline. Only the label and the entity type (class, datatypeproperty, objecproperty or individual) are considered. The entity

types are used as a filter to only be matched against each other, which proved to be a valuable heuristic in development. Matches are therefore mainly determined based on the entity label. In the first step of the pipelin, obvious matches are marked by a string matcher assuming n:m arity. Following steps try to apply increasingly less specific background knowledge in the form of embeddings trained on respectively less specific corpora, assuming only 1:1 arity. The specificity was assumed from the vocabulary size of a corpus. *Financial Services Business Ontology* (*FIBO*)[2] was used as the most specific source of background knowledge. It is an ontology specific to the Financial Services domain maintained by the EDM council, with the possibility for outside authors to contribute[3]. Two SAP-owned term definition/documentation sets *SAP Glossary* and *SAP Term* (both textual corpora) were used as more general business domain background knowledge source, where *Glossary* is smaller than *Term*.
Embeddings for *FIBO* were trained using Portisch's *jrdf2Vec*[3] implemenation of Ristoski et al.'s *RDF2Vec*[4], as well as Cochez et al's implemenation of their own *KGloVe*[1].

2 Results

todo

3 General Comments

todo

4 Conclusions

todo

[3] https://github.com/edmcouncil/fibo/blob/master/CONTRIBUTING.md

Bibliography

1. Cochez, M., Ristoski, P., Ponzetto, S.P., Paulheim, H.: Global RDF Vector Space Embeddings. In: d'Amato, C., Fernandez, M., Tamma, V., Lecue, F., Cudré-Mauroux, P., Sequeda, J., Lange, C., Heflin, J. (eds.) The Semantic Web – ISWC 2017, vol. 10587, pp. 190–207. Springer International Publishing, Cham (2017)
2. Council, E.: About FIBO. `https://edmcouncil.org/general/custom.asp?page=aboutfiboreview` (????), (accessed 2020-09-02)
3. Portisch, J.: Dwslab/jRDF2Vec. DWSlab (2020)
4. Ristoski, P., Paulheim, H.: RDF2Vec: RDF Graph Embeddings for Data Mining. In: Groth, P., Simperl, E., Gray, A., Sabou, M., Krötzsch, M., Lecue, F., Flöck, F., Gil, Y. (eds.) The Semantic Web – ISWC 2016, vol. 9981, pp. 498–514. Springer International Publishing, Cham (2016)