Towards Semantics-Based Ontology Similarity

Rudi Araújo and H. Sofia Pinto

INESC-ID
Rua Alves Redol 9, Apartado 13069, 1000-029 Lisboa, Portugal
{rnna,sofia}@algos.inesc-id.pt

Abstract. As the Semantic Web emerges the problem of semantic heterogeneity is becoming more acute. Ontology matching techniques aim at tackling this problem by establishing correspondences between elements of the ontologies. These techniques rely on distance metrics, often called (dis)similarity measures, to assess the similarity of elements within the ontologies. Most of these approaches are either terminological, structural and/or extensional. However, recently some proposals for semantics-based measures have been put forward. We reason that these latter should receive more attention, since semantics are one of the key advantages of ontologies. Therefore, we present a set of semantic ontology similarity measures.

1 Introduction and Motivation

Although the Semantic Web will be responsible for the proliferation of ontologies, it is unrealistic to expect that every Semantic Web agent shares the same set of ontologies. Therefore, the need for establishing a consensus among many cooperating agents arises. This is a key motivation for ontology matching.

Ontology matching techniques take as input a set of ontologies, and output a set of correspondences between elements of the ontologies. Distance metrics, often called (dis)similarity measures, are used to estimate the similarity of elements of the ontologies. There are four kinds of similarity measures: lexical, structural, extensional and semantics-based. The difference between the latter and the others lies on the fact that they are sensitive to the logical nature of the knowledge representation formalism in which the ontologies are formalized. Hence, they are enabled to resort to deduction services, such as consistency checking. Given that, on the one hand, modern representation languages have well-defined semantics, and on the other hand, one of the key advantages of ontologies lies on their semantics, we believe that this type of measures should deserve more attention from the ontology matching community.

This paper elaborates on our own work, presented in [2], where we define a purely semantics based similarity measure. Here, we propose a more refined approach and extend the proposed measure, which was only sensitive to concept constructs of $\mathcal{ALC}$ ($\mathcal{A}$ttributive $\mathcal{L}$anguage with $\mathcal{C}$omplement) – intersection $\sqcap$, conjunction $\sqcup$ and complement $\neg$, – to deal with role constructs – universal $\forall$ and existential quantifiers $\exists$. However, our measure does not yet cover the whole scope of $\mathcal{ALC}$. Though limited in expressivity, $\mathcal{ALC}$ is the foundation of more expressive Description Logics (DLs) and is already sufficiently expressive to formalize many practical ontologies.

Throughout the rest of the paper, we assume the standard DL notation [3], except for concept equivalence, $C \doteq D$, to distinguish from TBox equivalence, $\mathcal{T}_1 \equiv \mathcal{T}_2$ (meaning $\mathcal{T}_1 \models \mathcal{T}_2$ and $\mathcal{T}_2 \models \mathcal{T}_1$). We denote $\mathcal{C}^*$ as the closure of the set of concept names $\mathcal{C}$ under the DL constructs $\sqcap$, $\sqcup$ and $\neg$. We further assume that every set of concept names $\mathcal{C}$ and roles $\mathcal{R}$ are finite.

The paper is organized as follows: in section 2 we summarize related work. We introduce the theory underlying the proposed measures in section 3, followed by both the concept and role similarity measures, and a toy example. In section 4 we present the results for real-world examples. We discuss the measures based on our results in section 5 and finish with conclusions and future directions in section 6.

2 Related Work

The use of similarity measures is not limited to ontology matching. Other uses include unsupervised or semi-automatic ontology clustering [11], and automated ontology merging, as performed by PROMPT [10] and CHIMAERA [9].

Approaches to similarity measures in DLs can be found in [7, 5, 4, 8]. The measure presented in [7] is based on extracting and comparing *concept signatures*: the elements within the ontology that are related to one concept. The work presented in [5] is an extensional dissimilarity measure that estimates the difference between two concepts in different ontologies through the set of individuals they share. Borgida *et al.* [4] propose the adaptation of known similarity measures to DLs: feature-based models, the semantic-network approach and information-content models. Finally, the work presented in [8] defines a similarity measure for each DL construct and computes the similarity between two concepts by agreggating these values.

3 Theory

We start by presenting the general framework underlying the similarity measures. Though different from [2], the definitions presented here can be shown as equivalent.

3.1 Foundation

All measures presented here are based on the same idea: counting *characteristic concepts*. By characteristic concepts, we mean any concept, formed using standard concept constructs on a set of concept names, which cannot be any more specific. For example, given the concept names Man and Woman, the concepts $\neg\mathsf{Man} \sqcap \neg\mathsf{Woman}$ and $\neg\mathsf{Man} \sqcap \mathsf{Woman}$ are two of four possible characteristic concepts. Characteristic concepts can be seen as propositional models. The formal definition follows.

Definition 1 (Characteristic Concept). *Let $\mathcal{C}$ be a finite set of DL concept names. A concept $C \in \mathcal{C}^*$ is characteristic iff it is an intersection of n literals, with $n = |\mathcal{C}|$, and each concept name in $\mathcal{C}$ occurs exactly once in it. The set of all possible characteristic concepts w.r.t. $\mathcal{C}$ is $\zeta(\mathcal{C})$.*

Given a TBox $\mathcal{T}$ containing a set of concept names $\mathcal{C}$, we need to determine the *characteristic acceptance set*: the sub-set of $\zeta(\mathcal{C})$ with only *consistent* characteristic concepts w.r.t. $\mathcal{T}$.

