The Knowledge Engineering Review, Vol. 00:0, 1–24. © 2018, Cambridge University Press
DOI: 10.1017/S0000000000000000 Printed in the United Kingdom

A consensual dataset for complex ontology matching evaluation

Elodie Thiéblin[1], Michelle Cheatham[2], Cassia Trojahn[1], Ondřej Zamazal[3]

[1] *IRIT & Université de Toulouse 2 Jean Jaurès, Toulouse (France),* [2] *Wright State University, Dayton (USA),* [3] *University of Economics, Prague (Czech Republic)*
E-mail: elodie.thieblin@irit.fr, michelle.cheatham@wright.edu, cassia.trojahn@irit.fr, ondrej.zamazal@vse.cz

Abstract

Simple ontology alignments, largely studied in the literature, link one single entity of a source ontology to one single entity of a target ontology. One of the limitations of these alignments is, however, their lack of expressiveness, which can be overcome by complex alignments, which are composed of correspondences involving logical constructors or transformation functions. While most work on complex ontology matching has been dedicated to the development of complex matching approaches, there is still a lack of benchmarks on which the complex approaches can be systematically evaluated. The aim of this paper is to present the process of constructing the consensual complex Conference dataset, describing the design choices and the methodology followed for constructing it. We discuss the issues the experts were faced with during the process and discuss the lessons learned and perspectives in the field.

1 Introduction

Ontology matching is a task of generating a set of correspondences (i.e., an alignment) between the entities of different ontologies. This is the basis for a range of other tasks and applications, such as ontology evolution, query rewriting, and ontology integration. While the field has fully developed in the last decades, most works are still dedicated to generating simple correspondences between single ontology entities (e.g., $Author \equiv Writer$), mostly involving equivalence relations. However, with more and more ontologies being used for representing knowledge in many domains and being shared on the Linked Open Data (LOD) cloud, simple correspondences are not fully enough for covering the different kinds of heterogeneities (lexical, semantic, conceptual) in the ontologies to be linked. More expressiveness is achieved by complex correspondences (e.g., $IRIT\_Member \equiv Researcher \sqcap \exists belongsToLab.\{IRIT\}$), which can better express the relationships between entities of different ontologies.

Earlier works have introduced the need for complex alignments [Visser et al., 1997, Maedche et al., 2002], and different approaches for generating complex ontology alignments have been proposed in the literature afterwards. These approaches rely on diverse methods, such as correspondence patterns [Ritze et al., 2009, Ritze et al., 2010, Faria et al., 2018], knowledge-rules [Jiang et al., 2016], statistical methods [Parundekar et al., 2010, Parundekar et al., 2012, Walshe et al., 2016], competency questions for alignment [Thiéblin et al., 2018], or genetic programming [Nunes et al., 2011] and path-finding algorithms [Qin et al., 2007]. In other fields, such as relational databases, different approaches have been proposed so far [Dhamankar et al., 2004, He et al., 2004], including evaluation of alignments between hybrid structures such as ontologies and database schemes [Pinkel et al., 2017]. The reader can refer to [Thieblin et al., 2019] for a survey on complex matching in general. Here we focus on ontology matching.

While most work on complex ontology matching has been dedicated to the development of complex matching approaches, there is still a lack of benchmarks on which the complex approaches can be systematically evaluated. On the one hand, most existing proposals have been manually evaluated [Ritze et al., 2009], usually in terms of precision, or on approach-tailored datasets (e.g., one kind of correspondence only [Walshe et al., 2016]) on which recall is calculated. On the other hand, most efforts on systematic evaluation are still dedicated to matching approaches dealing with simple alignments. Although a large spectrum of matching cases has been proposed in the Ontology Alignment Evaluation Campaigns (OAEI)[1], e.g., involving synthetically generated or real world datasets with large or domain-specific ontologies, these datasets are mostly limited to alignments with simple correspondences.

Recently, the first OAEI complex track was proposed [Thiéblin et al., 2018a], opening new perspectives for the evaluation in the field. This track contained four datasets about different domains: Conference, Hydrography, GeoLink and Taxon. In particular, the complex Conference dataset results from a consensus between three raters manually generating the complex correspondences, with a special focus on the task of query rewriting. This consensual dataset extends the one presented in [Thiéblin et al., 2018b], where two (non-consensual) alignment sets for two task purposes (ontology merging and query rewriting) were proposed.

While most attention in the matching evaluation field is given to the description of datasets and the process of evaluating matching systems, the process of manual construction of reference alignments is rarely documented. However, this is a hard and time-consuming task that ideally should require multiple raters and the ability to reconcile the differences in the interpretation of ontology entities and their relations, between (usually) ill-defined natural language definitions. As stated in [Tordai et al., 2011], the manual creation of alignments is by no means an easy task and the ontology alignment community should be careful in the construction and use of reference alignments. The complexity of the problem becomes worse when dealing with complex correspondences.

The aim of this paper is to present the process of constructing the consensual complex Conference dataset and describe the design choices and methodology followed for constructing it. We explore the issues the experts were faced with during the process and discuss the lessons learned and perspectives in the field. The contributions of this paper can be summarized as follows:

- we extend the methodology from [Thiéblin et al., 2018b] for constructing complex alignments, with a focus on the query rewriting task. These guidelines can be adapted to the nature of the task or application.
- we present the consensual Complex correspondence dataset that results from the adoption of the proposed methodology by three domain experts with the same level of expertise on the domain of conference organization. While gathering annotators in the field is difficult, we argue that three annotators are reasonable for this task.
- we provide an evaluation of state-of-the-art matching systems on the consensual dataset, extending the evaluation that has been reported in the first OAEI complex track [Algergawy et al., 2018] by including additional complex matchers. We discuss their strengths and weaknesses.
- we provide lessons learned from this time-consuming and complex task, opening the room for further developments in the field.

More precisely, this paper extends the work from [Thiéblin et al., 2018b] by (i) reporting the process of construction of complex alignments by different annotators (only 1 annotator has been working on the previous datasets); (ii) focusing on alignments suitable for a query rewriting task; (iii) extending the evaluation with new matchers (AMLC), and (iv) extending the discussion on the lessons learned. We argue here that a single annotator provides only a single, non-absolute view and interpretation on the problem and several annotators are required instead. This is even more important when dealing with complex correspondences where the search space is higher. The

---

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

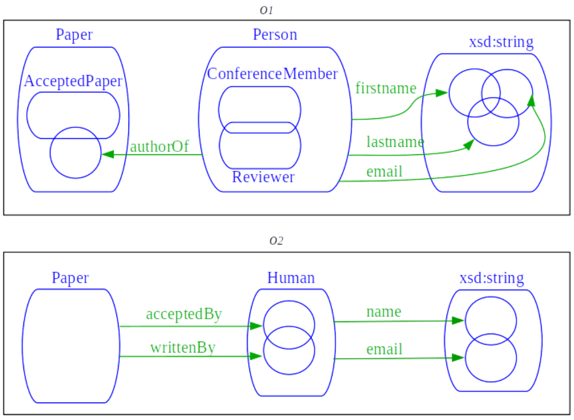


Figure 1 Fragment of two heterogeneous ontologies from the conference domain (from [Thiéblin et al., 2018b]).

improvement here is that the resulting alignment results from a consensual process, improving its quality.

The paper is organized as follows. After giving the background on ontology matching (§2) and discussing related work (§3), we describe the overall methodology to create the consensual alignments (§4). The consensus dataset is described (§5.2) and the evaluation of complex approaches presented (§6). We discuss the issues the experts faced and the mediation among annotators (§7) and then conclude with a discussion on the perspectives in the field (§8).

2 Background

Ontology matching [Euzenat and Shvaiko, 2013] is the process of generating an alignment A between two ontologies: a source ontology o_1 and a target ontology o_2. A is directional, denoted $A_{o_1 \to o_2}$. $A_{o_1 \to o_2}$ is a set of correspondences. Each correspondence is a triple $\langle e_{o_1}, e_{o_2}, r \rangle$. e_{o_1} and e_{o_2} are the members of the correspondence: they can be single ontology entities (classes, object properties, data properties, instances) of respectively o_1 and o_2 or constructions of these entities using constructors or transformation functions. r is a relation, e.g., equivalence ($\equiv$), subsumption ($\sqsubseteq$, $\sqsupseteq$), or disjointedness ($\perp$) between e_{o_1} and e_{o_2}.

