Semantic precision and recall for ontology alignment evaluation

Jérôme Euzenat

Semantic Precision and Recall for Ontology Alignment Evaluation

Jérôme Euzenat
INRIA Rhône-Alpes
Monbonnot, France
Jerome.Euzenat@inrialpes.fr

Abstract

In order to evaluate ontology matching algorithms it is necessary to confront them with test ontologies and to compare the results with some reference. The most prominent comparison criteria are precision and recall originating from information retrieval. Precision and recall are thought of as some degree of correction and completeness of results. However, when the objects to compare are semantically defined, like ontologies and alignments, it can happen that a fully correct alignment has low precision. This is due to the restricted set-theoretic foundation of these measures. Drawing on previous syntactic generalizations of precision and recall, semantically justified measures that satisfy maximal precision and maximal recall for correct and complete alignments is proposed. These new measures are compatible with classical precision and recall and can be computed.

1 Introduction

Ontology matching is an important problem for which many algorithms have been provided [Euzenat and Shvaiko, 2007]. Given a pair of ontologies, these algorithms compute a set of correspondences between entities of these ontologies, called alignment.

In order to evaluate such algorithms, they are confronted with ontologies to match and the resulting alignment (A) is compared to a reference alignment (R) based on some criterion. The usual approach for evaluating the returned alignments is to consider them as sets of correspondences and to apply precision and recall originating from information retrieval [van Rijsbergen, 1975] and adapted to the matching task. Precision and recall are the ratio of the number of true positive ($|R \cap A|$) on that of the retrieved correspondences ($|A|$) and those expected ($|R|$) respectively. These criteria are well understood and widely accepted.

There are problems with using precision and recall in this context. As reported in [Ehrig and Euzenat, 2005], these measures have the drawback to be of the all-or-nothing kind. An alignment may be very close to the expected result and another quite remote from it and both sharing the same precision and recall values. The reason for this is that the criteria only compare two sets of correspondences without considering if these correspondences are close or remote to each other: if they are not the same exact correspondences, they score zero. They both score identically low, despite their different quality.

Moreover, there can be semantically equivalent alignments which are not identical. A fair measure of alignment quality should rank these alignments with the same values (or at least, closer values than non equivalent alignments). It is thus necessary to design semantically grounded alignment measures instead of measures based on their syntactic equality.

In this paper we investigate some measures that generalize precision and recall in order to overcome the problems presented above. We first provide the basic definitions of alignments, semantics, precision and recall as well as a motivating example (§2). We then present the framework for generalizing precision and recall introduced in [Ehrig and Euzenat, 2005] (§3). From that point we investigate and propose new semantically justified evaluation versions of precision and recall. We discuss their properties and define complementary measures (§4). We show on the motivating example that the proposed measures improve on previous ones (§5).

2 Foundations

We first precisely define what are alignments through their syntax (§2.1) and semantics (§2.2) before introducing precision and recall adapted to alignments (§2.3).

We will consider ontologies as logics. The languages used in the semantics web such as RDF or OWL are indeed logics. The semantics of the ontologies are given through their set of models.

2.1 Alignments

The result of matching, called an alignment, is a set of pairs of entities $\langle e, e' \rangle$ from two ontologies o and o' that are supposed to satisfy a certain relation r with a certain confidence n.

Definition 1 (Alignment, correspondence)**.** *Given two ontologies o and o', an alignment between o and o' is a set of correspondences (i.e., 4-uples): $\langle e, e', r, n \rangle$ with $e \in o$ and $e' \in o'$ being the two matched entities, r being a relationship holding between e and e', and n expressing the level of confidence in this correspondence.*

For the sake of simplicity, we will here only consider correspondences as triples $\langle e, e', r \rangle$. The best way to compare results with confidence is to plot their precision/recall functions. The examples are only provided for simple ontologies, which are class hierarchies but do not depend on this simple language.