Definition 2 (Characteristic Acceptance Set). *Let $\mathcal{T}$ be a TBox containing a finite set of DL concept names $\mathcal{C}$. Given a set of characteristic concepts $S \subseteq \zeta(\mathcal{C})$, S is* accepted *by $\mathcal{T}$ iff every $C \in S$ is consistent in $\mathcal{T}$ (i.e., there exists a model $\mathcal{I}$ of $\mathcal{T}$ such that $C^{\mathcal{I}} \neq \emptyset$). The* characteristic acceptance set*, $Z(\mathcal{T}) \subseteq \zeta(\mathcal{C})$, is the maximal set accepted by $\mathcal{T}$.*

3.2 The Measures

In this section we introduce the measures for comparing concepts and roles. The former determines the similarity of two TBoxes disregarding any information concerning role restrictions.

Comparing Concepts Given two TBoxes, $\mathcal{T}_1$ and $\mathcal{T}_2$, we need to determine to which extent their acceptance sets overlap. In particular, if the ontologies are equivalent, then their acceptance sets are the same [1] and, therefore, their similarity assessment should be the highest. Although this assumption is rather intuitive for the maximum value, the same does not apply to the minimum value. The question of when two ontologies are totally dissimilar does not have a straight answer. One could consider that total dissimilarity corresponds to the case where the acceptance sets are disjoint. However, that depends on the interpretation given to similarity, since when the acceptance sets are disjoint, it means there is no *agreement* on a set of *consistent* characteristic concepts, while there may be agreement on a large number of *inconsistent* characteristic concepts.

Let us consider the following TBoxes that represent several typical cases of overlap between acceptance sets:

$$\mathcal{E}_1 = \{\mathsf{Man} \sqsubseteq \mathsf{Person} \sqcap \mathsf{Male}\}, \quad \mathcal{E}_2 = \{\mathsf{Man} \doteq \mathsf{Person} \sqcap \mathsf{Male}\},$$
$$\mathcal{E}_3 = \{\mathsf{Man} \sqsubseteq \mathsf{Person}, \mathsf{Man} \doteq \neg\mathsf{Person} \sqcup \neg\mathsf{Male}\} \ .$$

Their acceptance sets are shown in table 1 (with $Z_i = Z(\mathcal{E}_i)$). Each line of the table is a concept name, and each column is a characteristic concept, and a symbol at the intersection of a characteristic concept with a concept name indicates that the concept name occurs as a positive ($+$) or negative ($-$) literal.

Table 1. The characteristic acceptance sets for $\mathcal{E}_1$, $\mathcal{E}_2$ and $\mathcal{E}_3$.

$$\begin{array}{r|cccccccc}
 & \multicolumn{4}{c}{\overbrace{\qquad\qquad\qquad\qquad}^{Z_2}} & \multicolumn{2}{c}{\overbrace{\qquad\quad}^{Z_3}} & & \\
\hline
\mathsf{Person} & - & + & - & + & + & + & - & - \\
\mathsf{Man} & - & - & - & + & - & + & + & + \\
\mathsf{Male} & + & - & - & + & + & - & + & - \\
\hline
 & \multicolumn{5}{c}{\underbrace{\qquad\qquad\qquad\qquad\qquad}_{Z_1}} & & &
\end{array}$$

The *coverage* measure (γ) determines to which extent one of the ontologies covers the other, by measuring the percentage of overlap between the acceptance sets in proportion to the size of the acceptance set of one of the ontologies:

$$\gamma(\mathcal{T}_1, \mathcal{T}_2) = \frac{|Z(\mathcal{T}_1) \cap Z(\mathcal{T}_2)|}{|Z(\mathcal{T}_1)|} \ .$$

Applying coverage to the examples above yields $\gamma(\mathcal{E}_2, \mathcal{E}_1) = 100\%$, because $Z(\mathcal{E}_2)$ is a sub-set of $Z(\mathcal{E}_1)$. This measure is not symmetric, therefore inverting the order yields a different result, $\gamma(\mathcal{E}_1, \mathcal{E}_2) = 80\%$, because only 80% of $Z(\mathcal{E}_1)$ is contained within $Z(\mathcal{E}_2)$, since they only disagree on the consistency of one characteristic concept, $\neg\mathsf{Man} \sqcap \mathsf{Person} \sqcap \mathsf{Male}$ (which is consistent in $\mathcal{E}_1$ but not in $\mathcal{E}_2$). Given that $\mathcal{E}_2$ and $\mathcal{E}_3$ are disjoint, $\gamma(\mathcal{E}_2, \mathcal{E}_3) = \gamma(\mathcal{E}_3, \mathcal{E}_2) = 0\%$.

The *consistency agreement* measure (α_+) determines to which extent the ontologies agree on the set of consistent characteristic concepts. It measures the proportion of characteristic concepts consistent in both ontologies w.r.t. the set of characteristic concepts consistent in at least one of them:

$$\alpha_+(\mathcal{T}_1, \mathcal{T}_2) = \frac{|Z(\mathcal{T}_1) \cap Z(\mathcal{T}_2)|}{|Z(\mathcal{T}_1) \cup Z(\mathcal{T}_2)|} \ .$$

Contrary to γ, this measure is symmetric, thus $\alpha_+(\mathcal{E}_1, \mathcal{E}_2) = \alpha_+(\mathcal{E}_2, \mathcal{E}_1) = 80\%$. Again, since $\mathcal{E}_2$ and $\mathcal{E}_3$ are disjoint, their consistency agreement is 0%. We note that while coverage measures the overlap between $\mathcal{E}_1$ and $\mathcal{E}_3$ as 20% and 50%, the consistency agreement is lesser: $\alpha_+(\mathcal{E}_1, \mathcal{E}_3) = 16.67\%$. Therefore, consistency agreement measures the global overlap, while coverage only measures the overlap of one in relation to the other.

