Formally Measuring Agreement and Disagreement in Ontologies

Mathieu D'aquin

Formally Measuring Agreement and Disagreement in Ontologies*

Mathieu d'Aquin
Knowledge Media Institute, The Open University
Milton Keynes, MK7 6AA, UK
m.daquin@open.ac.uk

ABSTRACT

Ontologies are conceptual models of particular domains, and domains can be modeled differently, representing different opinions, beliefs or perspectives. In other terms, ontologies may disagree with some particular pieces of information and among themselves. Assessing such agreements and disagreements is very useful in a variety of scenarios, in particular when integrating external elements of information into existing ones. In this paper, we present a set of measures to evaluate the agreement and disagreement of an ontology with a statement or with other ontologies. While our work goes beyond the naive approach of checking for logical inconsistencies, it relies on a complete formal framework based on the semantics of the considered ontologies. The experiments realized on several concrete scenarios show the validity of our approach and the usefulness of measuring agreement and disagreement in ontologies.

Categories and Subject Descriptors

I.2.4 [ARTIFICIAL INTELLIGENCE]: Knowledge Representation Formalisms and Methods

General Terms

Measurement, Experimentation

Keywords

Ontologies, Agreement, Disagreement, Consensus, Controversy

1. INTRODUCTION

Ontologies are knowledge artifacts representing particular models of some particular domains. They are built within the communities that rely on them, meaning that they represent consensual representations inside these communities. However, several ontologies can cover the same domain, while being built by and for different communities. In these cases, different perspectives, points of view and opinions might be expressed on the same concepts and objects of the domain, i.e. ontologies might disagree.

Knowing to which extent an ontology agrees with another one, or with a particular statement, can be very useful in many scenarios. When using several ontologies in a common application, for example, disagreement can create unexpected results for the user. Also, when reusing statements from other ontologies while building a new one, there is a need for some formal measures to indicate whether or not there exists a general agreement concerning a candidate statement, in the current ontology or in other, external ontologies. Realizing that there is a high level of disagreement in available ontologies with a given statement, or that there is no clear cut between agreement and disagreement, can be very useful in pointing out potentially problematic statements or ontologies.

One way to detect whether there is a disagreement between two ontologies is to rely on the presence of logical contradictions. The two ontologies can be merged, based on mappings between their entities, and the resulting model be checked for inconsistencies and incoherences (the same approach can be used to check if an ontology disagrees with a statement). While this approach would certainly detect some forms of disagreement, it suffers from a number of limitations. First, it only checks whether the ontologies disagree or not. It does not provide any granular notion of disagreement and, if no contradictions are detected, it does not necessary means that the ontologies agree. Indeed, while two ontologies about two completely different, non overlapping domains would certainly not disagree, they do not agree either. More importantly, logical contradictions are not the only way for two ontologies to disagree. Indeed, there could also be conceptual mismatches, like in the case where one ontology declares that "Lion is a subclass of Species" and the other one indicates that "Lion is an instance of Species". Even at content level, logical contradictions would not detect some form of

*This work was supported by the European Commission in the IST project NeOn (IST-2006-027595).

K-CAP'09, September 1–4, 2009, Redondo Beach, California, USA.

disagreements. Indeed, the two statements “Human is a subclass of Animal” and “Animal is a subclass of Human” do not generate any incoherence. However, they disagree in the sense that, if put together, they generate results that were not expected from any of the two ontologies.

This other type of disagreement could be checked by using the notion of *conservative extension* [4]. Informally, a conservative extension of an ontology is an ontology where axioms have been added which do not ‘affect’ the relations that can be inferred about the entities of the original ontology. Therefore, merging two ontologies and checking whether the result is a conservative extension of both the original ontologies would give indications of some forms of disagreement. However, this would also only result in a ‘black box test’, not providing a granular notion of disagreement or a proper notion of agreement.

For these reasons, we developed a set of measures to assess both the level of agreement and the level of disagreement between a statement and an ontology or a set of ontologies, as well as between two ontologies. While these measures provide quantitative evaluations for agreement and disagreement, they rely on the formal properties of the ontologies and on the semantics of the ontology language to compare what is expressed by different statements and ontologies, detecting logical contradictions as well as other forms of contradictions.

