A Study in Empirical and 'Casuistic' Analysis of Ontology Mapping Results

Ondřej Šváb[1], Vojtěch Svátek[1] and Heiner Stuckenschmidt[2]

[1] Department of Information and Knowledge Engineering,
University of Economics, Prague, W. Churchill Sq. 4, 130 67 Praha 3, Czech Republic
e-mail: {svabo,svatek}@vse.cz
[2] Universität Mannheim, Institut für Informatik, A5, 6 68159 Mannheim, Germany
e-mail: heiner@informatik.uni-mannheim.de

Abstract. Many ontology mapping systems nowadays exist. In order to evaluate their strengths and weaknesses, benchmark datasets (ontology collections) have been created, several of which have been used in the most recent edition of the Ontology Alignment Evaluation Initiative (OAEI). While most OAEI tracks rely on straightforward comparison of the results achieved by the mapping systems with some kind of reference mapping created a priori, the 'conference' track (based on the *OntoFarm* collection of heterogeneous 'conference organisation' ontologies) instead encompassed multiway manual as well as automated analysis of mapping results themselves, with 'correct' and 'incorrect' cases determined a posteriori. The manual analysis consisted in simple labelling of discovered mappings plus discussion of selected cases ('casuistics') within a face-to-face consensus building workshop. The automated analysis relied on two different tools: the DRAGO system for testing the consistency of aligned ontologies and the *LISp-Miner* system for discovering frequent associations in mapping meta-data including the phenomenon of graph-based mapping patterns. The results potentially provide specific feedback to the developers and users of mining tools, and generally indicate that automated mapping can rarely be successful without considering the larger context and possibly deeper semantics of the entities involved.

1 Introduction

Ontologies can help integrate semantic views on real-world data. Unfortunately, designers of ontologies themselves apply different views of the same domain during ontology development. This yields semantic heterogeneity at ontology level, which is one of main obstacles to semantic interoperability. *Ontology mapping* (also called 'matching' or 'alignment') is the core component of approaches attempting to solve this problem. It consists in finding mappings (also called 'correspondences') among entities (classes, relations) from different ontologies. The set of mappings is called *alignment*. The process of mapping is followed by ontology merging, ontology transformation, data transformation etc. A survey of ontology mapping methods is e.g. in [11].

It is important to have means to *evaluate* the quality of mapping, and, consequently, the fitness of different methods and tools with respect to different domains and settings. Nowadays, the central approach to ontology mapping evaluation is based on the notion

of *reference alignment* ('gold standard'), defined a priori, to which the results obtained by the matching systems are compared. This typically yields measures borrowed from the discipline of Information Retrieval, such as *precision* (the proportion of mappings returned by the matching system that are also present in the reference mapping) and *recall* (the proportion of mappings present in the reference mapping that are also returned by the matching system). The correspondences in both the reference and experimental alignments are most often expressed as simple concept-concept (or relation-relation) pairs, interpreted as *logical equivalence*. Sometimes, alignments interpreted as *logical subsumption* (analogously to the same notion as omnipresent in ontology design), and/or with a non-Boolean value of *validity* are also considered. However, we might even be interested in more complex *alignment structures* (patterns), which could reveal interesting details about the relationship of the two ontologies—for example, the situation when an entity from one ontology can potentially be mapped on both a parent and a child from the other ontology.

In case there is no reference alignment (and providing it manually would be unacceptably tedious) or we are interested in more complex phenomena—say, *mapping patterns*—arising in ontology alignment, novel methods for mapping evaluation have to be devised. Let us outline four of them that are focal in this paper; while the first and the third are manual, the second and the fourth rely on automated procedures.