The previous measure only takes into account the set of consistent characteristic concepts. To measure the *inconsistency agreement*, the proportion of inconsistent characteristic concepts in both TBoxes w.r.t. the inconsistent characteristic concepts in at least one of them, we define the following measure:

$$\alpha_-(\mathcal{T}_1, \mathcal{T}_2) = \frac{|\zeta(\mathcal{C}) - (Z(\mathcal{T}_1) \cup Z(\mathcal{T}_2))|}{|\zeta(\mathcal{C}) - (Z(\mathcal{T}_1) \cap Z(\mathcal{T}_2))|} \ .$$

Even though all previous measures yield 0% for $\mathcal{E}_2$ and $\mathcal{E}_3$, this measure yields 25%, due to the fact that even though their acceptance sets are disjoint they agree that 2 out of 8 characteristic concepts are inconsistent.

The *agreement measure* (α) combines both consistency and inconsistency agreement, by measuring the proportion of characteristic concepts whose consistency is the same in both ontologies w.r.t. all possible characteristic concepts:

$$\alpha(\mathcal{T}_1, \mathcal{T}_2) = \frac{|\zeta(\mathcal{C}) - (Z(\mathcal{T}_1) \cup Z(\mathcal{T}_2))| + |Z(\mathcal{T}_1) \cap Z(\mathcal{T}_2)|}{2^{|\mathcal{C}|}} \ .$$

Agreement yields 87.50% for $\mathcal{E}_1$ and $\mathcal{E}_2$, since they agree on the consistency of 7 out of 8 characteristic concepts.

In [6] a set of similarity measures for the automatic evaluation of learnt ontologies against a gold-standard is proposed. In this context, if we apply the coverage measure, with, say, $\mathcal{T}_1$ as the gold-standard, then $\gamma(\mathcal{T}_1, \mathcal{T}_2)$ can be interpreted as *recall*, since it is the proportion of correct learnt characteristic concepts w.r.t. the total amount of correct characteristic concepts, while $\gamma(\mathcal{T}_2, \mathcal{T}_1)$ can be regarded as *precision*, since it measures the proportion of correct learnt characteristic concepts w.r.t. the set of all learnt characteristic concepts. We can thus define the F_1 measure:

$$F(\mathcal{T}_1, \mathcal{T}_2) = \frac{2 \times \gamma(\mathcal{T}_1, \mathcal{T}_2) \times \gamma(\mathcal{T}_2, \mathcal{T}_1)}{\gamma(\mathcal{T}_1, \mathcal{T}_2) + \gamma(\mathcal{T}_2, \mathcal{T}_1)} = \frac{2 \times |Z(\mathcal{T}_1) \cap Z(\mathcal{T}_2)|}{|Z(\mathcal{T}_1)| + |Z(\mathcal{T}_2)|} \ .$$

Comparing Roles Until this point we have only compared ontologies with concepts built using intersection ($\sqcap$), conjunction ($\sqcup$) and negation ($\neg$). However, this is very limited in expressiveness, given that standard DLs (at least as expressive as $\mathcal{ALC}$) allow building concepts with role constructs, namely universal quantification $\forall R.C$ and existential quantification $\exists R.C$. These constructs allow us to restrict the domain and range of a role:

$$\begin{aligned} \exists R.\top \sqsubseteq C & \qquad\qquad \text{(domain restriction)} \\ \top \sqsubseteq \forall R.C & \qquad\qquad \text{(range restriction)} \end{aligned} \tag{1}$$

The first axiom restricts instances of R to have as source an individual in C, while the second restricts instances of R to have as destination an individual contained in C. The power of these constructs is not limited, however, to general domain and range restriction. They can also be used to restrict the domain and range for a given concept D:

$$\begin{aligned} \exists R.D \sqsubseteq C & \qquad\qquad \text{(specific domain restriction)} \\ D \sqsubseteq \forall R.C & \qquad\qquad \text{(specific range restriction)} \end{aligned} \tag{2}$$

This means that the domain and range of R for individuals of D is C.

Given that $\mathcal{ALC}$ does not provide any other construct related to roles, it is reasonable to argue that the domain and range specifications characterize the semantics of a role in this language. Therefore, the task of computing role similarity can be reduced to comparing these domains and ranges. Since they are concepts, we can compare them using the measures introduced above.

Let us start by considering the case in (1) where the domain and range of a role consist of a single concept and they are not related to one another (i.e., it is not necessary to know the domain of a role to know its range). In this case, given a role R occuring in a TBox $\mathcal{T}$, we can define the concept δ_R and ρ_R as the domain and range of R respectively, i.e. such that $\mathcal{T} \models \exists R.\top \sqsubseteq \delta_R$ and $\mathcal{T} \models \top \sqsubseteq \forall R.\rho_R$. Given two TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$ in which the role R occurs, we can extract the domains of R, δ_R^1 and δ_R^2, and the ranges, ρ_R^1 and ρ_R^2. To compute the similarity of R, we can compare δ_R^1 with δ_R^2 and ρ_R^1 with ρ_R^2, which can be done by using the set of measures for concept similarity introduced above, but instead of using $Z(\mathcal{T}_1)$ and $Z(\mathcal{T}_2)$, we use only the sub-set of these that are subsumed by δ_R^1 and δ_R^2, to compare the domains, and by ρ_R^1 and ρ_R^2, to compare the ranges.