2. ASSUMPTIONS AND REQUIREMENTS

Based on the usage scenarios quickly sketched above and on the analysis of the limitations of the basic approaches relying on logical contradictions and conservative extensions, we devised an number of requirements that our measures should fulfill:

R1: Ontologies agree with themselves. This is a fairly obvious starting point, meaning that our measures should provide maximal results for agreement (and no disagreement) when checking an ontology against itself. It also means that we assume the ontologies we assess to be consistent, coherent, and homogeneous in terms of modeling.

R2: Covering different domains is not agreeing. We consider that there is a fundamental difference between agreeing and not talking about the same things. This means that, when applied on ontologies or statements that cover completely different domains (i.e. for which there is no mapping between the entities), our measures should indicate that there is no disagreement and no agreement.

R3: Different levels of agreement/disagreement. As already mentioned, there are different reasons for which an ontology might disagree with another one (at the level of the content, logically or not, at the level of the representation, etc.) These different kinds of disagreements should be considered with different levels of importance, leading to granular measures rather than binary tests.

R4: Independent from matching techniques. In order the check the agreement/disagreement of an ontology with respect to another ontology or a statement, corresponding entities should be related through mappings. However, the way these mappings are produced should not affect the computation of the agreement/disagreement measures. Therefore, we will assume the existence of mappings we can exploit, without considering any particular approach to produce such mappings. In our experiments, we will use a simple technique based on the lexical correspondence between the entities’ names.

R5: It is possible to agree and disagree at the same time. Two ontologies may agree on some statements and disagree on some others, meaning that, depending on the number of statements in each category and on the level of agreement/disagreement for each statement, the two ontologies might agree and disagree at different levels. It should even be possible for an ontology to both agree and disagree with one single statement. Indeed, we consider the case of an ontology that contains the statement “Lion is a subclass of Species” and for which we want to check whether it agrees with the statement “Lion is an instance of Species”. As already mentioned, this ontology disagrees at a modeling level with this statement. However, at content level, it agrees to a certain extent that Lion is included in Species.

3. BASIC FRAMEWORK

In this section, we consider an ontology O to be defined as a set of statements (corresponding to axioms in the description logic terminology). A statement is the expression of a relation between two entities of the ontology in the form of a triple $< subject, relation, object >$. In RDF-based ontology languages, *relation* corresponds to the identifier of a property (its URI), *subject* corresponds either to the URI of an entity or to an anonymous resource, and *object* corresponds to the URI of an entity, an anonymous resource or a literal. In the following, we base our definition of agreement and disagreement on whether or not an ontology conflicts with a relation linking two named entities. Hence, we chose to ignore the cases of anonymous entities (i.e., blank nodes in RDF) and literals. This does not mean that our measures could not be applied to ontologies containing such elements, but simply that we will only use statements where the subjects and objects are identifiers of entities.

As already mentioned, the atomic element on which ontologies can express agreement or disagreement is the statement. Indeed, an ontology can contain information that contradict, or on the contrary enforce the one expressed through a particular statement, and so can agree or disagree with it. We define two elementary functions representing respectively the agreement and the disagreement levels of an ontology O with respect to a statement $s = < subject, relation, object >$:

$$agreement(O, s) \rightarrow [0..1]$$
$$disagreement(O, s) \rightarrow [0..1]$$

We chose to use two distinct measures for agreement and disagreement so that an ontology can, at the same time and to certain extents, agree and disagree with a statement (cf. R3, R5). These two measures have to be interpreted together to indicate the particular belief expressed by the ontology O regarding the statement s. For example, if $agreement(O, s) = 1$ and $disagreement(O, s) = 0$, it means that O fully agrees with s and conversely if $agreement(O, s) = 0$ and $disagreement(O, s) = 1$, it fully disagrees with s. Now, agreement and disagreement can vary between 0 and 1, meaning that O can only partially agree or disagree with s (cf. R3) and sometimes both, when $agreement(O, s) > 0$ and $disagreement(O, s) > 0$ (cf. R5). Finally, another case is when $agreement(O, s) = 0$ and $disagreement(O, s) = 0$. This basically means that O neither agrees nor disagrees with s, for the reason that it does not express any belief regarding the relation encoded by s (cf. R2).