The ontologies used in the following examples are illustrated in Figure 1. In this paper, the complex correspondences are described using the $\mathcal{DL}$ syntax and the ontologies are graphically represented using the diagrammatic logic formalism defined in [Stapleton et al., 2014].

We consider two types of correspondences, depending on the type of their members:

- if the correspondence is **simple**, both e_{o_1} and e_{o_2} are atomic entities: one single entity is matched with another single entity, e.g., o_1*:Person* $\equiv$ o_2*:Human* is a simple correspondence.
- if the correspondence is **complex**, at least one of e_{o_1} or e_{o_2} involves a constructor or a transformation function. For example,
  1. o_1*:authorOf* $\equiv$ o_2*:writtenBy*$^-$ is a complex correspondence with the *inverseOf* constructor.
  2. o_1*:AcceptedPaper* $\equiv \exists$ o_2*:acceptedBy*$.\top$ is a complex correspondence with the *existential* constructor.
  3. o_2*:name* is the concatenation of the o_1*:firstname* and o_1*:lastname* is a complex correspondence with a transformation function[2].

---

[2] Transformation functions can not be formalized into $\mathcal{DL}$.

A complex alignment contains at least one complex correspondence. We will refer to approaches that generate simple alignments as *simple matchers* and to approaches that generate complex alignments as *complex matchers*.

3 Related work

This section discusses the main related work on complex ontology alignment evaluation and generation of reference alignments. Although some approaches rely on instances to discover alignments at the schema level as well as adopt different kinds of reference (alignments or queries), the focus here is on ontology matching rather than on entity matching.

3.1 Complex ontology alignment evaluation

Alignments generated by (simple) matchers have been evaluated in different ways [Do et al., 2002]. One classical way consists of comparing generated alignments to reference ones (gold standard). However, constructing such references is a time-consuming task that requires experts in the domain. In the absence of such resources or when dealing with large datasets, alternatives include manual labelling on sample alignments [Van Hage et al., 2007], computing the minimal set of correspondences (which can be used for computing all the other ones) for reducing the effort on manual validation [Maltese et al., 2010], or measuring the quality of alignments in terms of coherence measurements [Meilicke and Stuckenschmidt, 2008] or conservativity principle violation [Solimando et al., 2017]. Alternatively, an alignment can be assessed regarding its suitability for a specific task or application [Isaac et al., 2008, Hollink et al., 2008, Solimando et al., 2014]. Other approaches consider the generation of natural language questions to support end-users in the validation task [Abacha and Zweigenbaum, 2014] or validating correspondences on graph-based algorithms in a semi-automatic way [Serpeloni et al., 2011]. While those approaches have been primarily applied to simple alignments, complex alignment evaluation has been addressed to a lesser extent. To the best of our knowledge there is no current approach fully automating the evaluation of complex alignments.

The evaluation of most existing complex approaches has been done by manually calculating the precision of the alignments generated by the systems [Ritze et al., 2009, Ritze et al., 2010, Parundekar et al., 2012, Walshe et al., 2016]. With respect to the few complex alignment sets, most of them have been created to evaluate specific complex matching approaches, aiming at calculating recall. The approach of [Parundekar et al., 2012] estimated recall based on the recurring pattern between DBpedia and Geonames:$\exists$ *dbpedia:country.*$\{$`theCountryInstance`$\} \equiv$ $\exists$ *geonames:countryCode.*$\{$`theCountryCode`$\}$ where `theCountryInstance` is a country instance of DBpedia such as *dbpedia:Spain* and `theCountryCode` is a country code such as *"ES"*. They estimated the number of occurrences of this pattern between these ontologies and calculated the recall based on this estimation. In [Qin et al., 2007] a set of reference correspondences between two ontologies has been manually created, involving nine correspondences from which only two could not be expressed with simple ones. In [Walshe et al., 2016] the authors proposed an algorithm to create an evaluation data set that is composed of a synthetic ontology containing 50 classes with known *Class-by-attribute-value* correspondence pattern with DBpedia and 50 classes with no known correspondences with DBpedia. Both ontologies are populated with the same instances. More recently, a "compound" alignment set has been proposed between biology ontologies in [Oliveira and Pesquita, 2018]. These alignments involve entities from more than two ontologies. For example, $o_1{:}A \equiv o_2{:}B \sqcap o_3{:}C$ is a compound correspondence. These correspondences can be considered complex since one member contains a constructor, but they are out of the scope of our study. The closest approach to ours is from [Jiang et al., 2016], who extended the Conference dataset with complex alignments to evaluate their knowledge-rule based alignment approach. However, the methodology used for the construction of the dataset is not specified, and the dataset is not available online.

Approaches using complex correspondences for a given purpose (query rewriting for example), also propose alignment sets created for their needs, even though they have not been used for matcher evaluation. For instance, the authors of [Makris et al., 2012] present a set of complex correspondences used for query rewriting[3]. However, they are not in a reusable format and only concern a pair of ontologies. In [Thiéblin et al., 2017], complex correspondences between agronomic ontologies were manually created for query rewriting on the LOD cloud.

In the context of systematic evaluations, four datasets have been recently proposed in the first OAEI complex track [Thiéblin et al., 2018a]. These datasets cover different domains (conference, hydrology, geoscience and agronomy) and evaluation strategies. The Conference dataset refers to the consensus dataset described in Section 5.2. Precision and recall measures are manually calculated over the complex equivalence correspondences. The Hydrography dataset is composed of four source ontologies and a target ontology, and the evaluation is based on three subtasks: given an entity from the source ontology, identify all related entities in the source and target ontologies; given an entity in the source ontology and the set of related entities, identify the logical relation that holds between them; identify the full complex correspondences. The GeoLink dataset, as with the Conference dataset, was developed in consultation with domain experts from several geoscience research institutions. Evaluation was conducted as for the Hydrography dataset. Finally, the taxon dataset aims at evaluating alignments involving plant taxonomy. The evaluation is two-fold: first, the precision of the output alignment is manually assessed; then, a set of source queries are rewritten using the output alignment. Each rewritten target query is then manually classified as correct or incorrect. A source query is considered successfully rewritten if at least one of the target queries is equivalent to it (i.e., it is able to retrieve the same set of instances). The proportion of source queries successfully rewritten is then calculated. The evaluation over this dataset was open to all matching systems (simple or complex), but some queries can not be rewritten without complex correspondences.

Finally, in the domain of schema matching (database or XML schema), dedicated complex alignment datasets have been constructed to evaluate the approaches dealing with these schemata. In general, these datasets contain mostly transformation functions. For instance, the Illinois semantic integration archive [Doan, 2005] is a dataset of complex correspondences on value transformations (e.g., string concatenation) in the inventory and real estate domain. This dataset only contains correspondences between schemata with transformation functions. For the purpose of evaluating matching hybrid structures, the RODI Benchmark [Pinkel et al., 2017] proposes an evaluation over a given scenario involving R2RML correspondences between a database schema and an ontology. The benchmark relies on ontologies from the OAEI Conference dataset, Geodata ontology, and the Oil and Gas ontology. The schemata are either derived from the ontologies themselves or curated on the Web. RODI deals with R2RML alignment and uses reference SPARQL and SQL queries to assess the quality of the alignment. Recently, an approach for automatic generation of R2RML mappings has been evaluated on this benchmark [Mathur et al., 2018].

3.2 Consensual reference alignments

The creation of reference alignments is crucial in ontology matching evaluation. While different datasets have been constructed from manual analysis, involving a different number of experts and resulting in different levels of agreement, the focus has mostly been on describing the resulting dataset rather than on the details of the manual process. Guidelines for constructing reference alignments are in fact scarce in the field, though there are more general discussions on the qualities of a good benchmark in other research fields [Sim et al., 2003, Dekhtyar and Hayes, 2006].