The procedure introduced above can be generalized to deal with specific domain and range restrictions (2). The problem is that the range of a role depends on the domain (and vice-versa), so there is no single δ_R and ρ_R concepts. The idea is to construct a different TBox describing the domain and range restrictions of a given role. Given a TBox $\mathcal{T}$ and a role R, our approach is to construct a TBox $\mathcal{T}^R$ from $\mathcal{T}$, such that, for each concept name A occuring in $\mathcal{T}$, we have two concept names in $\mathcal{T}^R$: dA and rA. dA should be seen as the set of *role instances* which have an individual in A as source, and rA the ones that have an individual in A as destination. The domain and range concepts are related through subsumption according to the restrictions in $\mathcal{T}$. This technique in many ways resembles the reification technique, in the sense that individuals of $\mathcal{T}^R$ are not instances of concepts of $\mathcal{T}$, but instances of the role R. A formal definition of $\mathcal{T}^R$ follows after the following auxiliary definition.

First we define the domain/range description, which is a TBox describing the relations between the domain and range of a role. For example, the axiom $\mathsf{A} \sqsubseteq \forall R.\mathsf{B}$

is described in the domain/range description as $\mathsf{dA} \sqsubseteq \mathsf{rB}$, and should be interpreted as "the set of role instances of R that have origin in A is a sub-set of the set of role instances of R that have destination in B."[1]

Definition 3 (Domain/Range Description). *Let R be a role occuring in a TBox $\mathcal{T}$. The* domain/range description *of R w.r.t. $\mathcal{T}$, $\mathrm{dr}_{\mathcal{T}}(R)$, is defined as:*

$$\mathrm{dr}_{\mathcal{T}}(R) = \{\delta(C) \sqsubseteq \rho(D) | \mathcal{T} \models C \sqsubseteq \forall R.D\} \ ,$$

where $\delta(C)$ (resp. $\rho(D)$) is the same as C (resp. D) with every concept name A occuring in C (resp. D) replaced by dA (resp. rA).

Next we define the domain/range TBox of $\mathcal{T}$, which is essentially the union of three sets of axioms: (1) the domain/range description of a role R w.r.t. $\mathcal{T}$, (2) a set of axioms that maintains the semantics of the concepts when acting as domain of R and (3) another set of axioms when acting as range of R. This TBox should be seen as describing the role in terms of the relations between its domain and range.

Definition 4 (Domain/Range TBox). *Let $\mathcal{T}$ be a TBox and R a role. The* domain/range TBox *of $\mathcal{T}$ w.r.t. R, written $\mathcal{T}^R$, is defined as:*

$$\mathcal{T}^R = \delta(\mathcal{T}) \cup \rho(\mathcal{T}) \cup \mathrm{dr}_{\mathcal{T}}(R) \ .$$

In definition 4, the terms $\delta(\mathcal{T})$ and $\rho(\mathcal{T})$ ensure that the concepts' semantics are maintained when they act as domain and range of the role R. Finally, we define the role similarity measure, which is simply the employment of any of the previously defined measures to the domain/range TBoxes.

Definition 5 (Role Similarity). *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be ontologies, R a role and s a similarity measure, $s \in \{\gamma, \alpha_+, \alpha_-, \alpha, F\}$. The* role similarity measure *w.r.t. R under s, $\varrho_s^R : \mathcal{O} \times \mathcal{O} \to [0, 1]$, is defined as:*

$$\varrho_s^R(\mathcal{T}_1, \mathcal{T}_2) = s(\mathcal{T}_1^R, \mathcal{T}_2^R)$$

It is arguable that this solution is somewhat cumbersome, for three reasons:

1. the meaning of the prefixed concepts dA and rA is not clear;
2. it is necessary to define a domain/range TBox for each role, where the only variable is the domain/range description;
3. the computation of $\mathrm{dr}_{\mathcal{T}}(R)$ requires a full prover.

A more elegant solution would be to effectively reify the relations. This could be achieved by extending the formalism with a rule mechanism (e.g., $\mathcal{AL}$-log). In this case, we could simply add a rule base to the original TBoxes to relate the domain and range of any role. For example, the following rule could belong to such a rule base:

$$\mathrm{range}(\mathsf{marriedTo}, \mathsf{Woman}) \leftarrow \mathrm{domain}(\mathsf{marriedTo}, \mathsf{Man}) \ .$$

However, this solution does not fit as well with the previously defined measures. Moreover, it would be necessary to deal with a hybrid DL and rule-based formalism. In effect, we would be shifting from one problem to another.

Let us illustrate all measures with a more complex example. Consider the following ontologies, $\mathcal{T}_1$ and $\mathcal{T}_2$. Table 2 shows $Z(\mathcal{T}_1)$ and $Z(\mathcal{T}_2)$ (written Z_1 and Z_2).

[1] Note that $\exists R.D \sqsubseteq C$ is equivalent to $\neg C \sqsubseteq \forall R.\neg D$, so this definition captures domain as well as range restrictions.

$\mathcal{T}_1$	$\mathcal{T}_2$
$\neg\mathsf{Male} \sqsubseteq \mathsf{Female}$	$\mathsf{Female} \doteq \neg\mathsf{Male}$
$\mathsf{Man} \doteq \mathsf{Person} \sqcap \mathsf{Male}$	$\mathsf{Man} \doteq \mathsf{Person} \sqcap \mathsf{Male}$
$\mathsf{Woman} \doteq \mathsf{Person} \sqcap \mathsf{Female}$	$\mathsf{Woman} \doteq \mathsf{Person} \sqcap \neg\mathsf{Man}$
$\mathsf{MaleCat} \doteq \mathsf{Cat} \sqcap \mathsf{Male}$	$\mathsf{MaleCat} \sqsubseteq \mathsf{Cat} \sqcap \mathsf{Male}$
$\top \sqsubseteq \forall\mathsf{marriedTo}.\mathsf{Person}$	$\top \sqsubseteq \forall\mathsf{marriedTo}.\mathsf{Person}$
$\neg\mathsf{Person} \sqsubseteq \forall\mathsf{marriedTo}.\bot$	$\neg\mathsf{Person} \sqsubseteq \forall\mathsf{marriedTo}.\bot$
	$\mathsf{Man} \sqsubseteq \forall\mathsf{marriedTo}.\mathsf{Woman}$
	$\mathsf{Woman} \sqsubseteq \forall\mathsf{marriedTo}.\mathsf{Man}$

Table 2. The characteristic acceptance sets for $\mathcal{T}_1$ and $\mathcal{T}_2$.

