Empirical Knowledge Discovery over Ontology Matching Results

Ondřej Šváb-Zamazal and Vojtěch Svátek

University of Economics, Prague, Dept. Information and Knowledge Engineering
Nám. Winstona Churchilla 4, 130 67 Praha 3, Prague, Czech Republic
{ondrej.zamazal,svatek}@vse.cz

Abstract. Analysis of ontology alignments, as sets of correspondences between entities, can reveal knowledge to be later fed back to the alignment process. We report on data mining experiments over 3-year results of the 'conference' track of the Ontology Alignment Evaluation Initiative. The discovered hypotheses express relationships among the matching tools used, the nature of source ontologies, the confidence measure of the returned correspondences, their actual correctness, and, notably, the participation of the correspondences in mapping patterns.

1 Introduction

The heterogeneity of ontological conceptualisations on the Semantic Web is addressed by ontology matching tools, typically producing pairwise *alignments* consisting of numerous individual *correspondences*: pairs of corresponding entities, one from each of the two ontologies. As the alignments have to be not only discovered but also permanently stored, tested, searched over and updated, they are gradually becoming first-class citizens in the Semantic Web world [3]; they are even interpretable in terms of 'standard' ontological relationships such as equivalence or subsumption.

The correspondences are equipped with various characteristics related to their structural neighbourhood in both ontologies as well as to the process of their creation. It is natural to apply *inductive techniques* to discover hidden relationships among these characteristics. Such relationships provide feedback to setting up further *matching tasks* as well as to the improvement of the *matching algorithms*. In our earlier work we conceived data mining over ontology alignments as one of multiple methods of alignment evaluation [1, 7]. In this paper we extend this approach in terms of quantitative as well as qualitative aspects: more input datasets are examined, additional categories of mapping patterns are considered, more sophisticated analytical questions are posed to the mining tool, and the resulting hypotheses are more thoroughly interpreted.

Section 2 of the paper explains the origin of source data and the process of their preparation for the data mining process. Section 3 elaborates on a particular aspect of the data preparation: detection of so-called mapping patterns, yielding additional and potentially quite interesting data attributes. Section 4 briefly reviews the data mining tool used, namely, the 4ft-Miner procedure from the LISp-Miner toolbox. Section 5 presents the analytic questions posed to the mining tool, lists the strongest hypotheses discovered in return, and attempts to interpret these results in an aggregative manner. Finally, Section 6 surveys some related research, and Section 7 wraps up the paper.

2 Source Data

2.1 Origin of Data

The data used for mining were produced in the course of the *Ontology Alignment Evaluation Initiative* (OAEI), in the three consecutive runs (2006, 2007 and 2008) of one its tracks. The assignment to the participants of this track was based on ontologies from the *OntoFarm* collection.

The motivation for initiating the creation of the *OntoFarm* collection[1] (in Spring 2005) was the lack of 'manageable' material for testing ontology engineering (especially, mapping) techniques. As underlying domain, we chose that of *conference organisation*; motivations for this choice are elaborated in [8]. Each of the (small-to-medium sized) ontologies from the collection describes this domain from the point of view of a particular resource, which can be either a conference organisation *support tool* (yielding 'tool' ontologies, which are most frequent), experience of people with *personal participation* in conference organization (yielding 'insider' ontologies), or the content of *web pages* of concrete conferences (yielding 'web' ontologies). This results in the desired heterogeneity within a single domain, which to some degree emulates the real-world challenges faced by automated matching tools. The number of the ontologies has been constantly growing; between the first (2006) and last (2008) year of matching experiments considered in this paper it evolved from 10 to 15.

OAEI [2] is a coordinated international initiative that organises the evaluation of the increasing number of ontology matching systems. The main goal of OAEI is to to compare systems and algorithms on the same basis and to allow anyone for drawing conclusions about the best matching strategies. Every year there are several test cases (ontology pairs/collections) related to different domains, which emphasise different aspects of the matching needs; each of them constitutes a specific track of evaluation. As mentioned, the 'conference' track is based on the OntoFarm collection. The OAEI participants apply their matching systems on the test cases and send the resulting alignments, often including some numerical confidence of individual correspondences, to the OAEI organisers. The results are then evaluated in different ways, the most classical one being comparison with some ground truth (called reference alignment).

