Gap Analysis of Ontology Mapping Tools and Techniques

Najam Anjum*, Jenny Harding*, Bob Young* and Keith Case*
* Address Wolfson School of Mechanical and Manufacturing Engineering, Loughborough University, Loughborough, LE11 3TU

Abstract. Mapping between ontologies provides a way to overcome any dissimilarities in the terminologies used in two ontologies. Some tools and techniques to map ontologies are available with some semi-automatic mapping capabilities. These tools are employed to join the similar concepts in two ontologies and overcome the possible mismatches. Several types of mismatches have been identified by researchers and certain overlaps can easily be seen in their description. Analysis of the mapping tools and techniques through a mismatches framework reveals that most of the tools and techniques just target the explication side of the concepts in ontologies and a very few of them opt for the conceptualization mismatches. Research therefore needs to be done in the area of detecting and overcoming conceptualization mismatches that may occur during the process of mapping.

Keywords: Ontology Mapping, Ontology Mismatches, Ontology Mapping Tools and Techniques

1. Introduction

Ontologies have proven to be very helpful in explicitly defining concepts along with their relations and attributes in a formalized way. The characteristic of ontologies being sharable requires the formulation of techniques to allow seamless knowledge transfer between them. This problem of interoperability can be resolved by mapping ontologies. The tools and techniques available for mapping ontologies, however, are not fully automatic and most parts of the mapping process require human involvement. In order to make these tools more reliable and automatic, the mismatches that exist in ontologies need to be studied carefully and the tools available for their detection and resolution are require analysis from the mismatches perspective. In this paper an effort has been made to review the ontology mismathces identified by researchers. A framework is then developed from this review and this is then used to analyze some of the available mapping tools and techniques. The results of this analysis are discussed afterwards.

2. Mapping of Ontologies

Mapping is the process in which for each concept in the source ontology a corresponding concept with similar semantics in the target ontology is found (Ehrig & Staab, 2004). Typically a mapping process consists of three main stages. 1) Mapping discovery, 2)Mapping representation and 3)Mapping execution (Bruijn et al, 2006). Since there needs to be a similarity in the ontologies to be mapped, the mapping discovery stage corresponds to a search for this similarity. Once the similarities are detected a mapping plan is generated in the mapping representation stage and finally the mapping is executed.

Due to the heterogeneous nature of ontologies, the mapping process is subjec to mismatches in their components and building blocks. Being the specification of a conceptualization, an ontology is considered to consist of five components and sets of their definitions. In this scenario an ontology consists of a set of Class definitions, a set of Function definitions, a set of Relation definitions, a set of Instance definitions and a set of Axiom definitions (Visser et al, 1997). Differences in the way these five components are defined give ways in which ontologies can mismatch and these are now discussed.

2.1. Ontology Mismatches

Different types of mismatches have been defined by different authors. The most quoted classification is the one given by Visser et al (1997) who divided the semantic mismatches into two main types namely Conceptualization mismatches and Explication mismatches. Some other mismatches have also been identified. A brief overview of these mismatches is given below.

2.1.1. Conceptualisation Mismatches

These mismatches occur either due to a difference in the way conceptualisations are distinguished in two ontologies or the way they are related to each other in an ontology. Hence two different types of mismatches are Class mismatches and Relation mismatches.

Class Mismatch: This mismatch occurs due to the way classes in two ontologies are differentiated from each other. This mismatch can further be divided into two types namely a categorisation mismatch and aggregation level mismatch. A categorisation mismatch takes place when two similar classes in two ontologies contain different subclasses. An aggregation level mismatch arises when two ontologies define a similar class at different levels of abstraction.

Relation Mismatch: This mismatch happens due to the difference in relations and attributes of classes. Three further subdivisions of this type of mismatch are structure mismatch, attribute assignment mismatch and attribute type mismatch. Structure mismatch occurs when a conceptualisation is specified in two ontologies using a similar set of classes or subclasses but the structuring and relation setting is different. Attribute assignment mismatch occurs when two ontologies assign attributes to two similar classes differently. Attribute type mismatch comes into play when two ontologies in their classes contain similar instances but these instances differ in the way they are defined.

2.1.2. Explication Mismatches

Explication mismatches are due to the difference in the way conceptualisations are defined in an ontology. The definitions of classes, relations and instances are considered to be a 3-tuple of terms, definiens and concepts i.e. Def=<T,D,C> (Visser et al, 1997). An explication mismatch can arise when any of these three components of the 3-tuple in two ontologies are different in some way. The relation between the terms, definiens and concepts is that *definiens* use *terms* to define a *concept*. For example, the definition of a Pen can be ‘a writing device’ or it can be ‘a hollow cylinder filled with ink’. Both of these definitions attempt to describe the concept of a pen but they use different definiens and different terms. In the first one the definiens target the application of the pen while in the second one the structure of a pen is made the basis of its description. With these differences in terms, definiens and concepts, there can be six combinations of explication mismatches in ontologies. These are: concept mismatches (C-Mismatches), definiens mismatches (D-Mismatches), term mismatches (T-Mismatches), concept and definien mismatches (CD-Mismatches), concept and term mismatches (CT-Mismatches) and finally term and definien mismatches (TD-Mismatches). These mismatches are discussed below.