	$Z_1 - Z_2$				$Z_1 \cap Z_2$								$Z_2 - Z_1$	
Person	+	−	−	+	+	+	+	+	−	−	−	−	+	−
Man	+	−	−	+	+	+	−	−	−	−	−	−	+	−
Male	+	+	+	+	+	+	−	−	−	−	+	+	+	+
Woman	+	−	−	+	−	−	+	+	−	−	−	−	−	−
Female	+	+	+	+	−	−	+	+	+	+	−	−	−	−
Cat	+	+	−	−	+	−	−	+	−	+	−	+	+	+
MaleCat	+	+	−	−	+	−	−	−	−	−	−	+	−	−

Z_1

Z_2

Table 3 shows the results of applying the similarity measures to the example.[2] We observe that one of the main differences between $\mathcal{T}_1$ and $\mathcal{T}_2$ is that in the former every individual is required to be either Male or Female or both, but in the latter they have to be either one or the other exclusively. If this were the only difference, then $Z(\mathcal{T}_2)$ would be totally contained within $Z(\mathcal{T}_1)$, but the difference in the definition of Cats and MaleCats accounts for the two characteristic concepts in $Z(\mathcal{T}_2)$ and not in $Z(\mathcal{T}_1)$. The fact that more of $Z(\mathcal{T}_2)$ is contained within $Z(\mathcal{T}_1)$ than vice-versa is responsible for the lower coverage result between $\mathcal{T}_1$ and $\mathcal{T}_2$ than the inverse. Indeed, we can observe that more of what can be modeled in $\mathcal{T}_2$ can also be modeled in $\mathcal{T}_1$ than vice-versa, so these results are rather intuitive. The consistency agreement result comes from the number of agreed consistent characteristic concepts, 8, in proportion to the number of characteristic concepts in the table, 14. This result is also intuitive, since the ontologies have a reasonable overlap, i.e. most concepts are similar in both ontologies. On the other hand, the inconsistency agreement is the number of agreed inconsistent characteristic concepts, i.e., the ones that do not appear in table 2, $128 - 14 = 114$, in proportion to the number of inconsistent characteristic concepts in at least one of the ontologies, $128 - 8 = 120$. This shows that, although there is a considerable difference on what can be modeled in the ontologies, there is a substantially higher similarity concerning what can *not* be modeled in none of the ontologies. The result obtained for the agreement measure is a consequence of the fact that $8+114 = 122$ characteristic concepts, out of 128, have the same consistency in both TBoxes. It is slightly higher than the inconsistency agreement, since it also takes into account the consistency agreement.

The results for role similarity show us that there is a low coverage between $\mathcal{T}_1$ and $\mathcal{T}_2$, since $\mathcal{T}_1$ is much more permissible regarding the marriage relationship. In

[2] For the sake of readability, we refer to marriedTo as simply m.

Table 3. Results of applying the similarity measures to $\mathcal{T}_1$ and $\mathcal{T}_2$.

$\gamma(\mathcal{T}_1,\mathcal{T}_2)$	$\gamma(\mathcal{T}_2,\mathcal{T}_1)$	$\alpha_+(\mathcal{T}_1,\mathcal{T}_2)$	$\alpha_-(\mathcal{T}_1,\mathcal{T}_2)$	$\alpha(\mathcal{T}_1,\mathcal{T}_2)$	$F(\mathcal{T}_1,\mathcal{T}_2)$
66.67%	80.00%	57.14%	95.00%	95.31%	72.73%

$\varrho^{\mathsf{m}}_{\gamma}(\mathcal{T}_1,\mathcal{T}_2)$	$\varrho^{\mathsf{m}}_{\gamma}(\mathcal{T}_2,\mathcal{T}_1)$	$\varrho^{\mathsf{m}}_{\alpha_+}(\mathcal{T}_1,\mathcal{T}_2)$	$\varrho^{\mathsf{m}}_{\alpha_-}(\mathcal{T}_1,\mathcal{T}_2)$	$\varrho^{\mathsf{m}}_{\alpha}(\mathcal{T}_1,\mathcal{T}_2)$	$\varrho^{\mathsf{m}}_{F}(\mathcal{T}_1,\mathcal{T}_2)$
22.22%	66.67%	20.00%	99.80%	99.80%	33.33%

this TBox, both men and women can marry any person of any gender. In $\mathcal{T}_2$, men can only marry women and vice-versa. On the other hand, many of the possible marriage relations in $\mathcal{T}_2$ are covered in $\mathcal{T}_1$, which accounts for the higher coverage between $\mathcal{T}_2$ and $\mathcal{T}_1$. What keeps it from being 100% is the fact that women and men are exclusively female and male in $\mathcal{T}_2$, respectively, while in $\mathcal{T}_1$, men and women can have both genders, so the characteristic concepts subsumed by $\mathsf{dMan} \sqcap \mathsf{dWoman}$ or $\mathsf{rMan} \sqcap \mathsf{rWoman}$ are inconsistent in $\mathcal{T}_2^{\mathsf{marriedTo}}$ but consistent in $\mathcal{T}_1^{\mathsf{marriedTo}}$ (the domain/range TBoxes w.r.t. $\mathsf{marriedTo}$).