The actual values returned for both measures, when different from 0 and 1, are not very important. They correspond to different levels of dis/agreement and only an order between pre-defined levels is needed to interpret them. In this work, we consider two different levels of agreement and of disagreement different from 0 and 1, defined as follows:

$$0 < A1 < A2 < 1$$
$$0 < D2 < D1 < 1$$

Note again that the actual values for these levels are not significant, only the total order between them. However, to facilitate comparisons of measures, desirable properties are that $A1 + A2 = 1$, $D1 + D2 = 1$, $A1 + D1 = 1$ and $A2 + D2 = 1$.

To be able to compute a (non-null) level of dis/agreement, it is necessary that the ontology O covers the entities related by the statement s. Thus, we need to identify entities in O that correspond to the entities in the tested statement s. For this, we assume the existence of a function $match(O, e)$ that returns the identifier of an entity e' in O (cf. R4). To simplify the notation, we will call s' and o' the matching entities in an ontology O of the subject and object of a statement $< s, r, o >$ respectively.

Note that the *match* function does not need to be completely defined, meaning that it can happen that no matching entities are found in O for s or o. However, the following definitions will assume the existence of s' and o'. Hence we consider that $agreement(O, t) = 0$ and $disagreement(O, t) = 0$, with $t = < s, r, o >$, iff $match(O, s)$ or $match(O, o)$ are undefined.

4. MEASURING AGREEMENT AND DISAGREEMENT OF AN ONTOLOGY REGARDING A STATEMENT

Based on the basic definitions of the previous section, we can now devise a formal way to compute the agreement and disagreement measures between statements and ontologies. As mentioned before, these measures should essentially assess whether or not the relation between entities expressed in the considered statement s is 'validated' by the ontology. The central element to consider for our measures is therefore the type of the relation r in the considered statement $< s, r, o >$. This relation can either be a property included in the ontology representation language, i.e., one of $\{subClassOf, equivalentClass, domain, range, disjointWith, type, sameAs, differentFrom, subPropertyOf\}$, or a generic, user-defined property R (e.g., $isPartOf$, $isAuthorOf$).

To assess dis/agreement between this statement and the ontology O, we then need to compare the relation encoded in the statement, and the relations that are expressed between corresponding entities in O. In other terms, we need to extract from O statements that express relations between s' and o', the entities matching s and o. Note that these statements should not necessarily be declared in O, but can also be entailed from O. This leads to the definition of the *R-module* of an ontology O with respect to a statement $< s, r, o >$.

The **R-module** of an ontology O with respect to a statement $< s, r, o >$ is the list of statements of the form $< s', ?, o' >$ or $< o', ?, s' >$ that are entailed by O. Formally, it is defined by $RM(O, < s, r, o >) = \{st = < s', r', o' > \ or \ st = < o', r', s' > \mid match(O, s) = s' \ \wedge \ match(O, o) = o' \ \wedge \ O \models st\}$. Note that RM can be empty, meaning that the ontology does not express any relation between the two given entities.

The R-module represents a complete set of statements that O directly or indirectly expresses about the relations between s and o. In order to make it a minimal, concise summary of these relations, we introduce the notion of *non-redundant R-module*.

The **non-redundant R-module** of an ontology O with respect to a statement $< s, r, o >$ is a subset of the corresponding R-module such that none of the statements included can be inferred from other statements in the set. Formally, it is defined by $MRM(O, < s, r, o >) = \{st = < s', r', o' > \ or \ st = < o', r', s' > \mid st \in RM(O, < s, r, o >) \ \wedge \ RM(O, < s, r, o >) \setminus \{st\} \not\models st\}$. Note that, since we exclude statements involving the meta-model of the language, a non-empty R-module cannot lead to an empty non-redundant R-module.

In addition, according to requirement R1, we assume that $MRM(O, < s, r, o >)$ is a consistent and coherent set of axioms, so that it cannot contain for example both statements $< s, subClassOf, o >$ and $< s, disjointWith, o >$. Also, $MRM(O, < s, r, o >)$ should not contain any modeling conflict, meaning for example that it should not imply that an entity is at the same time a class and an individual, like in the two statements $< s, subClassOf, o >$ and $< s, type, o >$.