Figure 1 presents two ontologies together with five alignments R, A_1, A_2, A_3, and A_4. R is the reference alignment and can be expressed by the following equations:

$$R = \begin{cases} \text{Employee} = \text{Worker} & \text{Accounting} = \text{Headquarters} \\ \text{Production} \leq \text{Japan} & \text{Marketing} \leq \text{Spain} \end{cases}$$

The other alignments share the first two correspondences with the reference alignment and have additional ones:

$$A_1 = \begin{cases} \text{Employee} = \text{Worker} & \text{Accounting} = \text{Headquarters} \\ \text{Electronics} \leq \text{Japan} & \text{Computer} \leq \text{Spain} \end{cases}$$

$$A_3 = \begin{cases} \text{Employee} = \text{Worker} & \text{Accounting} = \text{Headquarters} \\ \text{Electronics} \leq \text{Worker} & \text{Computer} \leq \text{Worker} \end{cases}$$

$$A_4 = \begin{cases} \text{Employee} = \text{Worker} & \text{Accounting} = \text{Headquarters} \\ \text{Optics} \geq \text{Spain} & \text{Marketing} \geq \text{Saleforce} \end{cases}$$

Alignment A_2 contains more correspondences than the others, it is made of the following:

$$A_2 = \begin{cases} \text{Employee} \leq \text{Worker} & \text{Accounting} \leq \text{Headquarters} \\ \text{Employee} \geq \text{Worker} & \text{Accounting} \geq \text{Headquarters} \\ \text{Production} \leq \text{Japan} & \text{Marketing} \leq \text{Spain} \end{cases}$$

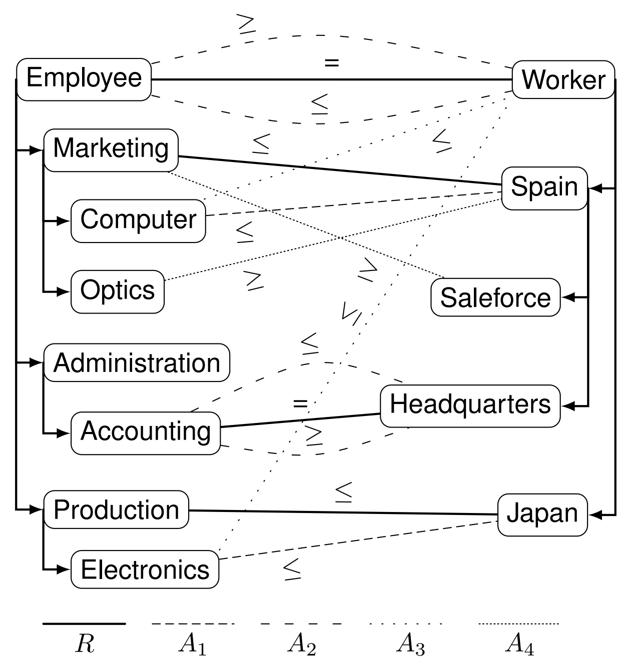

Figure 1: Five class alignments between two ontologies (only the classes involved in correspondences are displayed).

2.2 Semantics of alignments

In the line of the work on data integration [Ghidini and Serafini, 1998], we provide a first-order model theoretic semantics. It depends on the semantics of ontologies but does not interfere with it. In fact, given a set of ontologies and a set of alignments between them, we can evaluate the semantics of the whole system in function of the semantics of each individual ontology. The semantics of an ontology is given by its set of models.

Definition 2 (Models of an ontology)**.** *Given an ontology o, a model of o is a function m from elements of o to elements of a domain of interpretation Δ. The set of models of an ontology is noted $\mathcal{M}(o)$.*

Because the models of various ontologies can have different interpretation domains, we use the notion of an equalising function, which helps making these domains commensurate.

Definition 3 (Equilising function)**.** *Given a family of interpretations $\langle I_o, \Delta_o \rangle_{o \in \Omega}$ of a set of ontologies Ω, an equalising function for $\langle I_o, \Delta_o \rangle_{o \in \Omega}$ is a family of functions $\gamma = (\gamma_o : \Delta_o \longrightarrow U)_{o \in \Omega}$ from the domains of interpretation to a global domain of interpretation U. The set of all equalising functions is called Γ.*

When it is unambiguous, we will use γ as a function. The goal of this γ function is only to be able to (theoretically) compare elements of the domain of interpretation. It is simpler than the use of domain relations in distributed first order logics [Ghidini and Serafini, 1998] in the sense that there is one function per domain instead of relations for each pair of domains.

The relations used in correspondences do not necessarily belong to the ontology languages. As such, they do not have to be interpreted by the ontology semantics. Therefore, we have to provide semantics for them.

