Automatic Ontology Alignment Disambiguation based on Ontological Structural Dimension

Alexandre Gouveia, Nuno Silva and João Rocha
GECAD – Knowledge Engineering and Decision Support Research Group,
Instituto Superior de Engenharia do Porto, Porto, Portugal

Keywords: Ontology, Alignment, Disambiguation, Correspondences.

Abstract: Data transformation between repositories (data migration) demands great quality of ontology alignments, and as such ambiguous correspondences must be identified and corrected beforehand. In this paper we address the analysis and systematization of the ontology alignment disambiguation process, proposing the characterization of the ontology matching scenarios through ten dimensions. The characterization of the disambiguation scenarios according to the disambiguation solutions promotes the correct automatic adoption of the disambiguation actions. In this paper we focus on identifying and adopting structural resolution of ambiguities.

1 INTRODUCTION

Ontology mediation as a generic term gathers a set of techniques needed to achieve interoperability in semantically enabled systems, e.g. query rewriting and instance translation (data transformation). Automatic alignment systems (Euzenat & Shvaiko 2007) make use of automatic matching algorithms which evaluate the similarities between pairs of source and target ontologies' entities, exploring different dimensions of ontologies and reducing the user's participation. However, because most of the matching algorithms are insufficient and self-contradictory, the results obtained with automatic alignment systems are in fact below the requirement for ontology mediation (e.g. data integration, migration, data transformation) (Euzenat & Shvaiko 2007; Halevy et al. 2006), especially because there is not a direct and unique relation between these automatic alignments and the alignments that allow data transformation. On the other hand, manual systems, e.g. MAFRA Toolkit (Silva & Rocha 2004), MapForce (Altova 2012), Neon Toolkit (NeOn Foundation 2010), Snoogle (Snoogle 2007) use complex, time-consuming and yet error prone mapping processes that require extensive and profound (human) knowledge of the domain. It is therefore necessary to bridge the gap between automatically generated and data-integration-ready alignments. In order to deal with these issues, Meilicke and colleagues (Meilicke et al. 2007; Ritze et al. 2009) proposed improving automatically created mappings using logical reasoning, focused on the logical validity of the alignment requiring the use of expressive languages, leading to a substantial technological and performance overhead. However, one must keep in mind that ontology mediation alignments are ambiguous implying a strong, often implicit, semantic awareness. The work described in this paper focuses on improving the alignment quality for its application in ontology mediation, and particularly to data transformation. This application has specific requirements not addressed by the logic-based approaches.

The rest of the paper is organized as follows: the next section describes the problem from the conceptual point of view, highlighting the challenges. Section 3 describes the analysis and systematization that lead to the identification of the ambiguous scenarios and respective structure-based correction actions. Section 4 outlooks future research and development directions.

2 PROBLEM STATEMENT

Given two ontologies, an alignment is a set of correspondences between pairs of entities in the form of (e, e', r, n), where e and e' are ontology entities of the source and target ontologies

Gouveia A., Silva N. and Rocha J..
Automatic Ontology Alignment Disambiguation based on Ontological Structural Dimension.
DOI: 10.5220/0004000401640167
In *Proceedings of the 14th International Conference on Enterprise Information Systems* (ICEIS-2012), pages 164-167
ISBN: 978-989-8565-11-2

respectively, r is the relation held between the entities (e.g. equivalence, narrow) and n is the confidence value in the relation.

Consider the alignment described in Figure 1 where the correspondence between Person and Human is responsible for transforming every instance of O1:Person into O2:Human, and the correspondence between O1:name and O2:name transforms (copy) the value of name of every O1:Person into the name of the respective O2:Human. If not consider the correspondence between postal_address and postal_code, this is a data-integration-ready alignment. These relationships and constrains are of foremost importance in scenarios of ontology mediation.

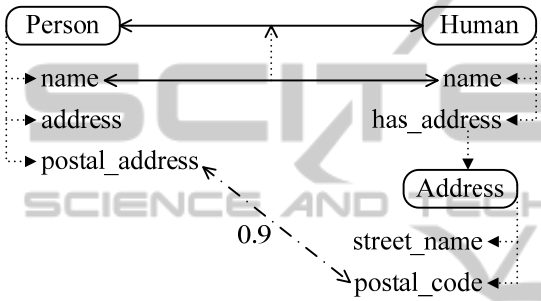

Figure 1: Alignment between two ontologies.

Not all automatically-generated alignments are data-integration-ready alignments. Again, consider Figure 1 where a correspondence is defined between postal_address and postal_code. This alignment is not data-integration-ready because the domain classes of these two properties are not mapped. These are referred to as ambiguous correspondences.

Analysis of automatically-generated alignments shows that these ambiguous cases are quite common preventing direct application in Ontology Mediation tasks. This analysis allowed the detection and identification of several problematic alignment situations:

- Quasi-matching (Figure 2), when in a correspondence between properties (p1, pA), the source property's domain concept (c1) is mapped with one or more target concepts (cA), but this is not domain concept of the mapped target property (pA).