Different strategies have been followed, including starting the alignment generation from scratch, relying on a set of initial alignments for gathering additional ones, and creating a

[3] `http://www.music.tuc.gr/projects/sw/sparql-rw/`

reference from validating and selecting a set of correspondences from automatically generated correspondences from a number of matching systems. In the first category, the creation of the first reference alignment of the Conference dataset dates back to 2008, when the track organizers created a reference alignment for all possible pairs of five of the conference ontologies. The reference alignments were based on the majority opinion of three evaluators and were discussed during a consensus workshop. This dataset has evolved over the years (as described in [Zamazal and Svátek, 2017]), with the feedback from the OAEI participants and has been revised in [Cheatham and Hitzler, 2014]. They re-examined the dataset with a focus on the degree of agreement between the reference alignments and the opinion of experts. A general method for crowdsourcing the development of more benchmarks of this type using Amazon's Mechanical Turk has been also introduced and shown to be scalable and to agree well with expert opinion.

With the aim of studying the way different raters evaluate correspondences, in [Tordai et al., 2011] experiments in manual evaluation have been carried out using a set of correspondences generated by different matchers between vocabularies of different types. Five raters evaluated alignments and talked through their decisions using the think aloud method. Their analysis showed which variables can be controlled to affect the level of agreement, including the correspondence relations, the evaluation guidelines and the background of the raters. That work refers as well to the different levels of agreements between annotators reported in the literature. While a perfect agreement between raters is reported in the Very Large Crosslingual Resources (VLCR) dataset in [Euzenat et al., 2009], [Halpin et al., 2010] reported a quite different observation when establishing *owl:sameAs* relationships in the LOD. These aspects have also been discussed in [Stevens et al., 2018] for the task of manually integrating top-level and domain ontologies.

Close to ours, [Zhou et al., 2018] propose a dataset where all correspondences were established as a collaborative effort between the data repository providers, the domain experts, and the ontology engineers involved in the modeling and deployment process of the GeoLink project. These correspondences include expressive (m:n) correspondences. However, the methodology followed to create this alignment has not been documented in their paper. The methodology we followed in this paper is detailed in the next section.

4 Methodology

This section describes the overall methodology we followed to create the consensual complex alignment dataset. As stated before, we adapted the methodology proposed in [Thiéblin et al., 2018b]. The methodology focuses on finding as many complex correspondences as possible with an equivalence relation according to the task purpose of the alignment. This methodology targets task-oriented creation of alignments, in particular, ontology merging and query rewriting.

Here, our choice was to focus on creating alignments for query rewriting. This design choice is motivated by the fact that query rewriting is a common task requiring alignments that does not restrict complex alignments to $\mathcal{SROIQ}$ expressiveness, as is the case for alignments applied to ontology merging. That task requires the resulting ontology to be coherent; in other words, reasoning on the merged ontology must be decidable. Therefore, the alignment should follow the $\mathcal{SROIQ}$ expressiveness and should not bring any incoherence. For query rewriting, the expressiveness of the correspondences is not limited. Transformation functions can be used as well as "complex roles" (which are limited in $\mathcal{SROIQ}$). Therefore, the coherence of an alignment intended for query rewriting can generally not be verified because a reasoning task is not decidable given its expressiveness.

Having the target task, the second step was to agree on the methodology to follow. For that, all three experts discussed the methodology and agreed on how to proceed for constructing complex correspondences targeting the task of query rewriting. Unclear points of the methodology and different interpretations of some steps have been discussed.

The overall methodology is articulated in the following steps:

1. Agree on the simple equivalence correspondences between o_1 and o_2 to rely on.
2. Individually create the complex correspondences based on the simple correspondences so that the complex correspondences fit the purpose of the alignment; and express the correspondences in First Order Logic (FOL).
3. Collaboratively validate the set of found complex correspondences.

In the following, we detail the main steps of the methodology.

4.1 Step 1: agree on equivalence correspondences

We argue that complex correspondences come as a complement to simple correspondences. The first step of the methodology was to decide the set of simple input alignments to use as a basis for anchoring the discovery of complex correspondences. Here, the available simple reference alignment (ra1) from the Conference dataset has been used. There are also other variants of ra1 in the Conference track, but since their differences are relatively small and all are derived from the (always open) ra1, this one was used as input for starting discussion towards agreeing on equivalence correspondences. The three annotators analyzed the simple alignments and some conflicts leading to incoherence were identified and discussed. Furthermore, missing correspondences were added to the original set in order to better guide the discovery of the complex correspondences. The output of this step was a consensual set of simple equivalent correspondences. This results in:

- 5 equivalence correspondences removed
  (e.g. *cmt:ConferenceMember* ≡ *ekaw:Conference_Participant*)
- 4 equivalence correspondences added
  (e.g. *cmt:ProgramCommitteeMember* ≡ *ekaw:PC_Member*)
- 7 modifications in the correspondence relation, changing equivalences to subsumptions)
  (e.g. *conference:Information_for_participants* $\sqsupseteq$ *ekaw:Programme_Brochure*, instead of an equivalence)

4.2 Step 2: individually create complex correspondences

From the simple consensus alignment established in the previous step, an agreement between the annotators was reached with respect to the following points:

- Find correspondences in both directions (1:n) and (m:1). By focusing on (1:n) and (m:1) correspondences, the size of the matching space (all possible correspondences) is reduced. The correspondences found by the annotators will then be easier to compare, as only their relation and target member will differ.
- Focus on equivalence correspondences because equivalence can be considered as the most informative relation.
- If no equivalences are found, (simple or complex) subsumptions are then considered. For subsumption correspondences, precision was favoured over recall. We chose to focus on more accurate correspondences rather than covering ones. For example, when matching *conference:Conference_contribution* with the *cmt* ontology, the correspondence *conference:Conference_contribution* $\sqsupseteq$ *cmt:Paper* will be preferred over *conference:Conference_contribution* $\sqsubseteq$ *cmt:Document.*

Having this in mind, each annotator manually generated a set of complex correspondences, following the steps, with o_1 the source ontology, and o_2 the target one.

1. For each entity e_1 of o_1 not in a simple equivalence correspondence, find a semantically equivalent construction from o_2 entities.

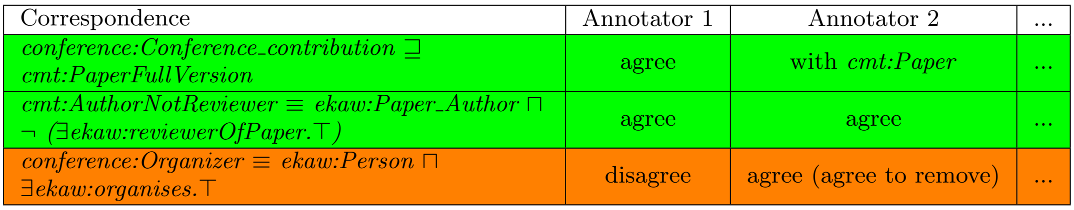
  - If no equivalence can be found, look for the closest entity or construction from o_2 subsumed by e_1.

2. Repeat the previous step for each entity of o_2 (constructions from o_1 entities).

All the correspondences have then been expressed in FOL. This choice is motivated by the fact it is a common representation language which has a good balance of expressiveness and readability.

4.3 *Step 3: collaboratively validate the complex correspondences*

In the third step of the process, the sets of correspondences generated by the three annotators have been merged together. Then, each annotator analyzed the merged set of correspondences, in an open evaluation process (i.e., knowing the name of the annotator of each correspondence), provided her/his comments and feedback on each correspondence, and classified them into *agree* or *disagree* categories. Annotators provided a justification in *disagree* cases. A fourth annotator also participated in this analysis. Figure 2 presents a fragment of the online shared spreadsheet used to create and comment the correspondences.



Correspondence	Annotator 1	Annotator 2	...
$conference{:}Conference\_contribution \sqsupseteq cmt{:}PaperFullVersion$	agree	with *cmt:Paper*	...
$cmt{:}AuthorNotReviewer \equiv ekaw{:}Paper\_Author \sqcap \neg\ (\exists ekaw{:}reviewerOfPaper.\top)$	agree	agree	...
$conference{:}Organizer \equiv ekaw{:}Person \sqcap \exists ekaw{:}organises.\top$	disagree	agree (agree to remove)	...