Definition 4 (Interpretation of alignment relations)**.** *Given r an alignment relation and U a global domain of interpretation, r is interpreted as a binary relation over U, i.e., $r^U \subseteq U \times U$.*

The definition of correspondence satisfiability relies on γ and the interpretation of relations. It requires that in the equalised models, the correspondences be satisfied.

Definition 5 (Satisfied correspondence)**.** *A correspondence $c = \langle e, e', r \rangle$ is satisfied for an equalising function γ by two models m, m' of o, o' if and only if $\gamma_o \cdot m \in \mathcal{M}(o)$, $\gamma_{o'} \cdot m' \in \mathcal{M}(o')$ and*

$$\langle \gamma_o(m(e)), \gamma_{o'}(m'(e')) \rangle \in r^U$$

This is denoted as $m, m' \models_\gamma c$.

For instance, in the language used as example, if m and m' are respective models of o and o':

$$m, m' \models_\gamma \langle c, c', = \rangle \text{ iff } \gamma_o(m(c)) = \gamma_{o'}(m'(c'))$$
$$m, m' \models_\gamma \langle c, c', \leq \rangle \text{ iff } \gamma_o(m(c)) \subseteq \gamma_{o'}(m'(c'))$$
$$m, m' \models_\gamma \langle c, c', \geq \rangle \text{ iff } \gamma_o(m(c)) \supseteq \gamma_{o'}(m'(c'))$$
$$m, m' \models_\gamma \langle c, c', \perp \rangle \text{ iff } \gamma_o(m(c)) \cap \gamma_{o'}(m'(c')) = \emptyset$$

Definition 6 (Satisfiable alignment)**.** *An alignment A of two ontologies o and o' is said satisfiable if and only if*

$$\exists m \in \mathcal{M}(o), \exists m' \in \mathcal{M}(o'), \exists \gamma \in \Gamma; \forall c \in A, m, m' \models_\gamma c$$

Thus, an alignment is satisfiable if there are models of the ontologies that can be combined in such a way that this alignment makes sense.

Definition 7 (Models of aligned ontologies)**.** *Given two ontologies o and o' and an alignment A between these ontologies, a model of these aligned ontologies is a triple $\langle m, m', \gamma \rangle \in \mathcal{M}(o) \times \mathcal{M}(o') \times \Gamma$, such that A is satisfied by $\langle m, m', \gamma \rangle$.*

In that respect, the alignment acts as a model filter for the ontologies. It selects the interpretation of ontologies which are coherent with the alignments. Note, this allows to transfer information from one ontology to another since reducing the set of models will entail more consequences in each aligned ontology.

In this paper we consider those consequences of aligned ontologies that are correspondences.

Definition 8 (α-Consequence of aligned ontologies)**.** *Given two ontologies o and o' and an alignment A between these ontologies, a correspondence δ is a α-consequence of o, o' and A (noted $A \models \delta$) if and only if for all models $\langle m, m', \gamma \rangle$ of o, o' and A, $m, m' \models_\gamma \delta$ (the set of α-consequences is noted by $Cn(A)$).*

For α-consequences, A_2 is strictly equivalent to R (i.e., $A_2 \models R$ and $R \models A_2$). In fact, the aligned ontologies with A_2 and R have exactly the same models.

It is noteworthy that, given an alignment, the α-consequences of this alignment can be larger than it. If the alignment is not satisfiable, then any correspondence is a α-consequence of it.

Such a formalism helps defining the meaning of alignments: it tells what are the consequences of ontologies with alignments. It is particularly useful for deciding if delivered alignments are consistent and for specifying what is expected from matching algorithms and how they should be designed or evaluated. It can be naturally extended to distributed systems in the sense of [Ghidini and Serafini, 1998], i.e., sets of ontologies and alignments.

2.3 Precision and recall

Precision and recall are commonplace measures in information retrieval. They are based on the comparison of an expected result and the effective result of the evaluated system. These results are considered as a set of items, e.g., the documents to be retrieved.

Since these measures are commonly used and well understood, they have been adapted for ontology alignment evaluation [Do *et al.*, 2002]. In this case, sets of documents are replaced by sets of correspondences, i.e., alignments. The alignment (A) returned by the system to evaluate is compared to a reference alignment (R).