As for the relation in the considered statement, the MRM of an ontology can contain any kind of relation, in particular properties of the language and user-defined relations. In the later case, we need to distinguish the cases of relations that match the one described in the statement (i.e., when $match(O, r) = r'$). In order to simplify the notations, we represent the set of statements in a MRM as a set of relations between s' and o'. In this notation, a statement such as $< s', subClassOf, o' >$ is represented by the relation $subClassOf$ and the statement $< o', subClassOf, s' >$ by the relation $subClassOf^{-1}$. Concerning user-defined properties, a statement $< s', r', o' >$ is represented by $matchRelation$ if $match(O, r) = r'$, r being the relation in the considered statement, and by $mismatchRelation$ if not.

One of the interesting properties of the set MRM in this notation is that, due to the fact that it is non-redundant, consistent and coherent, and that it does not contain conflicts at modeling level, there exists only a finite set of possible $MRMs$ whatever are the considered ontology and the considered statement. In other terms, there exist only a restricted number of allowed combinations of relations, as MRM sets such as $\{subClassOf, subClassOf^{-1}, equivalentClass\}$, $\{subClassOf, disjointWith\}$ or $\{subClassOf, type\}$ do not comply with the definition.

Finally, assessing the levels of agreement and disagreement between a statement and an ontology O consists in comparing the relation r expressed in the statement, and the corresponding MRM from O, to check whether they enforce or contradict each other. Considering that there are only a finite set of possible situations, the result of the agreement (respct. disagreement) measure can be completely defined by a matrix establishing the returned values depending on the relation r and on the set MRM. The matrixes we propose for our measures are shown in Table 1, but can be customized for specific needs. The methodology used to build these matrixes and the impact of using different parameters are out of the scope of this paper.

As an example, if we consider the statement $st =< Person, sameAs, Human >$ and as MRM for the ontology O the set $\{subClassOf\}$, we obtain from these matrixes the following values for the measures of agreement and disagreement:

$$agreement(st, O) = A1$$
$$disagreement(st, O) = D1$$

indicating that st and O both partially agree (because both believe there is an overlap between $Human$ and $Person$) and partially disagree (because they do not express the same level of overlap, and they use different modeling, considering classes in one case and instances in the other case).

5. MEASURING AGREEMENT AND DISAGREEMENT BETWEEN TWO ONTOLOGIES

Considering that ontologies are made of statements, extending the measures above to compute agreement and disagreement between two ontologies should be relatively straightforward, using the average of each measure for each statement of an ontology against the other ontology. However, there are two issues related to this approach. First, the measures would not be symmetric, as declared statements would be considered in one case, and entailed ones in the other case. Second, in cases where the ontologies only have a small overlap, the results would be lowered down by all the statements which would not find matches from one ontology to the other (resulting in null values for both agreement and disagreement).

For these reasons, we define the measures of agreement and disagreement between ontologies in the following way:

$$agreement(O_1, O_2) = \frac{\sum_{st \in ST_1} agreement(st, O_2) + \sum_{st \in ST_2} agreement(st, O_1)}{|ST_1| + |ST_2|}$$

$$disagreement(O_1, O_2) = \frac{\sum_{st \in ST_1} disagreement(st, O_2) + \sum_{st \in ST_2} disagreement(st, O_1)}{|ST_1| + |ST_2|}$$

where $ST_1 = \{< s, p, o > \in O_1 \mid match(s, O_2) \neq \emptyset \wedge match(o, O_2) \neq \emptyset\}$ and $ST_2 = \{< s, p, o > \in O_2 \mid match(s, O_1) \neq \emptyset \wedge match(o, O_1) \neq \emptyset\}$. In this way, only relevant statements are checked for agreement and disagreement (ST_1 and ST_2) and the measures are symmetric.

6. MEASURING CONSENSUS AND CONTROVERSIES

In some scenarios, it is useful to have measures on how much a statement is dis/agreed in a set of available ontologies, i.e., to which extent there is a consensus over this statement in these ontologies. Conversely, a related information concerns the level of controversy on the statement, i.e. whether there is a clear cut between agreement and disagreement.