Figure 2 Interface for correspondence annotation. Agree/disagree and a comment to argument if necessary

Once each annotator analyzed the whole set of correspondences, the decisions were discussed and the differences in interpretations were reconciled. This was an iterative process until full agreement (or disagreement) was reached for each correspondence. We ended up with a set of *agreed* correspondences, as further detailed in Section 5.2.

As expected, some correspondences were written differently by different annotators e.g., $conference{:}Conference\_part \equiv \exists ekaw{:}partOf.ekaw{:}Conference$ was also written as $conference{:}Conference\_part \equiv \exists ekaw{:}hasPart^{-}.ekaw{:}Conference$. These two correspondences are semantically equivalent as $ekaw{:}hasPart \equiv ekaw{:}partOf^{-}$. In the final consensus, we chose the expression with the smallest number of constructors. When different constructions were found equivalent by the annotators but were not explicitly semantically equivalent with respect to the ontology axioms, the target constructions were put together in a disjunction (union). For example, $ekaw{:}Accepted\_Paper \equiv \exists\ cmt{:}hasDecision.cmt{:}Acceptance$ and $ekaw{:}Accepted\_Paper \equiv \exists cmt{:}acceptedBy.\top$ were both *agreed* on by the annotators but nothing in *cmt* states that $\exists cmt{:}hasDecision.cmt{:}Acceptance \equiv \exists cmt{:}acceptedBy.\top$. Therefore, the following correspondence was chosen in the consensus alignment: $ekaw{:}Accepted\_Paper \equiv (\exists cmt{:}hasDecision.cmt{:}Acceptance) \sqcup (\exists\ cmt{:}acceptedBy.\top)$. More details are provided in the next section.

5 Consensus complex alignment set

Before describing the resulting consensus alignment set, we introduce the Conference dataset from which our dataset have been built.

5.1 Conference dataset

The Conference dataset[4] was proposed in [Šváb et al., 2005]. This dataset has been used to evaluate nearly all ontology matching systems developed since then [Cheatham and Hitzler, 2014] and it is quite a challenging dataset in the field [Zamazal and Svátek, 2017]. This dataset is composed of 16 ontologies on the conference organization domain and simple reference alignments between 7 of these ontologies. These ontologies were developed individually. We chose three ontologies among the ones in the reference simple alignment for their different number of classes (Table 1): *cmt*, *conference* (Sofsem), and *ekaw*. Here, we consider the set of simple reference alignments that results from the modifications made in the first step of the methodology.

Table 1 Number of entities by type of each ontology.

	cmt	conference	ekaw
Classes	30	60	74
Object properties	49	46	33
Data properties	10	18	0

5.2 Consensual complex dataset

As stated above, the methodology was applied by three experts in the domain to all six pairs involving the three ontologies. During the creation of the complex correspondences, some annotators did not exactly follow the methodology. The correspondences that they created were all annotated by the others annotators even if not compliant with the methodology. This was, in particular, related to the lack of direction in our current methodology regarding creation of (m:n) correspondences. This resulted in three alignments:

All Alignment containing all of the correspondences created by the annotators.

Methodo Alignment containing all of the correspondences created by the annotators and compliant with the methodology.

Logic methodo Alignment containing the correspondences with logic expressions as members created by the annotators and compliant with the methodology (all correspondences from Methodo except the value transformation function correspondences).

The observed agreements for the three datasets are shown in Table 2. Note that this agreement has been calculated over the consensus dataset. Overall, we observe a higher agreement, with a slight lower agreement for the *Methodo* and *Logic methodo* involving the pairs cmt-ekaw.

Table 2 Observed agreement between raters, for each pair of ontologies, for each version of the alignment (All, Methodo, Logic methodo).

	All (1)	Methodo (2)	Logic methodo (3)
cmt-conference	93%	87%	86%
conference-cmt	89%	91%	90%
cmt-ekaw	73%	67%	67%
ekaw-cmt	78%	100%	100%
conference-ekaw	79%	77%	77%
ekaw-conference	90%	91%	91%
Average	83%	85%	85%

---

[4] `http://oaei.ontologymatching.org/2016/conference/index.html`
`http://owl.vse.cz/ontofarm/`

The patterns were used *a posteriori* for analyzing the alignments, not as a basis for the correspondence creation. An extensive list of the patterns can be found in [Scharffe, 2009]. The meaning of the abbreviations used in the following tables is: *CAT*: $A \equiv \exists b.C$, *CAE*: $A \equiv \exists b.\top$, *CIAE*: $A \equiv \exists b^{-}.\top$, *CIAT*: $A \equiv \exists b^{-}.C$, *n*: negation, *dom*: domain restriction, *range*: range restriction, *dom/range*: domain and range restriction, *transfo*: transformation function on data properties, *c*: class, *rel*: object property, *prop*: data property, *chain*: a chain of properties (object properties and/or data properties), *inv*: inverse of an object property, *composite* or *compo*: different patterns in the same correspondence. The domain restriction and range restriction patterns are correspondence patterns from [Scharffe, 2009] and not OWL axiom primitives.

Table 3 presents examples of correspondences from the alignment sets and their type.

Table 3 Example of correspondences and their type (correspondence pattern).

Source entity	rel.	Target construction	type
cmt:ExternalReviewer	$\equiv$	$\exists$ *conference:invited_by*.$\top$	CAE
conference:Submitted_contribution	$\equiv$	$\exists$*cmt:submitPaper*$^{-}$.$\top$	CIAE
cmt:ProgramCommitteeMember	$\equiv$	$\exists$*conference:was_a_member_of.* *conference:Program_committee*	CAT
conference:Conference_part	$\equiv$	$\exists$*ekaw:hasPart*$^{-}$. *ekaw:Conference*	CIAT
ekaw:ScientificEvent	$\equiv$	*conference:Conference_part* $\sqcup$ *conference:Conference*	union(c)
ekaw:SubmittedPaper	$\sqsupseteq$	*conference:Submitted_contribution* $\sqcap$ *conference:Paper*	inters(c)
cmt:hasProgramCommitteeMember	$\equiv$	*conference:has_members.* *conference:Program_committee.*$\top$	dom(rel)
ekaw:reviewerOfPaper	$\equiv$	*conference:contributes* $\circ$ *conference:reviews*	chain(rel)
cmt:writeReview	$\equiv$	*ekaw:reviewWrittenBy*$^{-}$	inv(rel)

Table 4 shows the number of agreed correspondences per type in the *All* and *Methodo* agreed alignments. The *All* alignments contain globally correspondences with more patterns than the *Methodo* ones. There are more simple subsumptions in the *All* alignments than in the *Methodo* ones as most of them were filtered.

Table 4 All and methodology (met) number of correspondences per type of correspondence.

	cmt-conference		conference-cmt		cmt-ekaw		ekaw-cmt		conference-ekaw		ekaw-conference	
	all	met	all	met	all	met	all	met	all	met	all	met
simple eq	14	14	14	14	13	13	13	13	15	15	15	15
simple sub	9		11	9	6	1	8	6	20	10	18	9
inters(c)											4	4
inters(c,n(c))	2	1										
inters(compo(c))	1	1			2	2						
union(c)	1	1			1		1	1	4	3	4	3
union(compo(c))				2	3	2	2	2	1	1		
CIAT									3			
CAT	2	2	2				2			1	9	9
CIAE							1	1				
CAE	1	1	4	2			4	1		1		
dom(rel)	3	1	1	1	2	1			2	2	1	
range(rel)	3	1	1	1	2	2			2	2	3	1
dom/range(rel)	2	2		2	2				5	3		
chain(rel)	2	2									2	2
union(rel)			2			1	2	2	1		4	4
dom(inv(rel))									2			
inv(rel)					1	1	1	1				
dom(prop)			1	1								
transfo	1	1	2	2								

Table 5 Differences between the methodology-compliant consensus alignment and the query-rewriting alignment from [Thiéblin et al., 2018b]. It shows the number of correspondences which are identical, have been added or deleted, or whose relation (r) was changed from the query-rewriting alignment to obtain the consensus alignment.

	complex				simple			
	identical	added	deleted	r changed	identical	added	deleted	r changed
cmt-conference	11		4	2	13	1	1	1
conference-cmt	6			4	18	2	3	
cmt-ekaw	8	1			11	2		1
ekaw-cmt	6	1	3		11	3	1	
conference-ekaw	10	2	6	5	20		5	2
ekaw-conference	12	10	5	1	21		3	