2.2 Structure of Data Matrix

For the data mining experiments we represented the individual correspondences each by a record in a single data table. The base attributes of this table (metadata) are:

– name of the matching *system* that detected this (occurrence of) correspondence
– *confidence* assigned to the correspondence by the system
– types of ontologies ('tool', 'insider', 'web') mentioned as *resource* in the tables 4 and 5
– correctness *result* manually assigned to the correspondence ('+' correct, '-' incorrect, 't' trivial exact string matching).

---

[1] http://nb.vse.cz/~svabo/oaei2008
[2] http://oaei.ontologymatching.org

In addition, there is information about *patterns* (those described in the next section) in which the given correspondence participates. There are two data fields for each of the eight patterns; the first one contains the *correctness* evaluation result of the *other* correspondence within the pattern (note that there are exactly two correspondences in each of these simple patterns), and the second one contains the *confidence* assigned to this other correspondence by the system.

In this paper we analyse the datasets containing alignments from three consecutive editions of the 'conference' track within the OAEI campaign (i.e. OAEI-06, OAEI-07 and OAEI-08). In the OAEI-06 there are 5238 records. In the OAEI-07 there are 10574 records and in the OAEI-08 there are 5234 records.
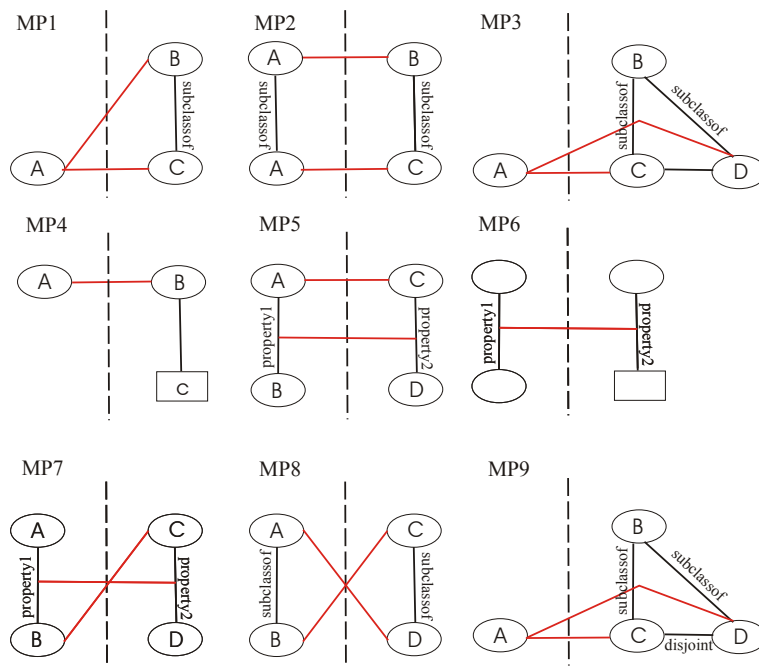

Fig. 1. Diagrams of mapping patterns

3   Introducing Mapping Patterns

As opposed to ontology design patterns,[3] which usually concern one ontology, mapping patterns deal with (at least) two ontologies. These patterns reflect the internal structure of ontologies as well as correspondences across the ontologies.

In this work we recognise three types of mapping patterns. Representatives of all three types (exactly those considered in this paper) are depicted in Figure 1.

---

[3] See e.g. http://ontologydesignpatterns.org.

3.1 Neutral Mapping Patterns

In our experiments we considered three patterns of this type:

– MP1 ('Parent-child triangle'): it consists of an equivalence correspondence between A classes and B and an equivalence correspondence between A and a child of B, where A and B are from different ontologies.
– MP2 ('Mapping along taxonomy'): it consists of simultaneous equivalence correspondences between parents and between children.
– MP3 ('Sibling-sibling triangle'): it consists of simultaneous correspondences between class A and two sibling classes C and D where A is from one ontology and C and D are from another ontology.