– Instead of formulating a reference alignment *a priori*, the mappings discovered by the system (which are often just a small fraction of the carthesian product of the sets of entities from two ontologies) can be *a posteriori* examined and *labelled* (as in/correct or possibly using a richer set of labels) by human evaluator/s. Such evaluation naturally lacks the rigour of 'blindfold' evaluation wrt. an (unbiased) reference alignment, and does not tell much about the *recall*. Still, the *precision* figure may be valuable[3]; and its subjective bias can be reduced via recourse to *multiple* (ideally, expert) evaluators.
– Automated *reasoning* over aligned ontologies. A complement to manual labelling of the mappings is the exploitation of an inference procedure that is usually considered as first step in exploiting ontologies: concept satisfiability testing. Clearly, mappings that incur inconsistency to ontologies (that were consistent as long as standalone) are potentially inadequate.
– A side-product of both manual labelling and automated consistency checking can be a list of 'interesting' (ambiguous, dubious, surprising etc.) mappings. In order to get an overview of typical *reasons* (or 'arguments') for success/failure of automated mapping, some individual 'interesting' cases (not only the entities mapped but also their context within the ontologies and possibly some metadata about the mapping process) can be submitted to a *discussion board*. The outcome of discussion is definitely of different nature than that of quantitative evaluation, but can lead to complementary feedback to the developers of mapping tools. This discussion can also help identify candidate *mapping patterns* to be quantitatively evaluated in further analysis (see next item). We can view this approach as analogous to *casuistic* medical studies/literature, which is also sometimes used as complement to the

---

[3] Cf. the discussion on prior and posterior precision in *ontology learning* [6].

nowadays dominant empirical (evidence-based) one. In the context of this paper, the 'casuistic' approach is incarnated in the *consensus building workshop*.

– Large-scale *mining* over the mapping results and associated meta-data. The input to the mining process can be not only the name of the *mapping system*, name and nature of the *ontologies* mapped, the *type* of mapping (such as equivalence/subsumption), its *validity estimate* (by the system) and the subjective *posterior evaluation* ('correct'/'incorrect'), but also the information whether the given mapping is part (and what part) of a certain *mapping pattern*. We believe that the hypotheses discovered via data mining (over ontology mapping data including information about mapping patterns), in particular, mining for *frequent associations*, can become useful feedback to the development and tuning of mapping tools, complementary to the feedback provided by Information Retrieval measures with respect to reference mapping.

The paper is structured as follows. Section 2 surveys the background of the current research: the underlying *ontology collection* and the *international initiative* (OAEI) within which the automated mapping experiments took place. Section 3 reports on the evaluation via *manual labelling*. Section 4 deals with *reasoning*-based evaluation, using the *Drago* distributed description logic (DDL) tool. Section 5 describes the *consensus building workshop* in which selected discovered mappings were discussed by humans. Section 6 presents a simple typology of *mapping patterns* of interest (preceded with a short expose on 'pattern' issues in ontological engineering in general). Section 7 is devoted to the *data mining* effort. Finally, section 8 surveys some related research, and section 9 wraps up the paper.

2 Project Background

2.1 *OntoFarm* Collection

The motivation for initiating the creation of the *OntoFarm*[4] collection (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*—among other, for the following reasons:

– Most ontology engineers are academics who themselves submit and review papers and organise conferences: there is zero overhead of acquiring the *domain expertise*.
– Organisation of a conference shares some aspects with (heavier-weighted) *business activities*: access restrictions, hard vs. soft constraints, temporal dependencies among events, evolution of the meaning of concepts in time etc.
– There is also a wide range of supporting *software tools* covering various aspects of conference organisation. Their domain assumptions can also be captured using ontologies (specific for each system). The process of matching the requirements of conference organisers with the capacities of such tools is analogous with that

---

[4] See `http://nb.vse.cz/~svabo/oaei2006`; the development of the collection is described in more detail in [12].

of matching the requirements of a business with the capacities of an off-the-shelf enterprise information system.

– In many cases, even the underlying *instance data* could be obtained, since legal restrictions are typically not as strong as e.g. in business or medicine.

The snapshot of the (constantly growing) collection used for the 2006 OAEI track, see below, consisted of ten OWL-DL ontologies, typically of the size of 30–80 concepts and 30–60 properties, most of them being endowed with DL axioms. The overview is in Table 1. Six among the ontologies ('tool' ontologies) were derived from different conference organisation *support tools* (for the review process, registration etc.), using their documentation and experiments with installed tools; two of them ('insider' ontologies) are based on the experience of people with *personal participation* in conference organization; finally, two of them ('web' ontologies) are merely based on the content of *web pages* of concrete conferences. The ontology designers (partly students of a course on Knowledge Modelling and partly experienced knowledge engineers) did not interact among themselves. This should guarantee that, although the ontologies themselves are to some degree 'artificial' (their development not being driven by an application need), their *heterogeneity* was introduced in a 'natural' way, that possibly simulating the heterogeneity of ontologies developed by different communitites in the real world.