Table 5 shows the differences between the methodology-compliant consensual alignment and the query-rewriting one from [Thiéblin et al., 2018b] (following the same methodology). One can notice that for some ontology pairs such as *cmt-ekaw*, few changes were made, whereas for others, such as *ekaw-conference*, we observe a higher number of changes. By comparing the alignments, for some cases, a change in the simple correspondences implies changes for the complex correspondences. This was the case for the *conference-ekaw* and *conference-cmt* correspondences, in which a simple equivalence correspondence (e.g., *cmt:Paper* ≡ *conference:Written_contribution*) was found in the consensus alignment to be a subsumption ($\sqsubseteq$), leading to complex correspondences with different relations from the query-rewriting alignment from [Thiéblin et al., 2018b]. Overall, the simple correspondences are more easily consensual than the complex correspondences. 79% of the simple correspondences from the consensus and the query-rewriting one [Thiéblin et al., 2018b] are identical whereas only 55% of the complex ones are. We argue here that the 45% of the non-identical correspondences refers to an extension of the original dataset, and in that sense we can argue that it is an improvement in terms of coverage of the space of possible correspondences. The quality is rather guaranteed by the fact that it has been manually created under a consensual process.

6 Evaluation of complex matchers

In order to perform an evaluation using the introduced consensual dataset we selected four approaches:

- Ritze2009: the pattern-based approach presented in [Ritze et al., 2009].
- Ritze2010: the lexical-based approach presented in [Ritze et al., 2010].
- Jiang2016: the approach, KAOM (Knowledge Aware Ontology Matching), based on Markov logic networks as described in [Jiang et al., 2016].
- AMLC2018: the approach used within the OAEI 2018 that is a variation of the AML matcher for complex matching [Faria et al., 2018].

The choice for these tools is motivated by the fact that: 1) they are the publicly available systems that could be run without errors; 2) they do not rely on instances, as the dataset is not equipped with instances.

The complex correspondences output by the matchers were manually compared to the methodology-compliant consensual alignment. For this evaluation we considered only equivalence correspondences. Further, the confidence of the correspondences was not taken into account. The input matchers used correspondences from the simple reference alignment (ra1); therefore we only evaluate the complex correspondences.

Table 6 shows precision, recall and F-measures per matcher and with regard to each ontology pair as well as on average. The best performance (0.42 of F-measure) was achieved by AMLC2018, which participated in OAEI 2018. Although other selected matchers only achieved F-measures around 0.10, they still managed to generate interesting true positives (TP) as well

Table 6 Precision (P), Recall (R) and F-measure for four selected matchers.

Tool	Ontology pair	P	F-measure	R
Ritze2009	conference-ekaw	0.00	0.00	0.00
	cmt-conference	0.00	0.00	0.00
	cmt-ekaw	0.50	0.30	0.20
	mean	0.17	0.10	0.07
Ritze2010	conference-ekaw	0.00	0.00	0.00
	cmt-conference	0.00	0.00	0.00
	cmt-ekaw	1.00	0.33	0.20
	mean	0.33	0.11	0.07
Jiang2016	conference-ekaw	0.06	0.05	0.05
	cmt-conference	0.00	0.00	0.00
	cmt-ekaw	0.14	0.12	0.10
	mean	0.07	0.06	0.05
AMLC2018	conference-ekaw	0.36	0.26	0.20
	cmt-conference	0.40	0.28	0.22
	cmt-ekaw	0.86	0.71	0.60
	mean	0.54	0.42	0.34

as interesting false positives (FP). In comparison with the task-oriented evaluation performed in [Thiéblin et al., 2018b], the matching system (KAOM) from the Jiang2016 approach found fewer TPs than in the case of the current evaluation (four against two TPs). In the case of the approaches Ritze2009 and Ritze2010, the number of TPs remains the same (two TPs).

We inspected the generated alignments in more detail per each matcher. The matcher from Ritze2009 generated several interesting incorrect correspondences, mostly for difficult concepts, e.g., *cmt:Meta-Review* $\equiv \exists$*ekaw:hasReview*$^{-}$*.ekaw:Review*. The complex correspondence states that a Meta-Review is a Review which reviews of something. This definition rather fits to any review. We should note that the concept of Meta-Review is underspecified in the cmt ontology (Meta-Review is merely defined as a subclass of Review) and thus it is difficult to grasp this concept based on the ontology. Another example of a difficult concept, which the matcher from Ritze2009 tried to match, is *AuthorNotReviewer* from the cmt ontology. In this case the difficulty comes from a negation present in its local name. While the attempt to match was promising it does not properly cope with the open world assumption principle: *cmt:AuthorNotReviewer* $\equiv \exists$*conference:contributes. conference:Reviewed_contribution*. The matcher from Jiang2016 generated the highest number of correspondences (39, see in Table 7). Many false positives happened due to a domain or range mismatch with regard to a property definition of domain and/or range in the source ontology and definition of domain and range stated in the complex correspondence, e.g., *cmt:writtenBy* $\equiv$ *ekaw:writtenBy.ekaw:Paper.ekaw:Paper_Author* where domain (range) of *writtenBy* property in *cmt* is defined as Review (Reviewer resp.). In many cases the matcher from Jiang2016 did not follow the right direction of the property used for the restriction, e.g., *conference:Review* $\equiv \exists$ *ekaw:hasReview.ekaw:Positive_Review* where domain of *hasReview* is Paper in the *ekaw* ontology.

AMLC2018 several times found an equivalence correspondence while the methodology-compliant consensual alignment had the correspondence with the subsumption relation. We think that it is particularly difficult to properly distinguish between equivalence and subsumption in some situations, e.g., *cmt:Rejected_contribution* $\equiv \exists$*conference:hasDecision. conference:Rejection*. Further, this matcher often generated complex correspondences where the property was used in the wrong direction, e.g., *conference:Presentation* $\equiv \exists$ *ekaw:presentationOfPaper*$^{-}.\top$.

Table 7 Number of different types of complex correspondences (correspondence patterns). Types are explained in Section 5.2. Other types include inv(rel), union(c) and correspondences with the universal quantifier.

Tool	Ontology pair	#CAE	#CAT	#dom/range	#chain	#Other	all
Ritze2009	conference-ekaw	0	3	0	0	0	3
	cmt-conference	0	2	0	0	0	2
	cmt-ekaw	0	4	0	0	0	4
	sum	0	9	0	0	0	9
Ritze2010	conference-ekaw	0	0	0	0	0	0
	cmt-conference	0	0	0	0	0	0
	cmt-ekaw	0	2	0	0	0	2
	sum	0	2	0	0	0	2
Jiang2016	conference-ekaw	0	4	8	4	1	17
	cmt-conference	0	1	9	2	3	15
	cmt-ekaw	0	3	2	0	2	7
	sum	0	8	19	6	6	39
AMLC2018	conference-ekaw	2	9	0	0	0	11
	cmt-conference	1	4	0	0	0	5
	cmt-ekaw	4	3	0	0	0	7
	sum	7	16	0	0	0	23

While it is common to assign confidence scores to simple correspondences, out of four selected matchers only AMLC2018 assigned a confidence to the output. As we expect more participants in the complex track of OAEI in the future, we also anticipate that they will assign confidence scores to complex correspondences.

The matchers differ not only in their performance with respect to the methodology-compliant consensual alignment but also with respect to the types of correspondences they found. Numbers of different types of correspondences are stated in Table 7. The matchers from Ritze2009 and Ritze2010 only consider the CAT type of complex correspondences (9 and 2 resp.). The AMLC2018 generated not only the CAT type of complex correspondences (16) but also the CAE type of complex correspondences (7). The most diverse types of complex correspondences were found by the matcher from Jiang2016. It further generated the dom/range type of complex correspondences (19) where restriction was applied on both domain and range and the chain type of complex correspondences (6). It also outputs four correspondences of inv(rel), one union(c) and one complex correspondence with a universal quantifier.

7 Lessons learned

In the following, the lessons learned from the effort of creating a consensus complex dataset and using it for evaluation of ontology matching systems are discussed together with future directions in the field.