3.2 Correspondence mapping patterns

These mapping patterns are inspired by correspondence patterns proposed in [6]; again, three of them are considered here:

– MP4: it is inspired by the 'class-by-attribute' correspondence pattern, where the class in one ontology is restricted to only those instances having a particular value for a a given attribute/relation.
– MP5: it is inspired by the 'composite' correspondence pattern. It consists of a class-to-class equivalence correspondence and a property-to-property equivalence correspondence, where classes from the first correspondence are in the domain or in the range of properties from the second correspondence.
– MP6: it is inspired by the 'attribute to relation' correspondence pattern where a datatype and an object property are aligned as an equivalence correspondence.

3.3 Error mapping patterns

Finally, error mapping patterns can disclose incorrect correspondences; our inventory consists of the following three:

– MP7: it is the variant of MP5 'composite pattern'. It consists of an equivalence correspondence between two classes and an equivalence correspondence between two properties, where one class from the first correspondence is in the domain and the other class from that correspondence is in the range of equivalent properties, except the case where domain and range is the same class.
– MP8: it consists of an equivalence correspondence between A and B and an equivalence correspondence between a child of A and a parent of B where A and B are from different ontologies. It is sometimes referred to as criss-cross pattern.
– MP9: it is the variant of MP3, where the two sibling classes C and D are disjoint.

3.4 Summary

Neutral mapping patterns are neither desirable or undesirable. Their presence does not by itself lead to incorrectness or incoherency of alignment. *Error mapping patterns* are mapping patterns that are undesirable because they contain some logical incoherency. Finally, *correspondence mapping patterns* are desirable patterns that can be seen as good design practise for modelling complex correspondences. Rigorous categorisation of patterns is still subject to investigation; the current distinction is rather intuitive.

In Table 1 there are numbers of occurrences of mapping patterns in results of participants of OAEI-06, in Table 2 for OAEI-07 and in Table 3 for OAEI-08. We already see that some patterns are more typical for some systems than for other. Proper quantification of this relationship as well as its combination with other characteristics of correspondences is however the task for a mining tool.

System	MP1	MP2	MP3	MP4	MP5	MP6	MP7	MP8	MP9
HMatch	339	132	323	261	123	122	53	45	203
Automs	0	28	0	52	6	105	52	0	0
COMA	0	61	30	214	190	115	55	0	4
OWL-CtxMatch	64	36	56	44	54	105	53	5	13
Falcon	0	63	0	134	74	113	56	0	0
RiMOM	54	228	83	520	278	160	63	6	24

Table 1. Occurrences of mapping patterns in OAEI-06 participants results

System	MP1	MP2	MP3	MP4	MP5	MP6	MP7	MP8	MP9
ASMOV	32	146	65	442	463	347	42	0	20
Falcon	0	91	0	328	136	309	46	0	0
Lily	0	487	0	1009	1382	446	50	22	0
OLA	0	139	17	382	356	369	43	0	2
OntoDNA	0	26	0	128	69	301	41	0	0
SEMA	217	57	153	83	6	293	41	6	112

Table 2. Occurrences of mapping patterns in OAEI-07 participants results

4   4ft-Miner Overview

The *4ft-Miner* procedure is the most frequently used procedure of the *LISp-Miner* data mining system [5]. 4ft-Miner mines for association rules of the form $\varphi \approx \psi/\xi$, where φ, ψ and ξ are called *antecedent*, *succedent* and *condition*, respectively. Antecedent and succedent are conjunctions of *literals*. Literals are derived from attributes, i.e. fields

System	MP1	MP2	MP3	MP4	MP5	MP6	MP7	MP8	MP9
ASMOV	0	255	0	261	467	132	14	0	0
DSSim	543	146	527	828	115	115	16	7	165
Lily	0	115	0	354	585	151	18	4	0