Name	Type	Number of Classes	Number of Properties	DL expressivity
EKAW	Insider	77	33	SHIN(D)
SOFSEM	Insider	60	64	ALCHIF(D)
SIGKDD	Web	49	28	ELI(D)
IASTED	Web	140	41	ALCIF(D)
Confious	Tool	57	57	SHIN(D)
PCS	Tool	23	38	ELUIF(D)
OpenConf	Tool	62	45	ALCIO(D)
ConfTool	Tool	38	36	SIF(D)
CRS	Tool	14	17	DL-Lite
CMT	Tool	36	59	ALCIF(D)

Fig. 1. Characteristics of ten *OntoFarm* ontologies.

2.2 *OAEI 2006* Initiative

The Ontology Alignment Evaluation Initiative[5] (OAEI) is a coordinated international initiative that organizes 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

---

[5] http://oaei.ontologymatching.org

[3]. The first OAEI evaluation campaign was presented at the workshop on Integrating Ontologies held in conjunction with the International Conference on Knowledge Capture (K-Cap) 2005. The outcomes of the 2006 campaign were then presented at the Ontology Matching (OM-2006) workshop at ISWC, in Athens, Georgia, USA. There were six different test cases (ontology pairs/collections), related to different domains, which emphasised different aspects of the matching needs; each of them constituted a specific track of evaluation. Four of the tracks ('benchmark', 'anatomy', 'jobs', 'directory') relied on some sort of pre-defined reference mappings to which those discovered by the systems could be compared (resulting in standard relevance measures such as precision/recall). The remaining ones ('food', 'conference') lacked such reference mappings, but the results were evaluated a posteriori. Here we concentrate on the 'conference' track, which was based upon the aforementioned *OntoFarm* collection and culminated at the OM-2006 *consensus building workshop*.

3 Initial Manual Evaluation

There were six participant groups to the 'conference' track, with mapping systems[6] named Automs, Coma++, OWL-CtxMatch, Falcon, HMatch and RiMOM. The alignments obtained were examined by the organisers, and each individual mapping was assigned a label. Results from the initial evaluation phase are on the result report page[7]; these consist in global statistics about the participants' results, which more-or-less reflect their quality.

The global statistics for each system amount to (among other):

– the distinction whether the mapping is true/false or is scaled between 0 and 1
– number of alignments (i.e. ontology pairs)
– number of individual mappings labelled as 'correct' vs. 'incorrect'
– number of 'interesting correct' mappings, namely, those that were subjectively 'not so easy to identify' at first sight (e.g. due to lack of string similarity)
– number of mappings that seemed to exhibit an interesting type of error (or problematic feature), specifically for: *subsumption* mistaken for equivalence, *sibling* concepts mistaken for equivalent ones, mutually *inverse* properties mapped on each other, *relation* mapped onto *class*
– *precision* as ratio of the number of all correct mappings to the number of all mappings, and *relative recall* as ratio of the number of all correct mappings to the number of correct mappings found by *any of the systems*

Additionally, some of the mappings that were retained as 'worth discussing' by both independent evaluators were then submitted to the consensus building workshop.

---

[6] Descriptions of the systems are in the OAEI 2006 papers available from `http://om2006.ontologymatching.org/#ap`.

[7] `http://nb.vse.cz/~svabo/oaei2006/`

4 Evaluation via Logical Reasoning

In addition to the manual evaluation of the generated mappings, we conducted an automatic analysis on a subset of the mappings. In this automatic analysis mappings between class names in the different ontologies were formalized in C-OWL [1] and the DRAGO system [10] was used to determine whether the mappings created by a particular system cause *logical inconsistencies* in one of the mapped ontologies. A more detailed description of the approach can be found in [7]. The analysis was performed on a subset of the ontologies. Mappings between the ontologies SOFSEM, IASTED, Confious and OpenConf were not considered because they could not be processed by DRAGO. Further, we restricted the analysis to four matching systems, namely Falcon, OWL-CTXmatch, COMA++ and HMatch. The analysis can easily be extended to include the other two participating systems, though.