To compute such measures, we need to assess of the global dis/agreement in the considered set of ontologies with a statement st. We represent the available ontologies as a set R of ontologies, which in practice can be implemented as an ontology repository. In our experiments we will use the set of ontologies collected by the

r	sClass	equiv.	dom.	range	disj.	type	same	diff.	sProp.	R
$\emptyset$	0	0	0	0	0	0	0	0	0	0
sClass	1	A2	0	0	0	A2	A1	0	A1	0
sClass^{-1}	A1	A2	0	0	0	0	A1	0	0	0
equiv.	A2	1	0	0	0	A1	A2	0	A1	0
disj.	0	0	0	0	1	0	0	A2	0	0
dom.	0	0	1	0	0	0	0	0	0	0
dom.$^{-1}$	0	0	0	0	0	0	0	0	0	0
range	0	0	0	1	0	0	0	0	0	0
range^{-1}	0	0	0	0	0	0	0	0	0	0
dom./range	0	0	1	1	0	0	0	0	0	0
dom.$^{-1}$/range^{-1}	0	0	0	0	0	0	0	0	0	0
sProp	A2	A1	0	0	0	A1	A1	0	1	0
sProp^{-1}	0	A1	0	0	0	0	A1	0	A1	0
type	A2	A1	0	0	0	1	A1	0	A1	0
type^{-1}	0	A1	0	0	0	A1	A1	0	0	0
same	A1	A2	0	0	0	A1	1	0	A1	0
match	0	0	0	0	0	0	0	0	0	1
match^{-1}	0	0	0	0	0	0	0	0	0	0
mmatch	0	0	0	0	0	0	0	0	0	0
same/match	A1	A2	0	0	0	A1	1	0	A1	1
same/match^{-1}	A1	A2	0	0	0	A1	1	0	A1	0
same/mmatch	A1	A2	0	0	0	A1	1	0	A1	0
diff.	0	0	0	0	A2	0	0	1	0	0
diff./match	0	0	0	0	A2	0	0	1	0	1
diff./ match^{-1}	0	0	0	0	A2	0	0	1	0	0
diff.mmatch	0	0	0	0	A2	0	0	1	0	0

r	sClass	equiv.	dom.	range	disj.	type	same	diff.	sProp.	R
$\emptyset$	D1	D1	D2	D2	D1	D1	D1	D1	D1	D2
sClass	0	D2	D1	D1	1	D2	D1	1	D2	D1
sClass^{-1}	D1	D2	D1	D1	1	D1	D2	1	D1	D1
equiv.	D2	0	D1	D1	1	D1	D2	1	D1	D1
disj.	1	1	D1	D1	0	1	1	D2	1	D1
dom.	D1	D1	0	0	D1	D1	D1	D1	D1	D1
dom.$^{-1}$	D1	D1	D1	D1	D1	D1	D1	D1	D1	D1
range	D1	D1	0	0	D1	D1	D1	D1	D1	D1
range^{-1}	D1	D1	D1	D1	D1	D1	D1	D1	D1	D1
dom./range	D1	D1	0	0	D1	D1	D1	D1	D1	D1
dom.$^{-1}$/range^{-1}	D1	D1	D1	D1	D1	D1	D1	D1	D1	D1
sProp	D2	D1	D1	D1	1	D1	D1	1	0	D1
sProp^{-1}	D1	D1	D1	D1	1	D1	D1	1	D1	D1
type	D2	D1	D1	D1	1	0	D1	1	D1	D1
type^{-1}	D1	D1	D1	D1	1	D1	D1	1	D1	D1
same	D1	D2	D1	D1	1	D1	0	1	D1	D1
match	D1	D1	D1	D1	D1	D1	0	0	D1	0
match^{-1}	D1	D1	D1	D1	D1	D1	0	0	D1	D1
mmatch	D1	D1	D1	D1	D1	D1	0	0	D1	D2
same/match	D1	D2	D1	D1	1	D1	0	1	D1	0
same/match^{-1}	D1	D2	D1	D1	1	D1	0	1	D1	D1
same/mmatch	D1	D2	D1	D1	1	D1	0	1	D1	D2
diff.	1	1	D1	D1	D2	1	1	0	1	D2
diff./match	1	1	D1	D1	D2	1	1	0	1	0
diff./match^{-1}	1	1	D1	D1	D2	1	1	0	1	D1
diff./mmatch	1	1	D1	D1	D2	1	1	0	1	D2

Table 1: Matrix definitions of the agreement (top) and disagreement (bottom) measures, depending on the relation r in the considered statement, and the set of relations in the MRM from the considered ontology O. R represents a user-defined property. Other relations are abbreviated.