Like in information retrieval, precision measures the ratio of correctly found correspondences (true positives) over the total number of returned correspondences (true positives and

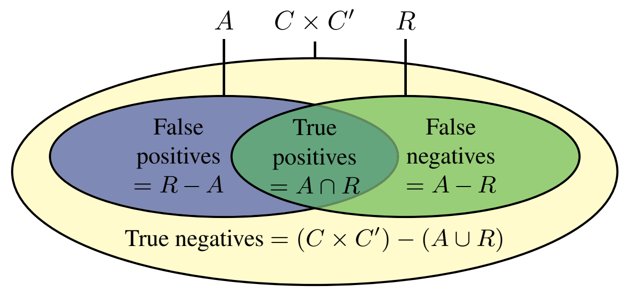

Figure 2: Two alignments seen as sets of correspondences and their relations.

false positives). In logical terms this is supposed to measure the correctness of the method. Recall measures the ratio of correctly found correspondences (true positives) over the total number of expected correspondences (true positives and true negatives). In logical terms, this is a completeness measure. This is displayed in Figure 2.

Definition 9 (Precision and recall)**.** *Given a reference alignment R, the precision of some alignment A is given by*

$$P(A, R) = \frac{|R \cap A|}{|A|}.$$

and recall is given by

$$R(A, R) = \frac{|R \cap A|}{|R|}.$$

Other measures, such as fallout, F-measure, noise, silence, etc. can be derived from precision and recall [Do *et al.*, 2002].

For the examples of Figure 1, the two first columns of Table 1 display precision and recall. As expected, evaluating the reference against itself yields a maximal precision and recall. All the other alignments share the two first correspondences with R and miss the two other ones so they have a precision and recall of 50% (but A_2 which is handicapped by having more than 4 correspondences).

The semantics of alignments has not been taken into account since an alignment, like A_2 equivalent to R score worse than an incorrect and incomplete alignment, like A_4.

3 Generalizing precision and recall

Even if they are well understood and widely accepted measures, precision and recall have the drawback that whatever correspondence has not been found is definitely not considered. As a result, they do not discriminate between a bad and a better alignment.

Indeed, when considering the example of Figure 1, alignment A_4 is arguably worse than the others, because its additional correspondences are measurably more different from the reference ones, but it scores the same as the other alignments.

As precision and recall are easily explained measures, it is useful to maintain the precision and recall structure when looking for new measures. This also ensures that measures derived from precision and recall (e.g., F-measure) still can be computed easily. [Ehrig and Euzenat, 2005] proposed an abstract generalization of precision and recall that has been instantiated with syntactic measures. This allows to take into account "near misses", i.e., incorrect correspondences that are close to the target (and that, for instance, can be more easily repaired).

Instead of comparing alignments set-theoretically, [Ehrig and Euzenat, 2005] proposes to measure the proximity of correspondence sets rather than the strict size of their overlap. Instead of taking the cardinal of the intersection of the two sets ($|R \cap A|$), the natural generalizations of precision and recall measure their proximity ($\omega(A, R)$).

Definition 10 (Generalized precision and recall)**.** *Given a reference alignment R and an overlap function ω between alignments, the precision of an alignment A is given by*

$$P_\omega(A, R) = \frac{\omega(A, R)}{|A|}$$

and recall is given by

$$R_\omega(A, R) = \frac{\omega(A, R)}{|R|}.$$

In order, for these new measures to be true generalizations, ω has to share some properties with $|R \cap A|$. In particular, the measure should be positive:

$$\forall A, B, \omega(A, B) \geq 0 \qquad \text{(positiveness)}$$

and should not exceed the minimal size of both sets:

$$\forall A, B, \omega(A, B) \leq min(|A|, |B|) \qquad \text{(maximality)}$$

This guarantees that the given values are within the unit interval [0 1]. Further, this measure should only add more flexibility to the usual precision and recall so their values cannot be worse than the initial evaluation:

$$\forall A, B, \omega(A, B) \geq |A \cap B| \qquad \text{(boundedness)}$$

Hence, the main constraint faced by the proximity is:

$$|A \cap R| \leq \omega(A, R) \leq min(|A|, |R|)$$

This is indeed a true generalization because, $\omega(A, R) = |A \cap R|$ satisfies all these properties.

4 Semantic precision and recall and other measures

Our main goal is to design a generalization of precision and recall that is semantically grounded. In consequence, those correspondences that are consequences of the evaluated alignments have to be considered as recalled and those that are consequence of the reference alignments as correct.