Table 2 shows the results of the reasoning-based analysis. Note that the precision only refers to mappings between *class* names and therefore naturally differs from the numbers at the result report page `http://nb.vse.cz/~svabo/oaei2006/`. The precision has been determined by a manual investigation of the mappings by three independent people (different from those doing essentially the same task for the sake of the result report page). In cases of a disagreement the correctness of a correspondence was decided by a majority vote. It however turned out that there was very little disagreement with respect to the correctness of correspondence. For only about 3% of the correspondences the result had to be determined by vote.

System	Inconsistent alignments	Avg. number of inconsistent concepts	Overall Precision
Falcon	4	1,5	89,7 %
OWL-CTXmatch	6	9,6	85,67 %
Coma	12	2,2	67,7 %
HMatch	9	5,5	63,7 %

Fig. 2. Results of Reasoning-Based Evaluation

The results of this evaluation are useful in two ways. First of all, we can see from the numbers that a low number of inconsistent alignments is an indicator for the quality of mappings (we also see that the actual number of concepts that become unsatisfiable is less relevant). The second benefit of this evaluation is the fact that the information about inconsistent concepts and mappings that caused these inconsistencies reveal obvious and also non-obvious errors in mappings. Some examples of obviously incorrect mappings produced by matching systems in the experiments are the following:

$$Document = Topic$$
$$Decision = Location$$
$$Reception = Rejection$$

The real benefit of this evaluation is its ability to find non-obvious errors in mappings that can only be detected taking the position of the mapped concepts in the concept hierarchy into account. In our experiments, we found a number of such errors. Examples include the following mappings:

$$Regular\_Paper = Regular$$
$$Reviewing\_event = review$$
$$Main\_office = Location$$

In the case of the first correspondence, *Regular* actually denotes the regular participation fee as opposed to the early registration. The error in the second correspondence is caused by the fact that *Reviewing_event* represents the process of reviewing whereas review denotes the review document as such. The last correspondence is not correct, because the concept *Main_office* actually represents the main office as an organizational unit rather than a location. Such mappings are candidates for a closer inspection in terms of a committee of experts that analyze the reason for the inconsistency and decide whether the problem is in the mapping or in the ontologies[8].

5 Consensus Building Workshop

The idea of consensus building workshop was to discuss some interesting mappings in detail. Such interesting mappings can be determined as a result of the manual and the automatic evaluation of the matching results, as shown above. In the case of the *manual evaluation* mappings where the evaluators where in doubt or where they disagreed on the correctness of a mapping are candidates for a consensus workshop. In the *automatic evaluation*, mappings that have been shown to cause concepts in the mapped ontologies to become inconsistent are such candidates, especially if the mappings have been annotated as being correct in the manual evaluation. Often, a decision whether a mapping is correct or not can be made quite easily in a committee of experts. In some cases, however, it turns out that deciding whether a mapping is correct or not is far from being trivial. In particular, it turns out that sometimes a detailed analysis of the mapped ontologies is necessary to come to a decision.

12 mappings were selected for discussion. In the following, we focus on examples that illustrate the kinds of arguments used in the discussion and the insights gained.

Person vs. Human At first sight the equivalence between the concepts person and human looks rather intuitive, it is however not obvious that the two concepts have the same intended meaning in different ontologies. First of all, the concept person can be interpreted in a legal context in which it also refers to organizations. Further, when we look at the hierarchies of the different ontologies, we see that the concepts have completely different sets of subconcepts depending on the scope of the ontology (compare figure 3.

---

[8] They were not submitted to the consensus building workshop as the results of logical reasoning were available too late wrt. workshop preparation.

(a) IASTED   (b) SIGKDD
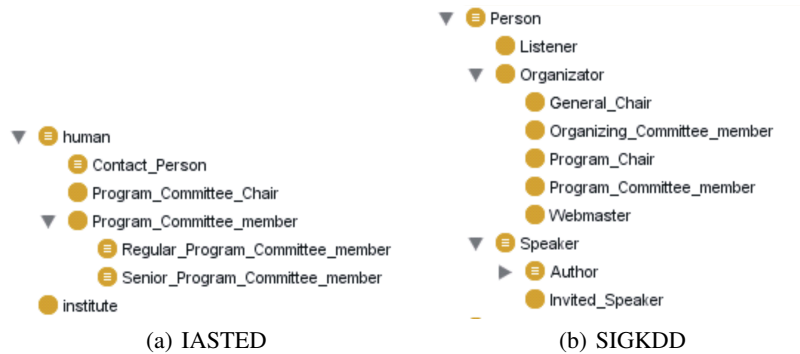
**Fig. 3.** Subtrees rooted at the concepts *Human* and *Person*

As we can see, the notion of *Person* in SIGKDD also contains subclasses not subsumed under *Human* in IASTED (e.g. speakers). As it is clear, however that both ontologies cover the same domain, it was decided that in this case the two concepts actually have the same intended meaning even though they do not share all subclasses.