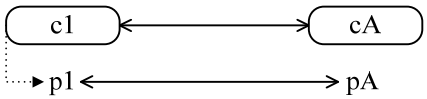

Figure 2: Quasi-matching.

- Semi-matching (Figure 3), when in a correspondence between properties (p1, pA), the source property's domain concept (c1) is not mapped.

Figure 3: Semi-matching.

- Poli-matching (Figure 4), when in a correspondence between properties (p1, PA), there is more than one correspondence between the source property's domain concepts (c1, c2) and the target property's domain concepts (cA) and simultaneously there are sub-concept relations between the mapped concepts (c2 subClassOf c1).

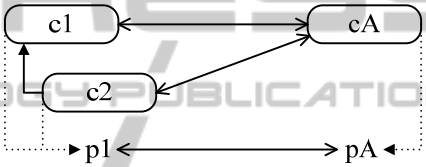

Figure 4: Poli-matching.

We analysed the conference track dataset found in OAEI 2011 Campaign (Euzenat et al. 2011). From the 2.292 correspondences between properties, there were a total of 821 whose domain concepts were not aligned (semi and quasi-matching) and 602 poli-matching, totalling 62% of ambiguous situations. It is worth noting that the reference alignments are themselves ambiguous.

3 SYSTEMATIZATION

The process used to build and characterize the scenarios led to the systematization of ten dimensions, captured in Table 1. For each dimension, several possibilities exist (every one referred to by a single letter). Some of these values are incompatible (X) and some imply other values (i).

Each of these scenarios is characterized by an acronym made up of ten letters (each one corresponding to the value of each dimension), e.g. CEHIMOSUXY in Figure 4.

The scenarios can be divided in two distinct groups according to its resolution: (i) Simple scenarios are those where the only action to take is the creation of the relation between the

Table 1: Incompatibilities and implications.

Characteristic		η		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Ontology dimension	the source property has η domain concepts	none	**A**	X	X	X				i	X			i	X	X				i	X	X	i	X	X	i	X	i	X
		one	**B**	X	X	X				i	X					X						X				i	X		
		many	**C**	X	X	X																							
	the target property has η domain concepts	none	**D**				X	X	X			i	X				i	X	X	i	X	X	i	X	X	i	X	i	X
		one	**E**				X	X	X			i	X						X						X			i	X
		many	**F**				X	X	X																				
	sub-concept relation between the source property's domain concepts	no	**G**							X	X															i	X		
		yes	**H**	X	X	i				X	X																		
	sub-concept relation between the target property's domain concepts	no	**I**									X	X															i	X
		yes	**J**				X	X	i			X	X																
Matching dimension	η source property's domain concepts are mapped	none	**K**											X	X	X				i	X	X	i	X	X	i	X	i	X
		one	**L**	X										X	X	X						X				i	X		
		many	**M**	X	X	i								X	X	X													
	η target property's domain concepts are mapped	none	**N**														X	X	X	i	X	X	i	X	X	i	X	i	X
		one	**O**				X										X	X	X						X			i	X
		many	**P**				X	X	i								X	X	X										
	η source property's domain concepts are mapped with the target property's domain concepts	none	**Q**																	X	X	X	i	X	X	i	X	i	X
		one	**R**	X			X							X			X			X	X	X	X			i	X		
		many	**S**	X	X	i	X							X	X	i	X			X	X	X	X						
	η target property's domain concepts are mapped with the source property's domain concepts	none	**T**																	i	X	X	X	X	X	i	X	i	X
		one	**U**	X			X							X			X			X			X	X	X			i	X
		many	**V**	X			X	X	i					X			X	X	i	X			X	X	X				
	sub-concept relation between the mapped source property's domain concepts	no	**W**																							X	X		
		yes	**X**	X	X	i	X			X	i			X	X	i	X			X	X	i	X			X	X		
	sub-concept relation between the mapped target property's domain concepts	no	**Y**																									X	X
		yes	**Z**	X			X	X	i			X	i	X			X	X	i	X			X	X	i			X	X

correspondences; and (ii) Composed scenarios are those where there is no correspondence between concepts with which one can relate the properties correspondence. Table 2 describes some scenarios regarding their type (i.e. Simple or Composed), the solutions and the respective resulting scenarios. The solution column represents the various possibilities to overcome the ambiguity. E.g.:

- discard: the properties correspondence is discarded;
- new1: create new correspondences between the domain concepts;
- new2: create new correspondences between (i) the direct-domain concepts, or (ii) the super-domain concepts;
- new3: create new correspondences between all the domain concepts;
- relate: relate the correspondence between properties with the correspondences between the domain concepts.

According to the characterization of scenarios and respective solutions, one can observe that the same alternative is found in different scenarios (e.g. new2). In those scenarios, the possible solutions are the same because the scenarios are similar (i.e. common subset of letters). For example, in any scenario having one of H or J and having Q in the acronym, i.e. *[HJ]*Q*, the solutions to adopt are new2, new3, or discard.