Table 3. Occurrences of mapping patterns in OAEI-08 participants results

of the underlying data matrix; unlike most propositional mining system, they can be (at runtime) equipped with complex *coefficients*, i.e. value ranges. The association rule $\varphi \approx \psi/\xi$ means that on the subset of data defined by ξ, φ and ψ are associated in the way defined by the symbol $\approx$. The symbol $\approx$, called *4ft-quantifier*, corresponds to some statistical or heuristic test over the four-fold contingency table of φ and ψ. In the experiments below, we only used the *above average difference* (short, AvgDiff) quantifier, which expresses the relative increase of the frequency of succedent for the subset of data corresponding to the antecedent in comparison to the the frequency of succedent in the whole dataset. It is combined with the *support* (short, Supp) quantifier, expressing the relative frequency of objects satisfying both the antecedent and succedent.

The task definition language of 4ft-Miner is quite rich, and its description goes beyond the scope of this paper. Let us only declare two features of the tool that is important for our mining experiments: it is possible to formulate a wide range of *analytic tasks*, from very specific to very generic ones, and the underlying data mining algorithm is very fast thanks to highly optimised bit-string processing [5].

5 Mining Process, Results and Interpretation

Several analytic tasks were formulated over the data, of which five are presented here, each in a separate subsection. For each task, we list the strongest hypotheses in textual form, separately for each year of OAEI; each subsection is then concluded by a discussion and interpretation. Strong hypotheses are also listed formally in Table 4 for the first two tasks and in Table 5 for the remaining three tasks (the pattern-oriented ones). The asterisk in a column always means that the particular attribute was not used. Columns for the condition part are omitted.

5.1 Analytic task #1

Which systems and for what confidence values produce in/correct correspondences more often than others?

OAEI-06

Hypothesis t1: Correspondences that are produced by system RiMOM and have maximal confidence (i.e. 1) are by 111% (i.e. more than twice) more often correct than correspondences produced by all systems with all confidence values (on average).

Hypothesis t2: Correspondences that are produced by system Falcon and have confidence between 0.8 and 1.0 are by 90% (i.e. almost twice) more often correct than correspondences produced by all systems with all confidence values (on average).

Table 4. Hypotheses for tasks 1 and 2

		Antecedent			Succedent	Values	
	System	Confidence	Resource1	Resource2	Result	Supp	AvgDff
t1	RiMOM	1.0	*	*	+	0.1	1.11
t2	Falcon	$< 0.8; 1.0 >$	*	*	+	0.07	0.9
t3	RiMOM	$< 0.01; 0.43)$	*	*	-	0.13	0.9
t4	Falcon	1.0	*	*	t	0.05	2.64
t5	OLA	1.0	*	*	t	0.02	2.5
t6	OntoDNA	1.0	*	*	t	0.04	2.31
t7	Lily	1.0	*	*	-	0.29	0.22
t8	ASMOV	1.0	*	*	-	0.18	0.14
t9	Lily	$< 0.48; 1.0 >$	*	*	+	0.06	0.4
t10	ASMOV	$< 0.27; 0.75 >$	*	*	+	0.05	0. 26
t11	ASMOV	$< 0.01; 0.48)$	*	*	-	0.25	0.2
t12	RiMOM	1.0	*	tool	+	0.05	1.11
t13	RiMOM	1.0	tool	*	+	0.05	1.08
t14	Falcon	*	*	tool	t	0.05	1.45
t15	Lily	1.0	web	tool	-	0.09	0.31
t16	ASMOV	1.0	web	tool	-	0.05	0.23
t17	DSSim	$(0.75; 1.0 >$	tool	web	+	0.04	0.62
t18	ASMOV	$(0.01; 0.27)$	tool	web	-	0.08	0.34
t19	Lily	*	tool	*	+	0.07	0.34

Hypothesis t3: Correspondences that are produced by system RiMOM and have confidence between 0.01 and 0.43 are by 60% more often incorrect than correspondences produced by all systems with all confidence values (on average).

OAEI-07

Hypothesis t4: Correspondences that are produced by system Falcon and have maximal confidence (i.e. 1) are by 264% (i.e. more than three times) more often trivially correct[4] than correspondences produced by all systems with all confidence values (on average).