2.1.3. Concept Description Mismatches:

Named as Modelling Convention mismatch by Chalupsky (2000), this type of mismatch comes under the category of *Class Mismatch* of Visser et al. This specific type, however, is not identified by them and therefore becomes an additional type of Class Mismatch. Concept description mismatch occurs when a concept is defined using different sub or super-classes. For example, Chalupsky (2000) states that to distinguish between tracked and wheeled vehicles, a choice one way is to make two subclasses of Vehicle as Tracked-Vehicle and Wheeled-Vehicle. Alternatively, an attribute of Wheeled can be defined with a relation of Traction-type.

2.1.4. Model Coverage and Granularity Mismatch:

This is another type of the class mismatch of Visser et al defined by Klien (2001) and Chaplusky (2000) as Model Coverage and Granularity mismatch. As the name suggests, this mismatch occurs when two ontologies define the same concept with different levels of granularity. For example, a list of names can come under a class Persons or to make it more detailed, the class Person can further be divided into Male and Female. This mismatch appears to be similar to the aggregation level mismatch of Visser et al (1997) but this similarity is not recognized by Klien (2001) and Chalupsky (2000).

2.1.5. Single vs. Multi-Valued Property

This is first of the three mismatches Qadir and colleagues (2007) claim to be different from the mismatches previously identified by other authors. This mismatch occurs when a data-type or object property is represented in the same class but take different number of values in two ontologies. The authors give an example of a class named Bank_Account. In the ontology of one bank, this class might take just one value because that bank doesn't allow its clients to have more than one account but in another, the class with same name might allow multiple values (i.e. to represent several different accounts) according to its policy.

2.1.6. Unique vs. Non-Unique Valued Property

This mismatch occurs when in one ontology a property can hold only one value that uniquely determines the subject, while in another ontology there can be multiple values but they cannot identify the subject uniquely (Qadir et al, 2007). Again quoting an example from the authors which explains the situation where in one ontology of a university, a student is identified by a unique rank number which is recognized by all departments while in another ontology the university requires multiple ranks corresponding to different departments and none of them individually determines the student uniquely.

2.1.7. Alignment Conflict among Disjoint Relations

A mismatch occurring when a disjoint relation in one ontology is not valid in the other. For example a class Student can be declared as disjoint with the class Employee in one ontology while in another, a student is also allowed to be an employee of an institution (Qadir et al, 2007).

2.2. The Mismatches Framework

Table 1 shows the framework formed by accumulating the possible ontological mismatches as described in the previous section. These mismatches are divided into two categories. The main list of semantic mismatches in relation to which all the other mismatches are analyzed is from Visser et al (1997). Their work is the most quoted one in the mismatches literature. Mismatches explained by other authors mostly overlap with those described by Visser and colleagues. For example categorization and aggregation level mismatches of Visser et al are similar to the scope differences of Wiederhold (1994) and scope mismatch of Klien (2001) and Qadir et al (2007). Similarly, the concept and definiens mismatches of Visser et al have a counterpart in the attribute scope mismatch of Wiederhold (1994) and homonym terms mismatch of Klien (2001). On the explication mismatch side, the concept and definiens mismatch of Visser et al (1997) has equivalents in Wiederhold (1994) and Klien (2001) with the names of Attribute Scopes and Homonym Terms mismatch respectively. Similarly, the Term mismatch and

Table 1: Comparison of Identified Ontology Mismatches

Mm - Mismatch

Mismatch Category		Visser et al (1997)	Wiederhold (1994)	Klien (2001)	Chaplusky (2000)	Qadir et al (2007)	Cummulative Mismatches
Conceptualization Mm	Class MM	Categorization Mm	Scope Differences	Scope Mm		Scope Mm	Categorization Mm
		Aggregation-level Mm					Aggregation-level Mm
				Concept Description	Modelling Conventions		Concept Description Mm
				Model Coverage and Granularity Mm	Model Coverage and Granularity Mm		Coverage Mm
						Single vs Multi valued property	Single vs Multi valued property
						Unique vs Non-unique valued property	Unique vs Non-unique valued property
	Relation MM	Structure Mm					Structure Mm
		Attribute-assignment Mm					Attribute-assignment Mm
		Attribute-type Mm					Attribute-type Mm
						Alignment conflict among disjoint relations	Alignment conflict among disjoint relations
Explication Mm	Concept & Term Mm						Concept & Term Mm
	Concept & Definiens Mm		Attribute Scopes	Homonym Terms Mm			Concept & Definiens Mm
	Concept Mm						Concept Mm
	Term & Definiens Mm			Synonym Terms Mm			Term & Definiens Mm
	Term Mm		Naming Differences				Term Mm
	Definiens Mm		Encoding Differences	Encoding Mm			Definiens Mm

Definiens mismatches of Visser et al (1997) are referred to as Naming Differences and Encoding Differences respectively in Wiederhold (1994).