PC_Member vs. Member_PC The concepts *PC_member* and *member_PC* are another example of mappings that seem to be trivially correct at first sight. In this case the question is whether the ontologies assumes the same set of people to belong to the program committee. A look at the hierarchies reveals that the two ontologies use a different interpretation of the set of people belonging to the PC. In particular in one case the *PC_chair* is assumed to be a member of the committee, in the other case not (compare figure 4). This seems to imply that the notion of *PC_member* in EKAW is more general than that in *ConfTool*. However, this is only the case if we assume that the concepts *Chair_PC* und *PC_Chair* are equivalent. Another possible interpretation is that the concepts *PC_member* and *Member_PC* are equivalent but *Chair_PC* and *PC_Chair* are different concepts, namely one denoting PC chairs that are members of the PC and the other denoting PC chairs that are not member of the PC. While both interpretations are possible, the majority of workshop participants favored the first interpretation where PC chairs are the same concepts.

Rejection vs. Reject Another mapping under discussion was the one between the concepts *Reject* and *Rejection*. It is clear that both are closely related to the outcome of the review of a submitted paper. Differences were only detected when looking at the subtrees of the superconcepts. While *Rejection* is a subconcept of *Decision*, *Reject* is defined as a subconcept of *Recommendation*. Understanding the difference between these two requires a deeper understanding of the process of reviewing, namely that a recommendation is the input for the final evaluation and the decision is the output.

Location vs. Place A similar situation could be observed in connection with the concepts *Location* and *Place*. Both concepts are closely related as they refer to some geographical entity. A closer look however reveals that they are used in a very different

(a) EKAW   (b) ConfTool
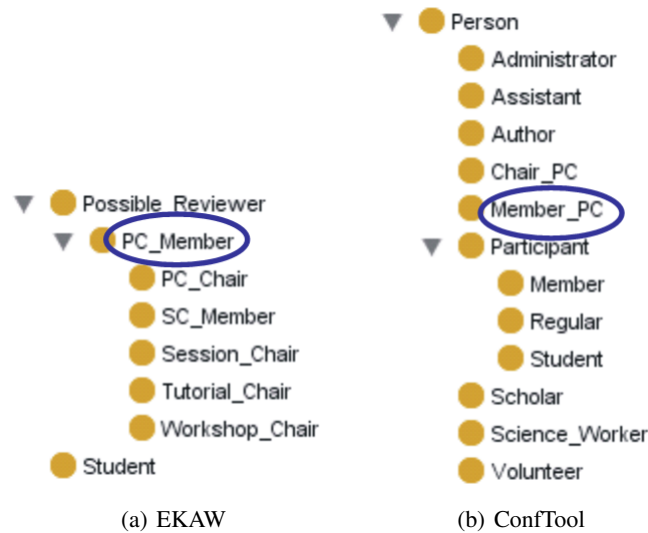
Fig. 4. Subtrees containing the concepts *PC_Member* and *Member_PC*

way. While *Location* refers to the country and city in which the conference is held, *Place* refers to buildings and parts of buildings in which certain conference-related events take place. The detection of this fundamental difference required a detailed analysis of the ontologies, in particular the range and domain restrictions of related properties in the ontologies.

5.1 Lessons Learned

The discussions at the consensus workshop revealed a number of insights about the nature of ontology matching and limitations of existing systems that provide valuable input for the design of matching tools. In the following we summarize the three most important insights gained.

Relevance of Context Probably the most important insight of the consensus workshop was that in many cases it is not enough to look at the concept names to decide whether a mapping is correct or not. In all of the examples above, the position of the concept in the hierarchy and in some cases also the scope of the complete ontology had to be taken into account. In some cases, a decision actually requires deep ontological arguments, for instance to distinguish between a recommendation and the actual decision made on the basis of this recommendation. For existing matching tools this means that the use of lexical matching techniques and often even of local structure matching is not sufficient. Matchers rather have to take the complete ontology and its semantics or even background knowledge about basic ontological distinctions into account. This observation is also supported by the results of the reasoning-based evaluation where automatically created mappings often turned out to cause inconsistencies in the ontologies.