For that purpose we will attempt to follow the guidelines introduced in [Ehrig and Euzenat, 2005] as far as possible. We add some more constraints to a semantic precision and recall which consider correctness and completeness of an alignment as their limit:

$$R \models A \Rightarrow P_{sem}(A, R) = 1 \qquad \text{(max-correctness)}$$

$$A \models R \Rightarrow R_{sem}(A, R) = 1 \qquad \text{(max-completeness)}$$

$$Cn(A) = Cn(R) \text{ iff } P_{sem}(A, R) = 1 \text{ and } R_{sem}(A, R) = 1 \qquad \text{(definiteness)}$$

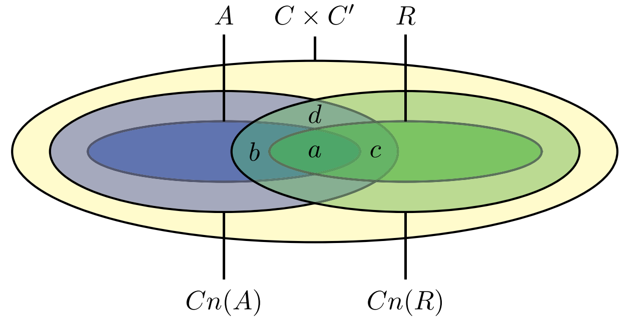

Figure 3: Two alignments and their relations through the set of their consequences.

The classical positiveness depends on $A = R$, it is replaced here by its semantic counterpart. In addition, we rephrase the previous properties by applying them to precision and recall instead of ω (M is any of precision and recall and M' its generalized counterpart):

$$M'(A, R) \geq 0 \qquad \text{(positiveness)}$$

$$M'(A, R) \leq 1 \qquad \text{(maximality)}$$

$$M'(A, R) \geq M(A, R) \qquad \text{(boundedness)}$$

4.1 Ideal model

The natural semantic extension of these measures consists of using the set of α-consequences (or deductive closure on the prover side) instead of $|A \cap R|$. This corresponds to taking as ω the size of the set identified by d instead of that identified by a in Figure 3.

In this case, the true positive becomes the correspondences that are consequences of both alignments and the usual definitions of true and false positives and negatives are only extended to alignment consequences.

Definition 11 (Ideal semantic precision and recall)**.** *Given a reference alignment R, the precision of some alignment A is given by*

$$P_{ideal}(A, R) = \frac{|Cn(R) \cap Cn(A)|}{|Cn(A)|} = P(Cn(A), Cn(R))$$

and recall is given by

$$R_{ideal}(A, R) = \frac{|Cn(R) \cap Cn(A)|}{|Cn(R)|} = R(Cn(A), Cn(R))$$

This ideal way of dealing with semantic precision and recall can be applied to any language with a semantics. It is not restricted to alignments and as soon as the notion of consequence is defined from an alignment, it can be applied.

These measures are different from the extensions of [Ehrig and Euzenat, 2005] because the divisor has changed. However, for a language with a well-defined semantics this measure is a natural extension of precision and recall. Most of the required properties are satisfied by these ideal measures:

Property 1. *P_{ideal} and R_{ideal} satisfy:*

- *positiveness;*
- *maximality;*
- *completeness-maximality;*
- *correctness-maximality;*
- *definiteness;*

P_{ideal} and R_{ideal} do not necessarily satisfy boundedness.

This extension has two drawbacks: (1) both numerator and divisor could be infinite, yielding an undefined result, and (2) contrary to the objective of [Ehrig and Euzenat, 2005], these measures do not guarantee to provide better results than precision and recall in general, i.e., we do not have $P(A,R) \leq P_{ideal}(A,R)$ nor $R(A,R) \leq R_{ideal}(A,R)$. This is because there is no direct relation between the size of an alignment (or a set of axioms) and the size of its α-consequences.

4.2 Semantic precision and recall

In order to deal with the problems raised by the infinite character of the set of α-consequences, a natural way would be to compare the deductive reductions instead of the deductive closures. Unfortunately, the deductive reduction is usually not unique. We thus propose to use the deductive closure bounded by a finite set so that the result is finite. It is based on different sets of true positives as :

$$TP_P(A,R) = \{\delta \in A; R \models \delta\} = A \cap Cn(R)$$

and

$$TP_R(A,R) = \{\delta \in R; A \models \delta\} = Cn(A) \cap R$$

These two sets correspond respectively to the sets b and c in Figure 3. They are obviously finite since they are the intersection of a set of α-consequences with a finite alignment. They are not, however, a real count of the true positive by any means.