3. Mapping Tools and Techniques

Table 2 lists some of the main techniques used to map ontologies. These techniques include frameworks like MAFRA, OIS, IFF and mapping methods and tools like GLUE, FCA Merge, ONION. These techniques are analyzed for the similarity measures they take to align ontologies and the way they verify the connections made between the mapped ontologies. For the purpose of brevity, a description of these techniques is not included here. The summary of the similarity and verification parameters that these techniques use can be seen in Table 2.

Table 2: Ontology Mapping Techniques

S.No.	Authors	Technique	Similarity Parameters	Verification Parameters
2	Maedch et al (2002)	MAFRA (MApping FRAmework)	Lexical Similarity Property Similarity (attributes or relations)	Object Identity Establishment Statistical Analysis of Transformations
3	Calvanese & Lenzerini (2001)	OIS (Ontology Integration System)	Replies to Queries (Views)	Completeness Soundness Exactness
5	Doan et al (2003)	GLUE	Concept Instances	Similarity Metrics (Probability of similarity of Instances)
6	Noy & Musen (2003)	I PROMPT	Class names	Any term-matching algorithm can be plugged in
7	Noy & Musen (2003)	AnchorPROMPT	Anchor Points	
8	Mitra & Wiederhold (2002)	ONION	Concept names	Context extracted from corpus based word relator
9	Stumme & Maedche (2001)	FCA-Merge	Concept names	Context extracted from corpus of domain specific documents
10	McGuinness et al (2000)	Chimaera	Term names, presentation names, term definitions, possible acronym and expanded forms, names that appear as suffixes of other names	Name resolution list and taxonomy resolution list

4. Analysis

Table 3 uses the mismatches framework developed from the review of typical ontological mismatches. The matrix formed here helps in analyzing the available

Table 3: Analysis of Mapping Tools and Techniques from the Mismatches Point of View

A – Automatic, U – Suggests solution to the user, M – Provides Mechanism, Mm - Mismatches

Semantic Mismatches		MAFRA		PROMPT		Anchor-PROMPT		GLUE		QOM		ONION		FCA-Merge		Chimera	
		Detection	Resolution	Detection	Resolution	Detection	Resolution	Detection	Resolution	Detection	Resolution	Detection	Resolution	Detection	Resolution	Detection	Resolution
Conceptualization Mm	Categorization Mm					M	U			M						M	
	Aggregation-level Mm					A	U			M						M	
	Concept Description Mm									M						M	
	Coverage Mm									M						M	
	Single vs Multi valued property							M		M						M	
	Unique vs Non-unique valued property							M		M						M	
	Structure Mm									M						M	
	Attribute-assignment Mm									M						M	
	Attribute-type Mm									M						M	
	Alignment conflict among disjoint relations									M						M	
Explication Mm	Concept & Term Mm									M						M	
	Concept & Definiens Mm (Homonyms)									M						M	
	Concept Mm									M						M	
	Term & Definiens Mm (Synonyms)	M		M	U	A	U	A	U	A	U	M	U	A	U	A	U
	Term Mm	M		M	U	A	U	A	U	A	U	M	U	A	U	A	U
	Definiens Mm	M		M	U	A	U	A	U	A	U	M	U	A	U	A	U

tools and techniques from the mismatches point of view. Three symbols are used here to denote the capability of a particular method to detect and resolve a mismatch as done by Klien (2001). *'A'* stands for automatic and represents a capability of automatically detecting or resolving a mismatch. *'U'* stands for user and symbolizes the suggestions a tool offers to the user to solve a particular mismatch, and *'M'* denotes the mechanism provided to the user, by a tool or technique, to detect or resolve a mismatch. Before any results are obtained from this analysis, it is necessary to clarify that the tools and techniques are designed to find out similarities while the mismatches literature stresses the dissimilarities that are present among ontologies. Hence, the fields filled in Table 3 indicate that a certain tool or technique overcomes a particular mismatch in one ontology by connecting a differently placed or named concept to a corresponding concept in another ontology.