4 Experimental Results

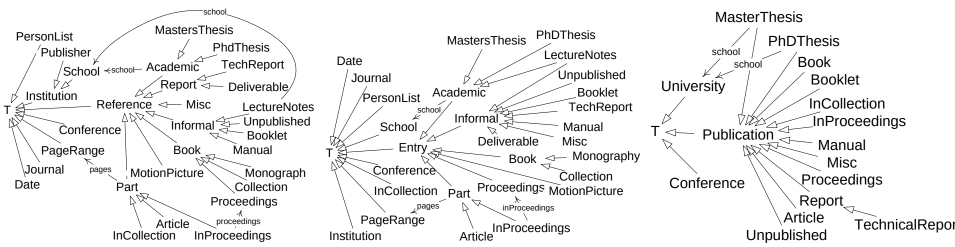
Evaluating similarity measures is a difficult task since it requires a standard definition of similarity, which is highly subjective. Therefore, it is hard to use common metrics such as precision and recall. In the following, we present and discuss two experiments conducted applying the similarity measures to two datasets: BibTeX and Directory.

BibTeX The first experiment used the BibTeX dataset extracted from OAEI.[3] We used a sub-set of three of those ontologies: the reference, $\mathcal{B}_{\mathsf{ref}}$, and the ones from Karlsruhe, $\mathcal{B}_{\mathsf{k}}$, and INRIA, $\mathcal{B}_{\mathsf{i}}$. Each ontology presents a different perspective on what bibliographic references are. All ontologies are formalized in OWL. Since the measures we propose are only applicable to ontologies with the same set of concepts, we need to align them. The ontologies are bundled with a set of alignments, between the reference ontology and the other two, which we use to merge them. These alignments contain not only equalities but also inclusion alignment relations. We discarded the latter since we focus on what is equal on both ontologies. Figure 1 shows a relevant section of the ontologies. Most concept names have the same (or similar) label in all ontologies, except $\mathsf{Reference}$, which is $\mathsf{Publication}$ and Entry in $\mathcal{B}_{\mathsf{k}}$ and $\mathcal{B}_{\mathsf{i}}$ respectively, and School, labeled as $\mathsf{University}$ in $\mathcal{B}_{\mathsf{k}}$.

In this experiment we compared the reference ontology to the other two. For that, we focus only on the concept names that are present in both ontologies (the reference one and the other), by trimming out the ones which are not involved in the alignment. Table 4 shows the results of applying the similarity measures to the ontologies.[4]

Coverage results show that $\mathcal{B}_{\mathsf{ref}} \models \mathcal{B}_{\mathsf{k}}$. In fact, we can observe that, if we focus on the concept names that appear in both ontologies, every axiom in $\mathcal{B}_{\mathsf{k}}$ is either

[3] `http://oaei.ontologymatching.org/2006/`.
[4] We refer to the school role as s.



Reference INRIA Karlsruhe

Fig. 1. BibTeX ontologies.

Table 4. Results of applying the measures to the BibTeX ontologies.

$\gamma(\mathcal{B}_{\mathsf{ref}}, \mathcal{B}_{\mathsf{k}})$	$\gamma(\mathcal{B}_{\mathsf{k}}, \mathcal{B}_{\mathsf{ref}})$	$\alpha_+(\mathcal{B}_{\mathsf{ref}}, \mathcal{B}_{\mathsf{k}})$	$\alpha_-(\mathcal{B}_{\mathsf{ref}}, \mathcal{B}_{\mathsf{k}})$	$\alpha(\mathcal{B}_{\mathsf{ref}}, \mathcal{B}_{\mathsf{k}})$	$F(\mathcal{B}_{\mathsf{ref}}, \mathcal{B}_{\mathsf{k}})$
100%	75.00%	75.00%	86.96%	90.63%	85.71%
$\varrho^{\mathsf{s}}_{\gamma}(\mathcal{B}_{\mathsf{ref}}, \mathcal{B}_{\mathsf{k}})$	$\varrho^{\mathsf{s}}_{\gamma}(\mathcal{B}_{\mathsf{k}}, \mathcal{B}_{\mathsf{ref}})$	$\varrho^{\mathsf{s}}_{\alpha_+}(\mathcal{B}_{\mathsf{ref}}, \mathcal{B}_{\mathsf{k}})$	$\varrho^{\mathsf{s}}_{\alpha_-}(\mathcal{B}_{\mathsf{ref}}, \mathcal{B}_{\mathsf{k}})$	$\varrho^{\mathsf{s}}_{\alpha}(\mathcal{B}_{\mathsf{ref}}, \mathcal{B}_{\mathsf{k}})$	$\varrho^{\mathsf{s}}_{F}(\mathcal{B}_{\mathsf{ref}}, \mathcal{B}_{\mathsf{k}})$
100%	44.99%	44.99%	94.96%	95.16%	62.06%
$\gamma(\mathcal{B}_{\mathsf{ref}}, \mathcal{B}_{\mathsf{i}})$	$\gamma(\mathcal{B}_{\mathsf{i}}, \mathcal{B}_{\mathsf{ref}})$	$\alpha_+(\mathcal{B}_{\mathsf{ref}}, \mathcal{B}_{\mathsf{i}})$	$\alpha_-(\mathcal{B}_{\mathsf{ref}}, \mathcal{B}_{\mathsf{i}})$	$\alpha(\mathcal{B}_{\mathsf{ref}}, \mathcal{B}_{\mathsf{i}})$	$F(\mathcal{B}_{\mathsf{ref}}, \mathcal{B}_{\mathsf{i}})$
85.44%	50.62%	46.60%	96.69%	96.79%	63.58%
$\varrho^{\mathsf{s}}_{\gamma}(\mathcal{B}_{\mathsf{ref}}, \mathcal{B}_{\mathsf{i}})$	$\varrho^{\mathsf{s}}_{\gamma}(\mathcal{B}_{\mathsf{i}}, \mathcal{B}_{\mathsf{ref}})$	$\varrho^{\mathsf{s}}_{\alpha_+}(\mathcal{B}_{\mathsf{ref}}, \mathcal{B}_{\mathsf{i}})$	$\varrho^{\mathsf{s}}_{\alpha_-}(\mathcal{B}_{\mathsf{ref}}, \mathcal{B}_{\mathsf{i}})$	$\varrho^{\mathsf{s}}_{\alpha}(\mathcal{B}_{\mathsf{ref}}, \mathcal{B}_{\mathsf{i}})$	$\varrho^{\mathsf{s}}_{F}(\mathcal{B}_{\mathsf{ref}}, \mathcal{B}_{\mathsf{i}})$
72.51%	20.50%	19.02%	99.88%	99.88%	31.97%