Hence, the solutions can be organized and their adoption decided according to the scenarios' characteristics. Figure 5 illustrates the solutions for the *[HJ]*Q* scenarios. For example, one may decide to adopt a new2-i solution in every *[HJ]*Q* scenario.

Table 2: Example of the characterization of scenarios.

Scenario	Type	Solution	Resulting Scenario
BEGIKNQTWY	C	new1	BEGILORUWY
		discard	-
BEGILORUWY	S	relate	-
CEHIKNQTWY	C	new2	CEHILORUWY
		new3	CEHIMOSUXY
		discard	-
CEHILORUWY	S	relate	-
CFGJMNQTWY	C	new2	CFGJMOSUWY
		new3	CFGJMPSVWZ
		discard	-

Defining the solution according to characteristics simplifies the parameterization of disambiguation. Based on the list of identified and characterized

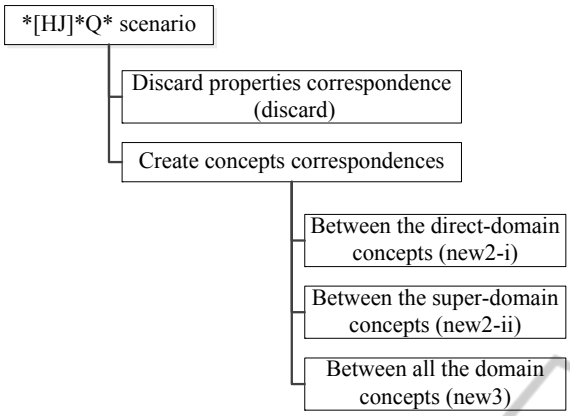

Figure 5: Solutions for the *[HJ]*Q* scenarios.

scenarios, we generated all the possible solutions for every situation. This systematization allowed the explicit selection of solution for every ambiguous situation, giving rise to a set of inter-related solutions referred to as strategies.

4 SUMMARY AND FUTURE WORK

This paper presented an analysis and systematization of the disambiguation process of automatically created ontology alignments, and proposed the characterization of ontology matching scenarios through ten dimensions. As a result, the solutions are identified per type and characteristics of scenarios. The scenarios are categorized according to the associated disambiguation and correction actions (i.e. discard the correspondence between properties, create relations between correspondences, or create correspondences between the domain concepts). This systematization followed a rigorous and extensive identification of incompatibilities and implications between the values of each dimension, and the exhaustive identification and characterization of all possible scenarios that can occur from the initial situation, according to the actions to be taken.

As a result of this systematization we developed a semi-automatic system that identifies, characterizes and solves the alignment scenarios, transforming them to a data-integration-ready alignment, based on a user-defined set of corrective actions (strategies). The experiments carried out demonstrate the completeness of the systematization, i.e. all cases of ambiguity are identified and addressed according to the previously defined strategy.

We are currently working on the identification of arbitrary type of correcting actions. In fact, the identification of an ambiguous situation may lead to arbitrarily complex actions, including triggering (new) matching efforts focused on solving the particular situation. Another trend concerns the application of the proposed systematization and disambiguation approach to more complex scenarios.

ACKNOWLEDGEMENTS

This work is supported by FEDER Funds through the "Programa Operacional Factores de Competitividade - COMPETE" program and by National Funds through FCT "Fundação para a Ciência e a Tecnologia" under the project: FCOMP-01-0124-FEDER-PEst-OE/EEI/UI0760/201 and through the project World Search (QREN11495) of FEDER.

REFERENCES

Altova, 2012. Altova MapForce – Graphical Data Mapping, Conversion, and Integration Tool. Available at: http://www.altova.com/mapforce.html.

Euzenat, J. et al., 2011. Results of the ontology alignment evaluation initiative 2011. In *Proc. of the 6th Int. Workshop on Ontology Matching*. Bonn, Germany.

Euzenat, J. & Shvaiko, P., 2007. *Ontology matching* 1st ed., Heidelberg, Germany: Springer-Verlag.

Halevy, A., Rajaraman, A. & Ordille, J., 2006. Data integration: the teenage years. *In Very Large Data Bases*. Seoul, Korea, pp. 9–16.

Meilicke, C., Stuckenschmidt, H. & Tamilin, A., 2007. Repairing ontology mappings. In *Proceedings of the 22nd National Conference on Artificial Intelligence*. AAAI Press, pp. 1408–1413.

NeOn Foundation, 2010. Neon Plugins - NeOn Wiki. Available at: http://neon-toolkit.org/.

Ritze, D. et al., 2009. A pattern-based ontology matching approach for detecting complex correspondences. In *Proceedings of the 4th International Workshop on Ontology Matching (OM)*. 8th International Semantic Web Conference (ISWC). Chantilly, USA.

Silva, N. & Rocha, J., 2004. Semantic Web complex ontology mapping. *Web Intelligence and Agent Systems Journal*, 1(3-4), p.235—248.

Snoogle, 2007. Snoggle - A Graphical, SWRL-based Ontology Mapper. Available at: http://snoggle.semwebcentral.org/.