Ontology matching: from PropBank to DBpedia

Exner, Peter; Nugues, Pierre

Published in:
SLTC 2012

2012

Ontology matching: from PropBank to DBpedia

Peter Exner **Pierre Nugues**

Department of Computer Science, Lund University, Sweden
peter.exner@cs.lth.se, pierre.nugue@cs.lth.se

1. Introduction

In ontology matching, disparate ontologies expressing similar concepts are aligned, enabling tasks such as data and ontology integration, query answering, data translation, etc. (Pavel and Euzenat, 2012). Common alignment methods used in state-of-the-art matching systems, based on similarity measurements, include:

- **Terminological**; comparison of the labels of entities.
- **Structural**; including internal comparison of entities and external comparison of relations with other entities.
- **Extensional**; analyzing the data instances in the ontology.
- **Semantic**; comparing the models of the entities.

While many systems such as those from Seddiqui and Aono (2009) and Cruz et al. (2009) use combinations of terminological and structural methods, the use of extensional and semantic methods in systems such as the one by Jean-Mary et al. (2009) have been largely unexplored (Pavel and Euzenat, 2012).

Similarly to these approaches, we use a combination of alignment methods to create mappings between PropBank (Palmer et al., 2005) predicates and DBpedia (Auer et al., 2007) properties. In particular, we identify predicate–argument structures from Wikipedia articles to extract triples and use a combination approach of extensional and semantical methods during the matching process to align the extracted triples with an exisiting DBpedia dataset.

2. System Description

Our system consists of different modules that perform the text processing tasks in parallel. Taking a set of Wikipedia articles as input, it produces PropBank-DBpedia ontology mappings. A generic semantic processing component based on a semantic role labeler (SRL) identifies the relations in the Wikipedia article texts. A coreference resolution module detects and links coreferring mentions in text and uses them to link the mentions located in the arguments of relations. Using a named entity linking module together with information inferred from the coreference chains, mentions are linked to a corresponding DBpedia URI. Finally, an ontology mapping module performs the final mapping of predicates from the PropBank nomenclature onto the DBpedia namespace.

3. Method

Using PropBank as a dictionary, our semantic parser annotates sentences with predicate–argument structures called rolesets. Our goal is to map more than 7,000 rolesets defined by PropBank, onto a more generalized roleset described by 1,650 DBpedia properties.

The matching process consists of the following steps:

1. Given a set of n-ary predicate–argument relations, we create binary subject–predicate–object relations by combinatorial generation.
2. The subject and object of the extracted relations are matched exactly to existing triples in the DBpedia dataset.
3. From the matching set of triples, links between PropBank roles and DBpedia properties are created. The mappings with the highest counts are selected.

We create then a generalized set of mappings using two procedures:

1. We generalize the subjects and objects of the extracted triples containing DBpedia URIs to 43 top-level DBpedia ontology classes.
2. We generalize the objects containing strings, dates, and numbers to the categories: String, Date, and Number respectively.

Most systems express mappings as alignments between single entities belonging to different ontologies. In addition, we also retain the related subject and object entities in such alignments and use them to express a more detailed mapping.

4. Results and Evaluation

In total, we processed 114,895 articles and we extracted 1,023,316 triples. The system mapped successfully 189,610 triples mapped to the DBpedia ontology. We singled out the unmapped triples using a predicate localized to a custom PropBank namespace. In Table 1, we can see that from the 189,610 extracted triples, 15,067 triples already exist in the DBpedia dataset. This means that our framework rediscovered 15,067 triples during the matching phase and introduced 174,543 new triples to the DBpedia namespace.