also present in $\mathcal{B}_{\mathsf{ref}}$ or is a logical consequence of it. The 75% result in the coverage between $\mathcal{B}_{\mathsf{k}}$ and $\mathcal{B}_{\mathsf{ref}}$ can be explained by the fact that $\mathsf{Proceedings}$ is a subclass of Book in the latter, but not in the former. Therefore, the set of characteristic concepts consistent in $\mathcal{B}_{\mathsf{k}}$ but not in $\mathcal{B}_{\mathsf{ref}}$ are the ones that contain $\mathsf{Proceedings} \sqcap \neg \mathsf{Book}$. These are also the only characteristic concepts which are not agreed by the ontologies, which explains the very high consistency agreement.

Regarding the similarity of the role school, although it would seem that it has the same domain and range in $\mathcal{B}_{\mathsf{ref}}$ and $\mathcal{B}_{\mathsf{k}}$, and, therefore, it should be assessed as totally similar, it is not exactly that case. In $\mathcal{B}_{\mathsf{ref}}$, the domain and range of school are modeled as general domain and range restrictions:

$$\exists \mathsf{school}.\top \sqsubseteq \mathsf{Academic} \sqcup \mathsf{LectureNotes} ,$$
$$\top \sqsubseteq \forall \mathsf{school}.\mathsf{School} .$$

This means that there cannot be any instance of the school role that does not start in $\mathsf{Academic}$ or $\mathsf{LectureNotes}$ or end in School. In $\mathcal{B}_{\mathsf{k}}$, we have:

$$\mathsf{MasterThesis} \sqcup \mathsf{PhDThesis} \sqsubseteq \forall \mathsf{school}.\mathsf{University} .$$

This only restricts the range of school for instances in the domain of $\mathsf{MasterThesis}$ or $\mathsf{PhDThesis}$. Since $\mathsf{MastersThesis}$ and $\mathsf{PhDThesis}$ are subclasses of $\mathsf{Academic}$, the former two axioms entail the latter one, but the contrary does not hold. The fact that $\mathsf{LectureNotes}$ is in the domain of the role does not affect the similarity assessment since this concept is not present in $\mathcal{B}_{\mathsf{k}}$, and is, therefore, trimmed out. Although the INRIA ontology also uses the general domain and range restriction, we can observe that $\mathsf{LectureNotes}$ has different semantics.

We note that the reference ontology is assessed as more similar to the INRIA one than to the Karslruhe one regarding their inconsistency. This results from the fact that the more concept names an ontology has, the stricter it is likely to be, since every concept name is usually involved in some axiom, which restricts its characteristic acceptance set. The inconsistency agreement and the agreement measures suffer from this problem. However, for a set of ontologies sharing exactly the same concepts, which is not the case with this dataset, these measures would be helpful in determining which ones were more similar.

Directory We also took the Directory ontologies from the OAEI dataset. These ontologies were extracted from the Google, Yahoo and Looksmart web directories. The source ontology comprises over 2,000 concept names while the target ontology more than 6,000. Both ontologies contain only inclusion statements (taxonomies) and are biased towards the task of easily finding a set of web sites. Since there was no reference alignment available, we performed one using a semi-automatic algorithm.

Due to space constraints, we only present the basic idea of the algorithm. It is composed of (1) an automatic phase and (2) a manual one. In (1) we use a derivation of the terminological overlap measure using a notion similar to the semantic cotopy [6]. For each concept name in each ontology, its cotopy is extracted from the ontology and augmented with its WordNet cotopy. The algorithm chooses the pairs of concept names that maximize the terminological overlap of their cotopies. Phase (2) is composed of two steps: filtering out wrong mappings and adding correct ones. The final result is a set of 580 concept mappings. Since there is no reference mapping available, we estimate the precision of the automatic phase of the algorithm w.r.t. our notion of correct or incorrect mappings: 51.56%.

We compared the ontologies w.r.t. the set of concept names involved in the extracted mapping. Because of the previously identified problem concerning the relation between the number of concept names and the inconsistency agreement, this measure, along with the agreement one, assessed the ontologies as almost 100% similar. In contrast, the ontologies were given a very low similarity assessment, close to 0%, w.r.t. coverage and consistency agreement. A closer inspection revealed that out of the 585 axioms present in each of the ontologies, 150 of them differed, revealing that at least such amount of concept names assume a different position in the taxonomy. If these differences are in higher-level concept names, that can have a deep impact on the similarity assessment. This leads us to conclude that either (1) the mapping is inaccurate or (2) the ontologies are indeed very different. Since the manual filter was introduced to improve the precision of the mapping, the first hypothesis is not very plausible.