The semantic precision and recall are based on these sets:

Definition 12 (Semantic precision and recall)**.** *Given a reference alignment R, the precision of some alignment A is given by*

$$P_{sem}(A,R) = \frac{|A \cap Cn(R)|}{|A|}$$

and recall is given by

$$R_{sem}(A,R) = \frac{|Cn(A) \cap R|}{|R|}$$

Both values are defined when the alignments are finite. Moreover, the considered values can be computed if there exists a complete and correct prover for the languages because there is always a finite set of assertions to check (i.e., $Cn(A) \cap R = \{\delta \in R; A \models \delta\}$).

Property 2. *P_{sem} and R_{sem} satisfy:*

- *positiveness;*
- *maximality;*
- *boundedness;*
- *completeness-maximality;*
- *correctness-maximality;*
- *definiteness;*

These measures satisfy positiveness and boundedness (since it is clear that $Cn(X) \supseteq X$). They have the classically expected values: $P_{sem}(R,R) = 1$ and $R_{sem}(R,R) = 1$. They do not satisfy anymore, if $A \cap R = \emptyset$, then $P_{sem}(A,R) = 0$ which is replaced by if $Cn(A) \cap Cn(R) = \emptyset$, then $P_{sem}(A,R) = 0$.

4.3 Compactness and independence

Now we can find that a particular alignment is semantically equivalent to some reference alignment. However, what makes that the reference alignment has been chosen otherwise? Are there criteria that enable to measure the quality of these alignments so that some of them are better than others?

Indeed, more compact alignments seem to be preferable. Compactness can be measured as the number of correspondences. So it is possible to measure either, this absolute number or the ratio of correspondences in the reference alignments and the found alignments:

$$Compactness(A,R) = \frac{|R|}{|A|}$$

There is no reason that this measure cannot be higher than 1 (if the found alignment is more compact than the reference alignment).

Compactness depends on the raw set of correspondences is however primitive and non semantically grounded (it is especially useful for complete and correct alignments). A more adapted measure is that of independence.

$$Ind(A) = \frac{|\{c \in A; A - \{c\} \not\models c\}|}{|A|}$$

which checks that alignments are not redundant.

It measures the ratio of independent correspondences in an alignment independently of a reference. If we want to measure independence with regard to the reference alignment, it is possible to measure the independence of correspondences that do not appear in the reference alignment:

$$Ind(A,R) = \frac{|\{c \in A - R; A - \{c\} \not\models c\}|}{|A - R|}$$

In the examples considered here independence measures return 1. for all alignments.

5 Examples

In order to compare the behavior of the proposed measure, we compare it with previously provided measures on the examples of Figure 1. Additionally to standard precision and recall we compare them with two measures introduced in [Ehrig and Euzenat, 2005]: symmetry considers as close a correspondence in which a class is replaced by its direct sub- or super-class; effort takes as proximity the measure of the effort to produce for correcting the alignment. A third measure, oriented, is not defined in this context (because of the presence of non-equivalent correspondences).

The results are provided in Table 1. As can be expected, the results of the semantic measures match exactly the correctness and completeness of the alignment. These results

ω	standard		symm.		effort		semantic		
A	P	R	P	R	P	R	P	R	Comp.
R	1.	1.	1.	1.	1.	1.	1.	1.	1.
A_1	.5	.5	.75	.75	.8	.8	1.	.5	1.
A_2	.33	.5	.5	.75	.5	.75	1.	1.	.66
A_3	.5	.5	.5	.5	.5	.5	1.	.5	1.
A_4	.5	.5	.5	.5	.65	.65	.5	.5	1.

Table 1: Precision and recall results on the alignments of Figure 1.

are far more discriminating than the other ones as far as A_4 is concerned. The equivalent alignment A_2 which did not stand out with the other measures is now clearly identified as equivalent. An interesting aspect for A_1 which is correct but incomplete, is that the other measures fail to recognize this asymmetry. A_1 and A_3 are both correct and incomplete, they thus have the same semantic values while A_1 is arguably more complete than A_3 (in fact $A_1 \models A_3$): this is accounted better by the syntactic measures.