A quick glimpse of this table reveals some empty fields representing a lack of available features in tools and techniques to detect and resolve conceptualization mismatches. Most of the tools and techniques provide a mechanism to the user to detect and resolve mismatches. It can be seen from table 3 that QOM (Quick Ontology Mapping) and Chimaera have a mechanism for the users to detect the conceptualization mismatches. This is because in QOM the breadth of scope of similarity measure allows this technique to cover all of the mismatches to be detected. In Chimaera, however, it is its detailed and user friendly interface that helps the user to manually detect any kind of mismatches. This on one hand shows that the available tools and techniques need to be made more automatic and on the other it indicates that these tools should be modified to target conceptualization mismatches. It is also clear from table 2 that the available tools and techniques mainly focus on finding the similarities rather than dissimilarities between the concepts in two ontologies and then establishing correspondences. So the main steps involved in every technique are:

1- Scanning ontologies for similar concepts,

2- Authenticating the similarity through different algorithms and tools,

3- Establishing correspondences.

The second step is the one which deals with the verification of knowledge in shared ontologies and it is here that the research so far is mainly directed towards the explication side of terminologies and concepts. The conceptualization side of the interpretation of terms and concepts is virtually void of any significant work Table 3 shows that only AnchorPROMPT provides an automatic detection of one of the conceptualization type of similarities and also suggests the possible correspondence that can be established between specific concepts in the ontologies to be mapped. The other two tools QOM and Chimaera just provide information about the structure of ontologies so that it becomes easier for the user to detect some conceptualization similarity.

The gap identified here suggests thatresearch is required to find ways through which different conceptualization mismatches can be detected and resolved in order to give accuracy to the process of mapping and thus verifying the knowledge being shared.

5. References

Bruijn, J. de., Ehrig, M., Feier, C., Marin-Recuerda, F., Scharffe, F. and Wieten, M., 2006, "Ontology Mediation, Merging and Aligning", In Davies J, Studer R, Warren P (eds), Semantic Web Technologies: Trends and Research in Ontology-based Systems, Wiley, UK, 2006

Calvanese, D. and Lenzerini, M., 2001. Ontology of integration and integration of ontologies, *In Proceedings of the 2001 Description Logic Workshop (DL*, 2001, pp10-19.

Chaplusky, H, 2000, "OntoMorph: A translation system for symbolic knowledge", Proc. 17th Intl. Conf. on Principles of Knowledge Representation and Reasoning (KR'2000)

Doan, ,Anhai, Madhavan, ,Jayant, Domingos, ,Pedro and Halevy, ,Alon, 2002. Learning to map between ontologies on the semantic web, *WWW '02: Proceedings of the 11th international conference on World Wide Web*, 2002, ACM pp662-673.

Ehrig, M. and Staab, S., 2004, "QOM – Quick Ontology Mapping", pp. 683-697.

Klien, M, 2001, "Combining and relating ontologies: an analysis of problems and solutions", Workshop on Ontologies and Information Sharing, IJCAI'01

Maedche, ,Alexander, Motik, ,Boris, Silva, ,Nuno and Volz, ,Raphael, 2002, "MAFRA - A MApping FRAmework for Distributed Ontologies", *EKAW '02: Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web*, 2002, Springer-Verlag pp235-250.

McGuiness, D.L., Fikes, R., RICE, J. and Wilder, S., 2000, "An environment for merging and testing large ontologies", Proc. 17th Intl. Conf. on Principles of Knowledge Representation and Reasoning (KR'2000)

Mitra, P. and Wiederhold, G., 2002, "Resolving terminological heterogeneity in ontologies", Workshop on Ontologies and Semantic Interoperability at the 15th European Conference on Artificial Intelligence (ECAI-2002)

Noy, N. F. and Musen, M. A., 2003, "The PROMPT suite: interactive tools for ontology merging and mapping", International Journal of Human-Computer Studies**,** Vol 59 , No. 6,pp 983 - 1024

Qadir, M. A., Fahad, M. and Noshairwan, M. W., 2007, "On conceptualization mismatches between ontologies", IEEE International Conference on Granular Compution (2007)

Stumme, G. and Maedche, A., 2001, "Ontology margining for federated ontologies on the semantic web", In Proceedings of the International Workshop for Foundations of Models for Information Integration (FMII-2001), pp 413-418

Visser, P.R.S., Jones, D.M., Bench-Capon, T.J.M. and Shave, M.J.R.,1997, An Analysis of Ontology Mismatches; Heterogeneity versus Interoperability In *AAAI1997 Spring Symposium on Ontological Engineering*, Stanford, USA.

Wiederhold, G, 1994, "An algebra for ontology composition", In Proceedings of 1994 Monterey Workshop on Formal Methods, pp 56-61