Semantic Relations All of the systems participating in the evaluation were restricted to detecting equivalences between concepts or relations respectively. It turned out that this restriction is a frequent source of errors. Often ontologies contain concepts that are closely related but not exactly the same. In many cases one concept is actually a subclass of the other. Heuristics-based matching tools will often claim these concepts to be equivalent, because they have similar features and similar positions in the hierarchy. As a result, the corresponding mapping often becomes inconsistent. We believe that matching tools that are capable of computing subsumption rather than equivalence relations are able to produce more correct and suitable mappings.

Alternative Interpretations The example of *PC_member* illustrates the fundamental dilemma of ontology matching, which tries to determine the intended meaning of concepts based on a necessarily incomplete specification. As a result, it is actually not always possible to really decide whether a mapping is correct or not. All we can do is to argue that a mapping is consistent with specifications in the ontologies and with the other mappings. In the example this leads to a situation where we actually have two possible interpretations each of which makes a different set of mappings correct. It is not completely clear how this dilemma can be handled by matching tools. The only recommendation we can give is in favor of using methods for checking the consistency of mappings as an indicator whether the mapping encodes a coherent view on the system.

6 Mapping Patterns

Apart from the low-scale 'casuistic' approach of the consensus building workshop, we so far considered two empirical techniques: manual labelling, which partly accounted for structural issues (such as subclassing mistaken for equivalence, see section 3), and automatic consistency checking, which however only returned results in terms of in/consistent concepts (and not about structure). A useful complement would then be an *automated* analysis technique also considering the *structure* of mappings. We believe that a state-of-the-art *data mining tool* can fit to our purposes (see section 7), provided we enrich the available data with explicit representation of structural *mapping patterns*.

6.1 Patterns in Ontological Engineering

Before starting to talk about mapping patterns, it could be useful to briefly discuss the notion of patterns as typically treated in ontological engineering research. We will consider three categories of patterns: content patterns, logical patterns and frequent errors.

Content patterns [4] use specific non-logical vocabulary and describe a recurring, often domain-independent state of affairs. An example is the "Descriptions&Situations" pattern, which reflects the typical way a situation (with various entities and events involved) is described using some representation. *Logical patterns*, in turn, capture the typical ways certain modelling problems can be tackled in a specific ontological language. An example is the "Classes as Property Values" pattern[9], which defines multiple

---

[9] http://www.w3.org/TR/swbp-classes-as-values/

ways to satisfy the need for using a class in place of a value of an OWL property. Finally, *frequent errors* (though not usually denoted as patterns, they are clearly so) describe inadequate constructions that are often used by unexperienced modellers [9].

All three mentioned types of patterns are used to describe modelling behaviour that considered as either 'desirable' (content and logical patterns) or 'undesirable' (frequent errors). They can be qualified as *design* patterns; indeed, ontology building is essentially an activity carried out by human intellect (at least at the level of defining logical axioms, which are hard to obtain via automated ontology learning). In contrast, *mapping patterns* that will be discussed further are by themselves neither desirable nor undesirable; their desirability depends on the correctness of the mappings. They don't result from a deliberate activity by humans but can be detected in data output by automated mapping systems.

6.2 Inventory of Simple Mapping Patterns

As opposed to ontology design patterns, which concern one ontology, mapping patterns deal with (at least) two ontologies. These patterns reflect the *structure of ontologies* on the one side, and on the other side they include *mappings* between elements of ontologies. A mapping pattern is a graph structure, where nodes are classes, properties or instances. Edges represent mappings, relations between elements (eg. domain and range of properties) or structural relations between classes (eg. subclasses or siblings).
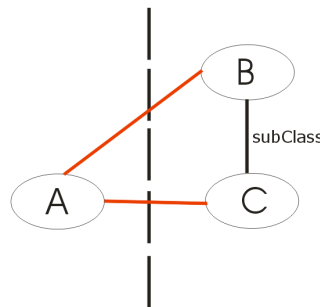

Fig. 5. Pattern 1 – 'Parent-child triangle'

The simplest (trivial) mapping pattern we do not consider here only contains one element from each of the two ontologies (let us call them O1 and O2), and a mapping between them. In our data mining experiments (see next section) we employed three slightly more complex mapping patterns.