Hypothesis t5: Correspondences that are produced by system OLA and have maximal confidence (i.e. 1) are by 250% (i.e. more than three times) more often trivially correct than correspondences produced by all systems with all confidence values (on average).

Hypothesis t6: Correspondences that are produced by system OntoDNA and have maximal confidence (i.e. 1) are by 231% (i.e. more than three times) more often trivially correct than correspondences produced by all systems with all confidence values (on average).

Hypothesis t7: Correspondences that are produced by system Lily and have maximal confidence (i.e. 1) are by 22% more often incorrect than correspondences produced

---

[4] In the OAEI 2007 evaluation, a specific category of 'trivially correct' correspondences, namely, those between entities whose names are identical strings, was considered.

by all systems with all confidence values (on average) conditioned on the data annotated with the reference alignment.[5]

Hypothesis t8: Correspondences that are produced by system ASMOV and have maximal confidence (i.e. 1) are by 14% more often incorrect than correspondences produced by all systems with all confidence values (on average) conditioned on the data annotated with reference alignment.

OAEI-08

Hypothesis t9: Correspondences that are produced by system Lily and have confidence between 0.48 and 1.0 are by 40% more often correct than correspondences produced by all systems with all confidence values (on average).

Hypothesis t10: Correspondences that are produced by system ASMOV and have confidence between 0.27 and 0.75 are by 26% more often correct than correspondences produced by all systems with all confidence values (on average).

Hypothesis t11: Correspondences that are produced by system ASMOV and have confidence between 0.01 and 0.48 are by 20% more often incorrect than correspondences produced by all systems with all confidence values (on average).

Discussion We can cluster the hypotheses t1, t4, t5 and t6, declaring that particular systems tend to produce correct correspondences (RiMOM-06, Falcon-07, OLA-07, OntoDNA-07). Furthermore, systems RiMOM-06 and ASMOV-08 tend to produce incorrect correspondences with low confidence (hypotheses t3 and t11). In the case of systems Falcon-06, Lily-08, and ASMOV-08 they deliver correct correspondences with high confidence (t2, t9 and t10). On the other hand, systems in hypotheses t7 and t8 produce incorrect correspondences (Lily-07, ASMOV-07) with high confidence. However both these hypotheses only hold on the subset of results for which a reference alignment exists. Considering those hypotheses (t9, t10 vs. t7, t8) we can conclude that systems ASMOV-08 and Lily-08 improved against their previous-year versions.

5.2   Analytic task #2

Which systems, for what confidence values and on what types of ontologies produce in/correct correspondences more often than others? (The difference from task #1 is in also considering the types of ontologies.)

OAEI-06

Hypothesis t12: Correspondences that are produced by system RiMOM and have maximal confidence (i.e. 1) and ontology2 is based on tool are by 111% (i.e. more than twice) more often correct than correspondences produced by all systems for all types of ontologies and with all confidence values (on average).

Hypothesis t13: Correspondences that are produced by system RiMOM and have maximal confidence (i.e. 1) and ontology1 is based on tool are by 108% (i.e. more than twice) more often correct than correspondences produced by all systems for all types of ontologies and with all confidence values (on average).

---

[5] I.e. only those records where we had the result from the a priori made reference alignment (1337 records for OAEI-07), in other cases we used a posteriori evaluation.

OAEI-07

Hypothesis t14: Correspondences that are produced by system Falcon and ontology2 is based on tool are by 145% (i.e. more than twice) more often trivially correct than correspondences produced by all systems for all types of ontologies (on average).

Hypothesis t15: Correspondences that are produced by system Lily and have maximal confidence (i.e. 1) and ontology1 is based on web and ontology2 is based on tool are by 31% more often incorrect than correspondences produced by all systems for all types of ontologies and with all confidence values (on average).

Hypothesis t16: Correspondences that are produced by system ASMOV and have maximal confidence (i.e. 1) and ontology1 is based on web and ontology2 is based on tool are by 23% more often incorrect than correspondences produced by all systems for all types of ontologies and with all confidence values (on average).