Subject	Predicate	Object	Mapping
dbpedia-owl:Person	bear.02.AM-LOC	dbpedia-owl:Place	dbpedia-owl:birthPlace
dbpedia-owl:Person	bear.02.AM-TMP	Date	dbpedia-owl:birthDate
dbpedia-owl:Person	retire.01.AM-TMP	Numeric	dbpedia-owl:activeYearsEndYear
dbpedia-owl:Person	marry.01.A1	dbpedia-owl:Person	dbpedia-owl:spouse
dbpedia-owl:Person	receive.01.A1	Thing	dbpedia-owl:award
dbpedia-owl:Person	manage.01.A1	dbpedia-owl:Organisation	dbpedia-owl:managerClub
dbpedia-owl:Person	serve.01.A1	dbpedia-owl:Organisation	dbpedia-owl:militaryBranch
dbpedia-owl:Place	locate.01.AM-LOC	dbpedia-owl:Place	dbpedia-owl:isPartOf
dbpedia-owl:Place	open.01.AM-TMP	Date	dbpedia-owl:openingDate
dbpedia-owl:Place	build.01.AM-TMP	Numeric	dbpedia-owl:yearOfConstruction
dbpedia-owl:Place	lie.01.A2	dbpedia-owl:Place	dbpedia-owl:locatedInArea
dbpedia-owl:Place	region.01.A1	dbpedia-owl:Place	dbpedia-owl:region
dbpedia-owl:Place	base.01.AM-LOC	dbpedia-owl:Place	dbpedia-owl:capital
dbpedia-owl:Place	include.01.A2	dbpedia-owl:Place	dbpedia-owl:largestCity
dbpedia-owl:Organisation	establish.01.AM-TMP	Numeric	dbpedia-owl:foundingYear
dbpedia-owl:Organisation	find.01.AM-TMP	Date	dbpedia-owl:formationDate
dbpedia-owl:Organisation	base.01.AM-LOC	dbpedia-owl:Place	dbpedia-owl:location
dbpedia-owl:Organisation	serve.01.A2	dbpedia-owl:Place	dbpedia-owl:broadcastArea
dbpedia-owl:Organisation	own.01.A1	dbpedia-owl:Organisation	dbpedia-owl:subsidiary
dbpedia-owl:Organisation	provide.01.A1	Thing	dbpedia-owl:product
dbpedia-owl:Organisation	include.01.A2	dbpedia-owl:Person	dbpedia-owl:bandMember

Table 2: Twenty one of the most frequent ontology mappings learned through bootstrapping.

Type	Count
DBpedia mapped triples (of which 15,067 already exist in DBpedia)	189,610
Unmapped triples	833,706
Total	1,023,316

Table 1: The extracted triples.

Table 2 shows some of the most frequent mappings learned during the matching process. Some general mappings, such as a person marrying another person corresponding to a spouse property, may hold for all cases. However, a mapping describing a person receiving a thing corresponding to an award property, requires a more detailed analysis since the thing received may represent items other than awards. Therefore, we believe that our ontology matching can be improved by a more fine grained approach to the subject-object generalization. In addition, by utilizing interlinking between DBpedia and other datasets such as LinkedMDB[1], we believe we can increase the amount of bootstrapping instances and thereby create more finely expressed mappings of higher quality.

5. Conclusion and Future work

From more than 114,000 articles from the English edition of Wikipedia, we have created a set of mappings aligning PropBank rolesets to DBpedia properties. We have expressed the mappings as a set of links from subject–predicate–object relations to DBpedia properties. In addition, the mappings have been generalized by classifying entities in subjects and objects to 43 top–level DBpedia classes.

We will improve this work by utilizing a more fine-grained approach to generalization by making full use of over 320 DBpedia classes as expressed by the ontology. We also intend to improve the matching process by interlinking related datasets, thereby increasing the amount of training instances used for alignment.

[1] `http://www.linkedmdb.org/`

6. References

Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. 2007. DBpedia: A nucleus for a web of open data. In *The Semantic Web*, volume 4825 of *Lecture Notes in Computer Science*, pages 722–735. Springer Berlin / Heidelberg.

Isabel F. Cruz, Flavio Palandri Antonelli, and Cosmin Stroe. 2009. Agreementmaker: efficient matching for large real-world schemas and ontologies. *Proc. VLDB Endow.*, 2(2):1586–1589, August.

Yves R. Jean-Mary, E. Patrick Shironoshita, and Mansur R. Kabuka. 2009. Ontology matching with semantic verification. *Web Semantics*, 7(3):235 – 251.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The Proposition Bank: an annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–105.

Shvaiko Pavel and Jerome Euzenat. 2012. Ontology matching: State of the art and future challenges. *IEEE Transactions on Knowledge and Data Engineering*, 99(PrePrints):1.

Md. Hanif Seddiqui and Masaki Aono. 2009. An efficient and scalable algorithm for segmented alignment of ontologies of arbitrary size. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7(4):344 – 356.