The first one is depicted in Figure 5. The left-hand side (class A) is from O1 and the right-hand side (class B and its subclass C) is from O2. There is a mapping between A and B and at the same time between A and C.

The second pattern is depicted in Figure 6. It is quite similar to the previous one, but now we consider a child and a parent from each ontology and simultaneous mappings between parents and between children.
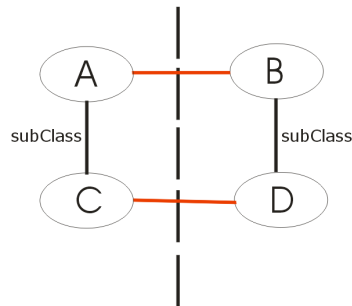
Fig. 6. Pattern 2 – 'Mapping along taxonomy'

The third mapping pattern we consider is depicted in Figure 7. It consists of simultaneous mappings between class A from ontology O1 and two sibling classes C and D from ontology O2.
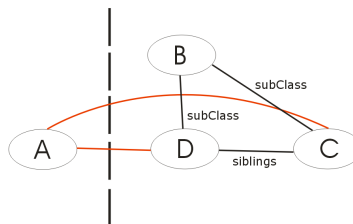
Fig. 7. Pattern 3 – 'Sibling-sibling triangle'

7 Evaluation via Data Mining

7.1 *4ft-Miner* Overview

The *4ft-Miner* procedure is the most frequently used procedure of the *LISp-Miner* data mining system [8]. *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 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.

The task definition language of *4ft-Miner* is quite rich, and its description goes beyond the scope of this paper. Let us only declare its two features important for our mining task: it is possible to formulate a wide range of so-called *analytic questions*, from very specific to very generic ones, and the underlying data mining algorithm is very fast thanks to highly optimised bit-string processing [8].

7.2 Using *4ft-Miner* for Mining over Mapping Results

For the purpose of data mining, a data matrix with each record capturing all information about one (occurrence of) correspondence was built. This elementary information amounted to: name of mapping *system* that detected this (occurrence of) correspondence; *validity* assigned to the correspondence by the system; types of *ontologies* ('tool', 'insider', 'web') on both sides of the correspondence; 'correctness' *label* manually assigned to the correspondence (cf. section 3). In addition, there is information about *patterns* (those from the previous section) in which the given correspondence participates. There are two data fields for each of the three patterns; the first one contains the 'correctness' *label* of the *other* mapping within the pattern (note that there are exactly two mappings in each of these simple patterns), and the second one contains the *validity* assigned to this *other* mapping by the system.

The *analytic questions* (i.e. task settings) we formulated for *4FT-Miner* were for example as follows[10]:

1. Which systems give higher/lower validity than others to the mappings that are deemed 'in/correct'?
2. Which systems produce certain mapping patterns more often than others?
3. Which systems are more successful on certain types of ontologies?

Due to limited space we do not list complete nor detailed results of the data mining process. We only present some interesting association hypotheses discovered.

For the first question, we found for example the following hypotheses:

– Correspondences output by Falcon with medium validity (between 0,5 and 0,8) are almost twice more often 'incorrect' than such correspondences output by all systems (on average).
– Correspondences output by RiMOM and by HMatch with high validity (between 0,8 and 1,0) are more 'correct' than such correspondences output by all systems (on average).

For the second question, we found for example the following hypotheses:

---

[10] Actually, the questions were even more generic; however, their generic form is less elegant when translated to natural language.

- Correspondences output by HMatch with medium validity (between 0,5 and 0,8) are more likely to connect a child with a class that is also connected (with high validity) with a parent of the child (Pattern 1) than such correspondences with all validity values (on average).
- Correspondences output by RiMOM with high validity (between 0,8 and 1,0) are more likely to connect class C with class D whose parent B is connected (with high validity) with A, which is parent of C (Pattern 2), than such correspondences with all validity values (on average).