7.1 Manual creation of alignments

In general, manually creating alignments is far from an easy task. This difficulty however increases when dealing with complex alignments, which are inherently more expressive. One of the main challenges is to fully understand the nuances between similar entities' definitions, which can lead to different interpretations in terms of the semantics of entities and the types of relations, e.g., equivalences or subsumptions. For example, the concept *reviewer of a paper* can either mean "*a person who reviews a paper*" or "*a person who is assigned to a paper*" which is slightly different. The propagation of these interpretations led to different interpretations of the whole set. This experience corroborates what has been stated in [Tordai et al., 2011] on the well-known vagueness of the boundary between polysemy and homonymy observed from studies in lexical semantics, where the classification of different types of polysemy is still a matter of debate among linguists. Humans rarely have problems disambiguating the meaning of words in a discourse context. However, in an ontology alignment task this context is usually much more limited than discourse. This is even worse when dealing with poor annotations in the ontologies to be aligned, as is the case in the Conference dataset, where not all entities have rich associated annotations. Having rich terminological layers in the ontologies could help as well as having some sample instance data.

Another issue in establishing complex alignments is related to the strong need for collaboration in order to reach a consensus. Reaching a consensus is done by measuring the level of agreement between annotators and keeping the correspondences with a high level of agreement. However, as observed in several works [Tordai et al., 2011, Halpin et al., 2010, Stevens et al., 2018], the level of agreement may diverge greatly. It is helpful to keep track of correspondences that have been previously considered and rejected (and the reasons why), to avoid repeated discussions on the same issues. Likewise, if a correspondence is accepted by the annotators as valid, the reasons for this should be captured. Part of the difficulty in achieving consensus is due to the various possible usages of an ontology alignment. A correspondence that is appropriate for an alignment intended for a query federation application may not be appropriate for an application that requires logical reasoning over the merged ontology. Keeping track of different application versions of an alignment between two ontology pairs is therefore important. We note that these requirements are not unique to establishing complex reference alignments – they also apply to simple alignments, but the complexity involved in the complex case makes these issues even more pressing.

Due to the difficulties discussed above and several others, the process of generating the complex alignments was time-intensive, and the nature of the work was considered somewhat tedious by the annotators. Many issues could potentially be mitigated in the future through improved tool support (though this is not only specific to the construction of complex alignments). Here, the generated alignments have been stored in spreadsheets. As reported in [Meilicke et al., 2009], even using an ad-hoc web-based tool can help support construction of alignments. More advanced tools could greatly facilitate the process. In particular, the following functionality may be useful in future efforts of this kind:

1. Automated selection (and ranking) of entities in one ontology related to a selected entity in the other; this might be provided based on an existing simple reference alignment and/or the output of a simple alignment system
2. Indication of when a newly suggested correspondence is already entailed by the existing set of correspondences and/or the ontologies themselves
3. Dependency tracking to monitor when a correspondence depends on a particular semantic interpretation of an ontology entity and/or another correspondence[5]
4. A collaborative mode supporting voting, comments, and version tracking
5. Ability to output a complex alignment in various popular representation formats (e.g. OWL, EDOAL, DL syntax)
6. Synthetic generation of instance data sufficient to evaluate a proposed alignment over the ontology pair from a query rewriting perspective

---

[5] For example, in the work here, the simple correspondence between *cmt:Reviewer* and *conference:Reviewer* was changed from equivalence to subsumption, which then caused a new complex correspondence, *conference:Reviewer* $\sqsupseteq$ *cmt:Reviewer* $\sqcup$ *cmt:ExternalReviewer* to be added to the alignment. If the original decision were reversed, the complex correspondence would also need to be removed.

Last but not least, the process of manual alignment creation could benefit from tools supporting traceability. In that sense, the M-Gov framework described in [Singh et al., 2017] could help in describing the metadata related to the users involved and their discussions during the generation of alignments.

7.2 Methodology

The methodology followed in this work to create the complex alignments required several refinements over the course of the project, due to different initial interpretations by the annotators involved. Key issues included the need to define the use case for the alignments under development (e.g., query rewriting versus ontology merging), the performance metric to optimize (e.g., precision rather than recall or F-measure), and the definition of "subsumption" with respect to data properties (e.g., is the meronomic relation of family name to full name considered a subsumption?). The methodology presented in Section 4 is to some degree specific to the decisions made on these issues. Future work on complex alignment generation that makes alternative choices on these matters will likely require variations on the approach presented here.

Even when future efforts to form consensus on complex alignments have similar goals to those pursued here, they might benefit from extensions to the methodology used here. In particular, the current methodology does not allow for the suggestion of (m:n) correspondences between two ontologies (i.e., correspondences in which both e_{o_1} and e_{o_2} both involve constructors or transformation functions). Existing work on aligning real-world ontologies suggests that such correspondences, while not close to the majority, occur naturally in many cases. Guidance could also be given on whether or not to include correspondences between the ontologies that are logically or intuitively true but may "add new knowledge" (i.e., that exist in an area at the border between ontology alignment and ontology engineering/creation). Furthermore, while we have made the assumption that the task impacts an alignment's expressiveness and therefore made the choice to target a query-rewriting application rather than, say, ontology merging, our work did not consider the specific application to which the alignment would be applied (e.g., conference paper management, conference attendees management). Taking into account the specific application purpose for an alignment may impact not only the expressiveness of the alignment but also its content.

Another potential area in which the current methodology could be extended is guidance on how to specify the confidence value of a correspondence. In fact, many existing simple alignment benchmarks also do not have meaningful confidence values associated with correspondences (they are all 1.0); however, this has been shown not to accurately reflect the degree of consensus on the correspondences [Cheatham and Hitzler, 2014]. Meaningful confidence values can be useful in evaluating ontology matchers (e.g., by penalizing a system more for missing an obviously correct correspondence than a controversial one). They can also be useful for system designers to determine when their system is producing meaningful but not optimal correspondences. There are several possible semantics for confidence values, including the degree of consensus or the degree of "correctness" of a correspondence. Developing a methodology for generating meaningful and useful confidence values for complex correspondences remains an important area of future work.

7.3 Consensus dataset

The consensus complex alignments presented here have the benefit of involving real-world ontologies that are fairly expressive; however, there are also several limitations to this dataset. In particular, the ontologies involved are relatively small and they are not populated, which impacts the evaluation of complex matchers that rely on large numbers of instances. Furthermore, because the ontologies considered here all model the same particular domain, the dataset may bias the evaluation of complex matchers. It is clear that a wider variety of datasets involving complex alignments should be made available to facilitate the development and evaluation of complex ontology alignment systems. These alignments should ideally involve ontologies of various sizes that cover a range of different domains. The work here leads us to note several potential difficulties in establishing such alignments, however. In particular, many ontology pairs contain valid correspondences that are not representable in either OWL DL or EDOAL, for example, relations between object properties in one ontology and datatype properties in the other, or many types of transformations. Since heterogeneous representation hinders an interoperability between real-world ontologies, this issue needs more work to be done in order to represent and utilize alignments that contain correspondences of this type. In addition, several elements of the methodology presented here were manually-intensive and will likely not scale well to large alignments. For instance, we began our process using a high quality set of simple correspondences as a starting point. If such an alignment is unavailable for an ontology pair, this step could be quite time consuming in itself. Additionally, we are not aware of any conflict resolution system for complex alignments (Alcomo [Meilicke, 2011] is only for simple alignments), so for the moment, this is a manual step.

7.4 Evaluation of complex matchers

With respect to the evaluation, the metrics used here are the classical precision, recall and F-measure. Applied to complex correspondences, they present some limits. First, the relation of the relationship (e.g., equivalence, subsumption, disjunction) is not taken into account. Second, the same correspondence can be expressed in different ways: in this evaluation, we manually compared two expressions; however, this approach is not scalable. Third, we could consider the confidence of correspondences (here we assume that all generated correspondences by matchers are 1.0). Finally, the evaluation is not task-centered in the sense that the alignments generated by the approaches were not applied to query rewriting or ontology merging.

