

A Comparison of Complex Correspondence Detection Techniques

Brian Walshe, Rob Brennan, Declan O’Sullivan

FAME & Knowledge and Data Engineering Group,
School of computer Science and Statistics,
Trintiy College Dublin.

{walshebr|rob.brennan|declan.osullivan}@scss.tcd.ie

Abstract. One to one correspondences between entities are not always sufficient to describe the true relationship between related entities in diverse ontologies, and complex correspondences are needed instead. We demonstrate the types of complex correspondence occurring between two LOD sources and compare techniques for discovering these complex correspondences.

1 Motivation and Background

Most alignment research focuses on one-to-one correspondences between named ontology elements [1], but these are not always sufficient for performing many integration tasks [2]. Data values, for example, may need some form of translation, or some form of condition may be required to scope a broader concept to correspond with a narrower one. These correspondences, which contain conditions or transformations, are known as *complex correspondences*.

There are many known patterns of complex correspondence [2]. Conditional correspondences – where instances of a concept in one ontology are related to a corresponding concept in the other ontology only if they have a particular value for a given attribute – include Class by Attribute Type (CAT), Class by Attribute Value (CAV), and Class by Attribute Existence (CAE). Similarly, Class by Attribute Path Correspondences (PATH) occur when some path of attributes must be followed before the scope of the more general concept can be narrowed. Correspondences where the value of an attribute must be altered in some way are called Attribute Transformation Correspondences (ATC).

In a sample of 50 concepts from YAGO2 [3], six of these concepts corresponded to equivalent concepts in the DBpedia [4] ontology, and 14 concepts required a Class by Attribute Value correspondence. Twenty-one concepts from YAGO2 corresponded with DBpedia concepts with broader scope which could not be narrowed with a correspondence pattern. Six YAGO2 concepts were aligned with DBpedia instances. We found no cases of CAT or PATH correspondences.

Approach	CAV	CAT	CAE	ATC	PATH
Pattern Fitting	Boolean values	Yes	No	No	Yes
MRDM	Yes	Yes	Yes	No	Yes
Model Fitting	Yes	Yes	Yes	Numerical	No

Table 1. Types of correspondence patterns each approach can detect.

2 Detecting Complex Correspondences

Approaches to detecting complex correspondences include a pattern based approach [5], multi relational data mining (MRDM) [6] and our model based approach [7]. Each approach differs in the particular types of correspondence it can detect, and these differences are outlined in table 1. The pattern based approach is the least flexible. For attribute value based patterns it is only capable of detecting cases where attributes have Boolean values. Each of the complex correspondences we found between DBpedia and YAGO2 use non-Boolean attributes, and so it could not detect these. The MRDM approach is more flexible, and is theoretically capable of finding most correspondence patterns listed in section 1, except value transformation patterns. Only the model fitting approach is capable of detecting value transformation correspondences. The current implementation can detect numerical transformations, but the approach could be extended to also detect transformations such as string splitting.

Acknowledgement: This research is supported by the Science Foundation Ireland (Grant 08/SRC/I1403) as part of the FAME Strategic Research Cluster.

References

1. Shvaiko, P., Euzenat, J.: Ontology matching: state of the art and future challenges. IEEE Transactions on Knowledge and Data Engineering. preprint, (2012).
2. Scharffe, F., Fensel, D.: Correspondence patterns for ontology alignment. Knowledge Engineering: Practice and Patterns. 83–92 (2008).
3. Suchanek, F.M., Kasneci, G., Weikum, G.: Yago: A large ontology from wikipedia and wordnet. Web Semantics: Science, Services and Agents on the World Wide Web. 6, 203–217 (2008).
4. Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., Hellmann, S.: DBpedia - A crystallization point for the Web of Data. Web Semantics: Science, Services and Agents on the World Wide Web. 154-165 (2009).
5. Ritze, D., Meilicke, C., Sváb-Zamazal, O., Stuckenschmidt, H.: A pattern-based ontology matching approach for detecting complex correspondences. Proc. of Int. Workshop on Ontology Matching (OM) (2009).
6. Qin, H., Dou, D., Lependu, P.: Discovering Executable Semantic Mappings Between Ontologies. 2007 OTM Confederated international conference on On the move to meaningful internet systems: CoopIS, DOA, ODBASE, GADA, and IS. pp. 832-849 (2007).
7. Walshe, B.: Identifying Complex Semantic Matches. 9th Extended Semantic Web Conference. pp. 849-853 (2012).