OAEI-08

Hypothesis t17: Correspondences that are produced by system DSSim and have confidence between 0.75 and 1.0 and ontology1 is based on tool and ontology2 is based on web are by 62% more often correct than correspondences produced by all systems for all types of ontologies with all confidence values (on average).

Hypothesis t18: Correspondences that are produced by system ASMOV and have confidence between 0.01 and 0.27 and ontology1 is based on tool and ontology2 is based on web are by 34% more often incorrect than correspondences produced by all systems for all types of ontologies with all confidence values (on average).

Hypothesis t19: Correspondences that are produced by system Lily and ontology1 is based on tool are by 34% more often correct than correspondences produced by all systems for all types of ontologies (on average).

Discussion There are two conspicuous clusters of hypotheses. The first suggests that 'tool' ontologies are possibly aligned better than other types (hypotheses t12, t13, t14 and t19). The second suggests that aligning 'web' ontologies and 'tool' ontologies is risky. This could be explained by the fact that conference websites use similar terms as conference tools but with a different semantic flavour.

5.3 Analytic task #3

Which systems produce certain neutral mapping patterns more often than others?

OAEI-06

Hypothesis m1: Correspondences that are produced by system HMatch are by 217% (i.e. three times) more often part of MP1 than correspondences produced by all systems (on average).

OAEI-07

Hypothesis m2: Correspondences that are produced by system SEMA are by 2192% (i.e. 22 times) more often part of MP1 than correspondences produced by all systems (on average).

Hypothesis m3: Correspondences that are produced by system OLA are by 179% (i.e. almost three times) more often part of MP2 than correspondences produced by all systems (on average).

Table 5. Hypotheses for tasks 3, 4 and 5

	Antecedent		Succedent		Values	
	System	Confidence	ResultMP	ResultMP	Supp	AvgDff
m1	HMatch	*	MP1	*	0.06	2.17
m2	SEMA	1.0	MP1	*	0.02	21.92
m3	OLA	1.0	MP2	*	0.01	1.79
m4	DSSim	$< 0.75; 1.0 >$	MP1	MP2	0.05	1.68
m5	ASMOV	$< 0.01; 0.27)$	MP2	*	0.03	0.72
m6	OWL-CtxMatch	*	MP5	*	0.01	0.31
m7	COMA	*	MP4	*	0.03	0.26
m8	Falcon	*	MP4	*	0.03	0.34
m9	DSSim	*	MP4	*	0.16	0.54
m10	Lily	*	MP5	*	0.11	0.5
m11	HMatch	*	MP9	*	0.04	2.55
m12	SEMA	*	MP9	*	0.01	20.93
m13	DSSim	*	MP9	*	0.03	1.68

OAEI-08

Hypothesis m4: Correspondences that are produced by system DSSim and have confidence between 0.75 and 1.0 are by 168% (i.e. almost three times) more often part of MP1 and 2 than correspondences produced by all systems with all confidence values (on average).

Hypothesis m5: Correspondences that are produced by system ASMOV and have confidence between 0.01 and 0.27 are by 72% (i.e. almost twice) more often part of MP2 than correspondences produced by all systems with all confidence values (on average).

5.4 Analytic task #4

Which systems produce certain correspondence mapping patterns more often than others?

OAEI-06

Hypothesis m6: Correspondences that are produced by system OWL-CtxMatch are by 31% more often part of MP5 than correspondences produced by all systems (on average).

Hypothesis m7: Correspondences that are produced by system COMA are by 26% more often part of MP4 than correspondences produced by all systems (on average).

OAEI-07

Hypothesis m8: Correspondences that are produced by system Falcon are by 34% more often part of MP4 than correspondences produced by all systems (on average).

OAEI-08

Hypothesis m9: Correspondences that are produced by system DSSim are by 54% more often part of MP4 than correspondences produced by all systems (on average).

Hypothesis m10: Correspondences that are produced by system Lily are by 50% more often part of MP5 than correspondences produced by all systems (on average).

Discussion Regarding neutral mapping patterns, system ASMOV found MP1 for correspondences with low confidence, while DSSim with high confidence. Hypotheses m7, m8, m9 refer to usage of context for matching.