Nevertheless, the evaluation of the matchers shows that there is room for improvements in complex alignment generation. In our evaluation, we observe that good precision is often achieved at the expense of recall. The approaches [Ritze et al., 2009, Ritze et al., 2010] only found correspondences in the *cmt-ekaw* pair. Even though they both achieved a good precision performance (0.5 and 1.0), they had a low recall (0.2). The approach of [Jiang et al., 2016] is the only one of the four approaches that considers object property restrictions, which are needed for query rewriting. However, many of the output correspondences were incorrect. The approach of [Faria et al., 2018] could discover the most correct correspondences overall but their F-measure on the *conference-ekaw* and *cmt-conference* pairs (0.26 and 0.28) are open to improvement.

8 Conclusion and perspectives

This paper has presented a consensual dataset supporting the task of evaluating complex matching approaches. This dataset has been used in the evaluation of complex matchers in OAEI 2019. While most works in the literature are focused on describing the datasets themselves, less attention has been given to the methodological aspects of the (manual) creation of (complex) reference alignments. We described the design choices and the methodology followed for constructing it. In particular, the issues the experts faced during the process have been discussed and the lessons learned and perspectives in the field have been pointed out. Generating such alignments is, in fact, a time-consuming and sometimes tedious task that requires different human annotators and that has to be guided by a “consensual” methodology. However, starting from that does not guarantee uniform interpretations along the process, as we could observe in this experience.

Summing up our findings, we highlight that:

- ontology interpretations and their propagation has a strong impact in the generated correspondences;

- reaching the consensus needs a strong collaboration; keeping track of the usage of the alignments and of the evaluation metric to optimise helps in their construction;
- there is a strong relation between the kind of task and the expressiveness of the correspondences;
- existing alignment representation languages do not cover all possible constructions and transformations;
- there is a lack of tools supporting the whole process; automatic evaluation is also an open issue.

In the future, the following objectives should be considered: a) to populate the ontologies in order to be able to apply complex approaches relying on instances; b) to propose an automatic evaluation strategy of complex alignments based on instances comparison rather than on the syntactic or semantic comparison of correspondences; c) to extend the methodology and the dataset itself in order to cover (m:n) correspondences and to address the task of ontology merging; and d) to develop a tool able to support the managing of complex alignments (creation, evolution, visualisation, versioning, collaboration, etc.); e) to work on extending existing alignment representation languages in order to cover the whole space of representation possibilities; f) studying the possibility of computing minimal complex correspondences (considering logical inference) from which the other ones can be generated, helping the task of manually creating such correspondences.

References

[Abacha and Zweigenbaum, 2014]Abacha, A. B. and Zweigenbaum, P. (2014). Means: une approche sémantique pour la recherche de réponses aux questions médicales. *TAL*, 55(1):71–104.

[Algergawy et al., 2018]Algergawy, A., Cheatham, M., Faria, D., Ferrara, A., Fundulaki, I., Harrow, I., Hertling, S., Jiménez-Ruiz, E., Karam, N., Khiat, A., Lambrix, P., Liand, H., Montanelli, S., Paulheim, H., Pesquita, C., Saveta, T., Schmidt, D., Shvaiko, P., Splendiani, A., Thiéblin, E., Trojahn, C., Vatasčinová, J., Zamazal, O., and Zhou, L. (2018). Results of the ontology alignment evaluation initiative 2018. In *OM-2018: Proceedings of the Twelfth International Workshop on Ontology Matching*.

[Cheatham and Hitzler, 2014]Cheatham, M. and Hitzler, P. (2014). Conference v2. 0: An uncertain version of the oaei conference benchmark. In *International Semantic Web Conference*, pages 33–48. Springer.

[Dekhtyar and Hayes, 2006]Dekhtyar, A. and Hayes, J. H. (2006). Good benchmarks are hard to find: Toward the benchmark for information retrieval applications in software engineering. In *Proceedings of the 22th International Conference on Software Maintenance*.

[Dhamankar et al., 2004]Dhamankar, R., Lee, Y., Doan, A., Halevy, A., and Domingos, P. (2004). iMAP: discovering complex semantic matches between database schemas. In *Proceedings of the 2004 ACM SIGMOD international conference on Management of data*, pages 383–394. ACM.

[Do et al., 2002]Do, H.-H., Melnik, S., and Rahm, E. (2002). Comparison of schema matching evaluations. In *Net. ObjectDays: International Conference on Object-Oriented and Internet-Based Technologies, Concepts, and Applications for a Networked World*, pages 221–237. Springer.

[Doan, 2005]Doan, A. (2005). The illinois semantic integration archive.

[Euzenat et al., 2009]Euzenat, J., Ferrara, A., Hollink, L., Isaac, A., Joslyn, C., Malaisé, V., Meilicke, C., Nikolov, A., Pane, J., Sabou, M., Scharffe, F., Shvaiko, P., Spiliopoulos, V., Stuckenschmidt, H., Sváb-Zamazal, O., Svátek, V., dos Santos, C. T., Vouros, G. A., and Wang, S. (2009). Results of the ontology alignment evaluation initiative 2009. In *Proceedings of the 4th International Workshop on Ontology Matching (OM-2009) collocated with the 8th International Semantic Web Conference (ISWC-2009) Chantilly, USA, October 25, 2009.*

[Euzenat and Shvaiko, 2013]Euzenat, J. and Shvaiko, P. (2013). *Ontology Matching.* Springer Berlin Heidelberg.

[Faria et al., 2018]Faria, D., Pesquita, C., Balasubramani, B. S., Tervo, T., Carrio, D., Garrilha, R., Couto, F. M., and Cruz, I. F. (2018). Results of aml participation in oaei 2018. In *OM-2018: Proceedings of the Twelfth International Workshop on Ontology Matching.*

[Halpin et al., 2010]Halpin, H., Hayes, P. J., McCusker, J. P., McGuinness, D. L., and Thompson, H. S. (2010). When owl:sameas isn't the same: An analysis of identity in linked data. In Patel-Schneider, P. F., Pan, Y., Hitzler, P., Mika, P., Zhang, L., Pan, J. Z., Horrocks, I., and Glimm, B., editors, *The Semantic Web – ISWC 2010*, pages 305–320, Berlin, Heidelberg. Springer Berlin Heidelberg.

[He et al., 2004]He, B., Chang, K. C.-C., and Han, J. (2004). Discovering complex matchings across web query interfaces: a correlation mining approach. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 148–157. ACM Press.

[Hollink et al., 2008]Hollink, L., Van Assem, M., Wang, S., Isaac, A., and Schreiber, G. (2008). Two Variations on Ontology Alignment Evaluation: Methodological Issues. In *5th European Semantic Web Conference*, pages 388–401.

[Isaac et al., 2008]Isaac, A., Matthezing, H., van der Meij, L., Schlobach, S., Wang, S., and Zinn, C. (2008). Putting Ontology Alignment in Context: Usage Scenarios, Deployment and Evaluation in a Library Case. In *5th European Semantic Web Conference*, pages 402–417.

[Jiang et al., 2016]Jiang, S., Lowd, D., Kafle, S., and Dou, D. (2016). Ontology matching with knowledge rules. In *Transactions on Large-Scale Data-and Knowledge-Centered Systems XXVIII*, pages 75–95. Springer.

[Maedche et al., 2002]Maedche, A., Motik, B., Silva, N., and Volz, R. (2002). Mafra—a mapping framework for distributed ontologies. In *International Conference on Knowledge Engineering and Knowledge Management*, pages 235–250. Springer.

[Makris et al., 2012]Makris, K., Bikakis, N., Gioldasis, N., and Christodoulakis, S. (2012). SPARQL-RW: transparent query access over mapped RDF data sources. In *15th International Conference on Extending Database Technology*, pages 610–613. ACM.

[Maltese et al., 2010]Maltese, V., Giunchiglia, F., and Autayeu, A. (2010). Save up to 99% of your time in mapping validation. In *On the Move to Meaningful Internet Systems, OTM 2010 - Confederated International Conferences: CoopIS, IS, DOA and ODBASE, Hersonissos, Crete, Greece, October 25-29*, pages 1044–1060.

[Mathur et al., 2018]Mathur, S. N., O'Sullivan, D., and Brennan, R. (2018). Milan: Automatic generation of r2rml mappings. In *Proceedings of the 26th Irish Conference on Artificial Intelligence and Cognitive Science, Dublin, Ireland, 7th December*, pages 1–12.