5 Discussion

The results for the BibTeX ontologies show that the proposed measures bear relevant information concerning the similarity of the ontologies. Moreover, the different measures are to some extent independent, which is an important feature of similarity measures identified in [6]. From the results we can conclude that the measures are more trustworthy when applied to ontologies with the same set of concept names (as opposed to the BibTeX dataset, where $\mathcal{B}_{\mathsf{ref}}$ shares less concept names with $\mathcal{B}_{\mathsf{k}}$ than with $\mathcal{B}_{\mathsf{i}}$). This comes from the fact that the more concept names the ontologies share, the more likely there are differences in them.

One of the problems we clearly identify is the growth of the inconsistency agreement in proportion to the number of concept names. Therefore, this measure, along with the agreement one, is only useful for ontologies with few concept names or when comparing a set of (three or more) ontologies with exactly the same set of concept names. In the BibTeX results presented, the set of concept names differed among the three ontologies, so the results of comparing $\mathcal{B}_{\mathsf{ref}}$ to $\mathcal{B}_{\mathsf{i}}$ are not directly comparable to the similarity results between $\mathcal{B}_{\mathsf{ref}}$ and $\mathcal{B}_{\mathsf{k}}$, because $\mathcal{B}_{\mathsf{i}}$ and $\mathcal{B}_{\mathsf{k}}$ don't share the same concept names. If all ontologies shared the same set of concept names, and even if the results of the inconsistency agreement are very high, they are in the same scale, and can thus be compared with each other.

Given that our measures require a previous mapping, instead of comparing characteristic concepts, an alternative approach would be to compare their set of consequences. Though countable, we should stress that this set is infinite. Nonetheless, the acceptance set can be seen as a finite set of consequences.

Contrary to related work, the concept similarity measures focus on the ontologies as wholes, which may hinder their usefulness. Local similarity measures are often used to support the matching process of individual concepts. Therefore, the concept similarity measures presented here are not directly useful in this task. However, they can be used to determine the *global a posteriori* quality of a given mapping.

We should also note that the concept similarity is not affected by roles. These measures assess the following TBoxes as equal:

$$\mathcal{T}_1 = \{\mathsf{A} \sqsubseteq \forall R.A\},\ \mathcal{T}_2 = \{\mathsf{A} \sqsubseteq \forall R.\neg A\}\ ,$$

although the role similarity measures do not. This is naturally undesirable and should be revised in the future. Since the role similarity measures focus on each role individually, they can be used to perform role matching given a previous mapping of concepts.

The question of which measure is the most useful is only sensible in view of the purpose it is used for. If we are comparing a learnt ontology against a gold-standard, as discussed earlier, the coverage operator corresponds to precision and recall, and can therefore be used as such. If we are only concerned with what can be represented with the ontologies, then the agreement operator is a good indication of similarity. Naturally, the different operators can be combined within a unique operator that suits a particular purpose, but this is outside the scope of this work.

Regarding efficiency, although $|\zeta(\mathcal{C})|$ is exponential to $|\mathcal{C}|$, it can be shown that #SAT solvers, used to compute the size of acceptance sets, perform well in practice [2]. Finally, although it is not shown here, the measures satisfy the set of criteria identified in [6].

6 Conclusions and Future Work

In this paper we proposed a set of similarity measures to assess the similarity of ontologies based on the notion of characteristic concept. We applied them in two experiments using real ontologies yielding intuitive and very promising results.

Extending the measures to more expressive DLs is one of the directions for future work. Further developments should tackle the problem regarding the weak relation between the concept and role similarities. Other future developments include integration of the measures in a system, for example to perform (semi-)automatic evaluation of learnt ontologies against a gold-standard. It should also be interesting to research into the usage of characteristic concepts for the assessment of the similarity of concept names instead of ontologies. This could then be used to create a novel matching algorithm.

References

1. R. Araújo and H. S. Pinto. SEMilarity theory. Technical report, INESC-ID, 2007. `http://algos.inesc-id.pt/~rnna/semilarity-theory.pdf`.
2. R. Araújo and H. S. Pinto. SEMilarity: Towards a model-driven approach to similarity. In *Proc. of DL2007*, volume 250 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2007.
3. F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, and P. F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2003.
4. A. Borgida, T. Walsh, and H. Hirsh. Towards measuring similarity in description logics. In *Proc. of DL2005*, volume 147 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2005.
5. C. d'Amato, N. Fanizzi, and F. Esposito. A dissimilarity measure for ALC concept descriptions. In *Proc. of ACM SAC 2006*, pages 1695–1699. ACM, 2006.
6. K. Dellschaft and S. Staab. On how to perform a gold standard based evaluation of ontology learning. In *Proc. of ISWC2006*, volume 4273 of *LNCS*, pages 228–241. Springer, 2006.
7. B. Hu, Y. Kalfoglou, H. Alani, D. Dupplaw, P. H. Lewis, and N. Shadbolt. Semantic metrics. In *Proc. of EKAW2006*, volume 4248 of *LNCS*, pages 166–181. Springer, 2006.
8. K. Janowicz. Sim-DL: Towards a semantic similarity measurement theory for the description logic ALCNR in geographic information retrieval. In *On the Move to Meaningful Internet Systems. Proc., Part II*, volume 4278 of *LNCS*, pages 1681–1692. Springer, 2006.
9. D. McGuinness, R. Fikes, J. Rice, and S. Wilder. An environment for merging and testing large ontologies. In *Proc. of KR2000*, pages 483–493. Morgan Kaufmann, 2000.
10. N. F. Noy and M. A. Musen. Prompt: Algorithm and tool for automated ontology merging and alignment. In *Proc. of AAAI/IAAI 2000*, pages 450–455. AAAI Press / The MIT Press, 2000.
11. V. A. M. Tamma, P. R. S. Visser, D. Malerba, and D. M. Jones. Computer assisted ontology clustering for knowledge sharing. In *Proc. of ECML'2000/ML net Workshop on Machine Learning in the New Information Age*, pages 75–83, 2000.