No redundancy was present in the selected alignments, so they all score 100% in independence. A_2 is less compact than R (and the others) as expected.

6 Related work

Relevant related work is that of [Langlais *et al.*, 1998] in computational linguistics, [Sun and Lin, 2001] in information retrieval and [Ehrig and Euzenat, 2005] in artificial intelligence. All rely on a syntactic distance between entities of the ontologies (or words or documents). They have been compared in [Ehrig and Euzenat, 2005]. However, these work tackle the problem of measuring how far is a result from the solution. Here, the semantic foundations only consider valid solutions: the semantic definitions account for the semantically equivalent but syntactically different results.

This explains why, with regard to [Ehrig and Euzenat, 2005], we used different but compatible properties not fully relying on a proximity measure ω. Hence, it should be possible to combine the semantic precision and recall with the proposed syntactic relaxed measures.

The results of Table 1 show how the semantic approach compares with [Ehrig and Euzenat, 2005] on a small example. One important difference is that the syntactic measures can be easily computed because they only work on the syntax of the alignments while the semantic measures require a prover for the ontology languages (and even for the alignment semantics).

7 Discussion

Evaluation of matching results is often made on the basis of the well-known and well-understood precision and recall measures. However, these measures are based on the syntax of alignments and are not able to take into account the semantics of the formalism.

In this paper we provided a semantics for alignments based on the semantics of ontologies and we designed semantic precision and recall measures that take advantage of this semantics. The definition of these measures is thus independent from the semantics of ontologies. We showed that these measures are compliant with (an adapted version of) constraints put on precision and recall generalization of [Ehrig and Euzenat, 2005] and that they behave as expected in the motivating example.

In particular, the new measures have the expected behavior:

– they are still maximal for the reference alignment;
– they correspond to correctness and completeness;
– they help discriminating between irrelevant alignments and not far from target ones.

The examples have been given based on the alignment of simple taxonomies with very limited languages. However, all the definitions are independent from this language. We plan to implement these measures and apply them to larger sets of data (results of the OAEI[1] evaluation campaigns for instance). This requires the use of a correct and complete prover for the considered ontology languages.

Acknowledgements

This work has been partially supported by the Knowledge Web European network of excellence (IST-2004-507482). Thanks to Mark Ehrig for pushing me on this track, to Stefano Zanobini for complaining about OAEI-2005 results and to Antoine Zimmermann for his comments.

References

[Do *et al.*, 2002] Hong-Hai Do, Sergei Melnik, and Erhard Rahm. Comparison of schema matching evaluations. In *Proc. GI-Workshop "Web and Databases", Erfurt (DE)*, 2002.

[Ehrig and Euzenat, 2005] Marc Ehrig and Jérôme Euzenat. Relaxed precision and recall for ontology matching. In Ben Ashpole, Jérôme Euzenat, Marc Ehrig, and Heiner Stuckenschmidt, editors, *Proc. K-Cap 2005 workshop on Integrating ontology, Banff (CA)*, pages 25–32, 2005.

[Euzenat and Shvaiko, 2007] Jérôme Euzenat and Pavel Shvaiko. *Ontology matching*. Springer, Heidelberg (DE), 2007. to appear.

[Ghidini and Serafini, 1998] Chiara Ghidini and Luciano Serafini. Distributed first order logics. In Franz Baader and Klaus Ulrich Schulz, editors, *Frontiers of Combining Systems 2*, pages 121–139, Berlin, 1998. Research Studies Press.

[Langlais *et al.*, 1998] Philippe Langlais, Jean Véronis, and Michel Simard. Methods and practical issues in evaluating alignment techniques. In *Proc. 17th international conference on Computational linguistics, Montréal (CA)*, pages 711–717, 1998.

[Sun and Lin, 2001] Aixin Sun and Ee-Peng Lin. Hierarchical text classification and evaluation. In *Proc. IEEE international conference on data mining*, pages 521–528, 2001.

[van Rijsbergen, 1975] Cornelis Joost (Keith) van Rijsbergen. *Information retrieval*. Butterworths, London (UK), 1975. http://www.dcs.gla.ac.uk/Keith/Preface.html.

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