These two hypotheses seem to have a natural interpretation (at the level of patterns, perhaps not so at the level of mapping systems). Pattern 1 represents a potential mapping conflict (aka love triangle with a father and a son competing for the same woman), i.e. increasing the validity of one may lead to decreasing the validity of the other. On the other hand, Pattern 2 seems to evoke positive feedback between the two mappings (as might be the case when a father is interested in the mother of his son's girlfriend).

A feature of the *OntoFarm* collection that was clearly beneficial for the 'data mining' approach to mapping evaluation was the fact that it contains (far) more than two ontologies that can be matched. Thanks to that, mapping patterns frequently arising because of the specific nature of some ontologi/es could be separated from mapping patterns that are frequent in general by its nature.

8 Related Work

To our knowledge, there has been no systematic effort in posterior analysis of ontolology mappings without reference alignment involving multiple methods like in our research. There are only projects with which we share some isolated aspects.

Mapping patterns are considered in [5]; however, they focus on 'heterogeneous mappings' (class to property) as special kind of pattern. We also considered this, but it appeared too infrequently (essentially, it was only output by the Coma++ system) to allow for meaningful data mining.

Data mining of a kind was also used for ontology mapping by Ehrig [2]. However, unlike our approach, this was supervised Machine Learning rather than mining data for frequent associations.

9 Conclusion and Future Work

The purpose of the current study was to examine multiple, principally different methods of posterior evaluation of ontology mappings, focussing on the situation when there is no reference mapping available and/or we want to get deeper insight into the nature of mappings. Our results could have at least two potential uses: to give the authors of individual *mapping systems* feedback on strong and weak points of the systems (going far beyond the usual precision/recall statistics), and to contribute to better insight of the whole *research community* into the possible argumentation used in the ontology mapping process.

In the future, we would like to more thoroughly compare the outcomes of the different methods used (manual labelling, board discussion, data mining, logical reasoning). We would also like to consider a richer variety of ontology mapping patterns. An important task is also to increase the size and improve the quality of *OntoFarm* collection, which would presumably be used in the next OAEI edition.

References

1. Bouquet, P., Giunchiglia, F., van Harmelen, F., Serafini. L., Stuckenschmidt, H.: C-OWL: Contextualizing ontologies. In: Proc. ISWC 2003, Springer 2003.
2. Ehrig M., Staab S., Sure Y.: Bootstrapping Ontology Alignment Methods with APFEL In: Proceedings of ISWC, Galway, Ireland, 2005.
3. Euzenat, J., Mochol, M., Shvaiko, P., Stuckenschmidt, H., Šváb, O., Svátek, V., van Hage, W. R., Yatskevich, M.: First Results of the Ontology Alignment Evaluation Initiative 2006. In: International Workshop on Ontology Matching collocated with the 5th International Semantic Web Conference ISWC-2006 , November 5, 2006: GA Center, Athens, Georgia, USA.
4. Gangemi, A.: Ontology Design Patterns for Semantic Web Content. In: The Semantic Web - ISWC 2005, 4th International Semantic Web Conference, Springer LNCS 3729.
5. Ghidini, C., Serafini, L.: Reconciling concepts and relations in heterogeneous ontologies. In: Proc. ESWC 2006, Budva, Montenegro, 2006.
6. Kavalec, M., Svátek, V.: A Study on Automated Relation Labelling in Ontology Learning. In: P.Buitelaar, P. Cimiano, B. Magnini (eds.), Ontology Learning and Population, IOS Press, 2005, 44-58.
7. Meilicke, C., Stuckenschmidt, H., Tamilin, A.: Improving Automatically Created Mappings Using Logical Reasoning. In: Ontology Matching 2006, Workshop at ISWC 2006.
8. 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
9. Rector, A., et al.: OWL Pizzas: Practical Experience of Teaching OWL-DL: Common Errors and Common Patterns. Proc. EKAW 2004, LNAI3257, Springer-Verlag, 63-81.
10. Serafini, L., Tamilin, A.: DRAGO: Distributed Reasoning Architecture for the Semantic Web. In: Proc. of the Second European Semantic Web Conference (ESWC'05), Springer-Verlag, 2005.
11. Shvaiko P., Euzenat J.: A Survey of Schema-based Matching Approaches. *Journal on Data Semantics*, 2005.
12. Šváb O., Svátek V., Berka P., Rak D., Tomášek P.: OntoFarm: Towards an Experimental Collection of Parallel Ontologies. Poster Session at ISWC 2005.