5.5 Analytic task #5

Which systems produce certain error mapping patterns more often than others?

OAEI-06

Hypothesis m11: Correspondences that are produced by system HMatch are by 255% (i.e. more than three times) more often part of MP9 than correspondences produced by all systems (on average).

OAEI-07

Hypothesis m12: Correspondences that are produced by system SEMA are by 2093% (i.e. almost 22 times) more often part of MP9 than correspondences produced by all systems (on average).

OAEI-08

Hypothesis m13: Correspondences that are produced by system DSSim are by 168% (i.e. almost three times) more often part of MP9 than correspondences produced by all systems (on average).

Discussion According to abovementioned hypotheses with error mapping patterns we can conclude which systems could be improved in terms of avoiding inconsistent correspondences. From this point of view, we can say that application of error mapping patterns would improve the systems' performance of HMatch, SEMA, and DSSim. None of these systems explicitly describes whether they use some kind of *verification* phase during an ontology matching process. On the other hand, the ASMOV system (from the OAEI-07 and the OAEI-08) verifies alignments in terms of consistency. We can expect that other OM tools also verify their results but they are not always clear at this point.

6 Related Work

Data mining of a kind was used for ontology matching by Ehrig [2]. However, unlike our approach, this was supervised machine learning rather than mining data for frequent associations.

The relationship between matching tools and various features of the matching task was studied by Mochol&Jentzsch [4] in the context of matching tool recommender development. The rule base was created manually, based on analysis of literature describing the tools. The focus of their work is on the predictive task, i.e. efficient recommendation. In this sense their approach is perfectly complementary to our, empirical and descriptive one.

7 Conclusion and Future Work

Inductive knowledge discovery techniques are a promising means for getting insight into the large sets of correspondences output by the ontology matching tools. They can provide the tool developers as well as end users with systematic feedback, leading to improvement of the tools as well as to the selection of the most suitable tool for a certain task. Association mining, as performed by 4ft-Miner, has proven adequate for this problem.

In the future we also plan to exploit *other inductive procedures* that are part of the LISp-Miner toolbox. An interesting option would be to use the related SD4ft-Miner procedure, which allows to discover features in which one set of objects (here, for example, correspondences output by one system) most differs from another set (correspondences output by another system). We also plan to design a methodology for exploiting the mining results in the *matching tool recommendation* process as described in [4].

Acknowledgments

References

1. Caracciolo C., Euzenat J., Hollink L., Ichise R., Isaac A., Malais V., Meilicke Ch., Pane J., Shvaiko P., Stuckenschmidt H., Šváb-Zamazal O., Svátek V.: First results of the Ontology Alignment Evaluation Initiative 2008. In: OM-2008 at ISWC-2008.
2. Ehrig M., Staab S., Sure Y.: Bootstrapping Ontology Alignment Methods with APFEL In: Proceedings of ISWC, Galway, Ireland, 2005.
3. Euzenat, J., Mocan, A., Scharffe, F.: Ontology Alignments, An Ontology Management Perspective. In: Ontology Management, Springer 2007, pp.177–206.
4. Mochol, M., Jentzsch, A.: Towards a rule-based matcher selection. In: Proc. EKAW 2008.
5. Rauch, J., Šimůnek, M.: An Alternative Approach to Mining Association Rules. In: Lin, T. Y., Ohsuga, S., Liau, C. J., Tsumoto, S. (eds.), Data Mining: Foundations, Methods, and Applications, Springer-Verlag, 2005, pp. 211–232
6. Scharffe F., Fensel D.: Correspondence Patterns for Ontology Alignment. In: Proc. EKAW-2008.
7. Šváb O., Svátek V., and Stuckenschmidt H.. A study in empirical and 'casuistic' analysis of ontology mapping results. In: Proceedings of ESWC 2007.
8. Šváb O., Svátek V., Berka P., Rak D., Tomášek P.: OntoFarm: Towards an Experimental Collection of Parallel Ontologies. In: Poster Session at ISWC 2005.