[Meilicke, 2011]Meilicke, C. (2011). *Alignment incoherence in ontology matching*. PhD thesis, Universität Mannheim.

[Meilicke and Stuckenschmidt, 2008]Meilicke, C. and Stuckenschmidt, H. (2008). Incoherence as a basis for measuring the quality of ontology mappings. In *3rd International Conference on Ontology Matching-Volume 431*, pages 1–12.

[Meilicke et al., 2009]Meilicke, C., Stuckenschmidt, H., and Šváb-Zamazal, O. (2009). A reasoning-based support tool for ontology mapping evaluation. In *European Semantic Web Conference*, pages 878–882. Springer.

[Nunes et al., 2011]Nunes, B. P., Mera, A., Casanova, M. A., Breitman, K. K., and Leme, L. A. P. (2011). Complex Matching of RDF Datatype Properties. In *6th ISWC workshop on ontology matching*.

[Oliveira and Pesquita, 2018]Oliveira, D. and Pesquita, C. (2018). Improving the interoperability of biomedical ontologies with compound alignments. *Journal of Biomedical Semantics*, 9(1).

[Parundekar et al., 2010]Parundekar, R., Knoblock, C. A., and Ambite, J. L. (2010). Linking and building ontologies of linked data. In *ISWC*, pages 598–614. Springer.

[Parundekar et al., 2012]Parundekar, R., Knoblock, C. A., and Ambite, J. L. (2012). Discovering concept coverings in ontologies of linked data sources. In *ISWC*, pages 427–443. Springer.

[Pinkel et al., 2017]Pinkel, C., Binnig, C., Jiménez-Ruiz, E., Kharlamov, E., May, W., Nikolov, A., Sasa Bastinos, A., Skjæveland, M. G., Solimando, A., Taheriyan, M., Heupel, C., and Horrocks, I. (2017). RODI: Benchmarking relational-to-ontology mapping generation quality. *Semantic Web*, 9(1):25–52.

[Qin et al., 2007]Qin, H., Dou, D., and LePendu, P. (2007). Discovering executable semantic mappings between ontologies. In *On the Move to Meaningful Internet Systems*, pages 832–849. Springer.

[Ritze et al., 2009]Ritze, D., Meilicke, C., Šváb Zamazal, O., and Stuckenschmidt, H. (2009). A pattern-based ontology matching approach for detecting complex correspondences. In *4th ISWC workshop on ontology matching*, pages 25–36.

[Ritze et al., 2010]Ritze, D., Völker, J., Meilicke, C., and Šváb Zamazal, O. (2010). Linguistic analysis for complex ontology matching. In *5th workshop on ontology matching*, pages 1–12.

[Scharffe, 2009]Scharffe, F. (2009). *Correspondence Patterns Representation*. PhD thesis, Faculty of Mathematics, Computer Science and University of Innsbruck.

[Serpeloni et al., 2011]Serpeloni, F., Moraes, R., and Bonacin, R. (2011). Ontology mapping validation. *International Journal of Web Portals*, 3(3):1–11.

[Sim et al., 2003]Sim, S. E., Easterbrook, S., and Holt, R. C. (2003). Using benchmarking to advance research: A challenge to software engineering. In *Proceedings of the 25th International Conference on Software Engineering*, ICSE '03, pages 74–83, Washington, DC, USA. IEEE Computer Society.

[Singh et al., 2017]Singh, A., Debruyne, C., Brennan, R., Meehan, A., and O'Sullivan, D. (2017). Extension of the m-gov ontology mapping framework for increased traceability. In *Proceedings of the 12th International Workshop on Ontology Matching co-located with the 16th International Semantic Web Conference (ISWC 2017), Vienna, Austria, October 21*, pages 49–60.

[Solimando et al., 2017]Solimando, A., Jimenez-Ruiz, E., and Guerrini, G. (2017). Minimizing conservativity violations in ontology alignments: Algorithms and evaluation. *Knowledge and Information Systems*, 51(3):775–819.

[Solimando et al., 2014]Solimando, A., Jiménez-Ruiz, E., and Pinkel, C. (2014). Evaluating ontology alignment systems in query answering tasks. In *ISWC 2014 International Conference on Posters & Demonstrations*, pages 301–304.

[Stapleton et al., 2014]Stapleton, G., Howse, J., Bonnington, A., and Burton, J. (2014). A vision for diagrammatic ontology engineering. In *International Workshop on Visualizations and User Interfaces for Knowledge Engineering and Linked Data Analytics*, pages 1–13.

[Stevens et al., 2018]Stevens, R., Lord, P., Malone, J., and Matentzoglu, N. (2018). Measuring expert performance at manually classifying domain entities under upper ontology classes. *Journal of Web Semantics.*

[Thiéblin et al., 2017]Thiéblin, E., Amarger, F., Hernandez, N., Roussey, C., and Trojahn, C. (2017). Cross-querying lod datasets using complex alignments: An application to agronomic taxa. In *Research Conference on Metadata and Semantics Research*, pages 25–37. Springer.

[Thiéblin et al., 2018a]Thiéblin, É., Cheatham, M., dos Santos, C. T., Zamazal, O., and Zhou, L. (2018a). The first version of the OAEI complex alignment benchmark. In *Proceedings of the ISWC 2018 Posters & Demonstrations, Industry and Blue Sky Ideas Tracks co-located with 17th International Semantic Web Conference (ISWC 2018), Monterey, USA, October 8th - to - 12th, 2018.*

[Thiéblin et al., 2018b]Thiéblin, É., Haemmerlé, O., Hernandez, N., and Trojahn, C. (2018b). Task-oriented complex ontology alignment: Two alignment evaluation sets. In *The Semantic Web - 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3-7, 2018, Proceedings*, pages 655–670.

[Thieblin et al., 2019]Thieblin, E., Haemmerle, O., Hernandez, N., and Trojahn, C. (2019). Survey on complex ontology matching. *semantic Web Journal (to appear).*

[Thiéblin et al., 2018]Thiéblin, E., Haemmerlé, O., and Trojahn, C. (2018). Complex matching based on competency questions for alignment: a first sketch. In *Ontology Matching Workshop.*

[Tordai et al., 2011]Tordai, A., van Ossenbruggen, J., Schreiber, G., and Wielinga, B. (2011). Let's agree to disagree: On the evaluation of vocabulary alignment. In *Proceedings of the Sixth International Conference on Knowledge Capture*, K-CAP '11, pages 65–72, New York, NY, USA. ACM.

[Van Hage et al., 2007]Van Hage, W. R., Isaac, A., and Aleksovski, Z. (2007). Sample evaluation of ontology-matching systems. In *EON*, volume 2007, pages 41–50.

[Visser et al., 1997]Visser, P. R., Jones, D. M., Bench-Capon, T. J., and Shave, M. (1997). An analysis of ontology mismatches; heterogeneity versus interoperability. In *AAAI 1997 Spring Symposium on Ontological Engineering, Stanford CA., USA*, pages 164–72.

[Walshe et al., 2016]Walshe, B., Brennan, R., and O'Sullivan, D. (2016). Bayes-recce: A bayesian model for detecting restriction class correspondences in linked open data knowledge bases. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 12(2):25–52.

[Zamazal and Svátek, 2017]Zamazal, O. and Svátek, V. (2017). The Ten-Year OntoFarm and its Fertilization within the Onto-Sphere. *Web Semantics: Science, Services and Agents on the World Wide Web*, 43:46–53.

[Zhou et al., 2018]Zhou, L., Cheatham, M., Krisnadhi, A., and Hitzler, P. (2018). In *The Semantic Web - ISWC 2018 - 17th International Semantic Web Conference, Monterey, CA, USA, October 8-12, 2018, Proceedings, Part II*, pages 273–288.

[Šváb et al., 2005]Šváb, O., Svátek, V., Berka, P., Rak, D., and Tomášek, P. (2005). Ontofarm: Towards an experimental collection of parallel ontologies. *Poster Track of ISWC*, 2005.

9 Acknowledgments

We would like to thank Lu Zhou for his help on the creation of the correspondences. Ondřej Zamazal was supported by the CSF grant no. 18-23964S and by long term institutional support of research activities by Faculty of Informatics and Statistics, University of Economics, Prague.