Watson system[1]. We then compute the agreement and disagreement measures between a statement *st* and this repository R as a simple average, but taking only into account ontologies matching the entities of *st*:

$$agreement(st, R) = \frac{\sum_{O \in rR} agreement(st,O)}{|rR|}$$
$$disagreement(st, R) = \frac{\sum_{O \in rR} disagreement(st,O)}{|rR|}$$

where $st =< s, p, o >$ and $rR = \{O \in R \mid match(s, O) \neq \emptyset \wedge match(o, O) \neq \emptyset\}$.

The level of consensus concerning a statement *st* corresponds to the level of certainty on whether ontologies in R agree or disagree with *st*. We say that there is a high level of (positive consensus) if the overall agreement about this statement is high and the overall disagreement is low. Thus, we define the measure of consensus in a set of ontologies R upon a statement *st* as follows:

$$consensus(st, R) = agreement(st, R) - disagreement(st, R)$$

this measure should be interpreted in the following way:

- if $consensus(st, R) > 0$ then it represents the level of 'positive' consensus in R about *st*, meaning that there is a consensus on agreement.
- if $consensus(st, R) < 0$ then it represents the level of 'negative' consensus in R about *st*, meaning that there is a consensus on disagreement.

We consider the notion of controversy to be the inverse from the one of consensus: there is a high level of controversy on a given statement when there is no clear cut between agreement and disagreement, i.e. there is a low level of consensus. Therefore, the measure of controversy in a set of ontologies R upon a statement *st* can simply be computed in the following way:

$$controversy(st, R) = 1 - |consensus(st, R)|$$

7. USAGE SCENARIOS AND EXPERIMENTS

To test the behavior of our measures, we experiment on computing them on three concrete scenarios based on the Watson collection of ontologies. In these experiments, we use a simple lexical matching technique, and define $A1 = 0.25$, $A2 = 0.75$, $D2 = 0.25$ and $D1 = 0.75$.

7.1 Assessing Statements in the Watson Plugin

Watson [1] is a gateway to the Semantic Web. It collects, indexes and gives access to online knowledge available in ontologies and semantic documents. It is a Semantic Web search engine, but focusing on supporting applications in exploiting the Semantic Web through a set of high level APIs.

To help ontology engineers benefit from Watson, we developed the Watson Plugin for ontology editors [2]. This plugin allows the developer of an ontology to query Watson for entities corresponding to the ones in the built ontology, and to automatically integrate selected statements from these external entities into the ontology.

[1] `http://watson.kmi.open.ac.uk`

In this tool, it is essential to be able to assess statements, to identify the ones that can be safely integrated into the built ontology from the ones that require more attention. Here, we consider the case of an ontology engineer who needs to integrate to her ontology the class *SeaFood*. The Watson Plugin retrieves a set of statements for this class from other ontologies. We then compute the overall agreement and disagreement measures in Watson for each of these statements, as well as the consensus and controversy measures. The results are summarized in Table 2. Note that, in this table, 'lexically similar' statements are grouped together (the table indicates the number of ontologies where each statement appears).

As can be seen from these results, the 4 first statements are fully agreed with by ontologies in Watson, meaning that all the ontologies containing both entities of each statement express exactly the same relation as the one of the statement. The 3 next statements also have a very high level of agreement, and a very low level of disagreement. This is mainly due to a few ontologies containing the right entities, but not necessarily relating them. Hence, there is a high level of consensus on these statements. Finally the 2 last statements are the ones for which there is the highest level of controversy. The last one is by far the most disagreed with (which correlate with the high level of agreement of the other one contradicting it), clearly indicating that it should be considered carefully whether it can be integrated into the ontology being developed.

Another interesting example is the one of the statement $< river, subClassOf, sea >$, which gives a high level of disagreement (0.766). The disagreement is not 1 in that case, because only very few ontologies express explicitly contradicting relations. However, in this case, the level of agreement is 0: There is no ontology to actually agree with this statement.

7.2 Checking Automatically Discovered Relations

A similar scenario to the one presented above concerns automatically discovered relations between ontological entities, in particular in the task of ontology matching [3]. Indications about the level of dis/agreement concerning the derived relations can help filter out possibly problematic relations and so improve the precision of matching. For this to be achieved, we need to show that there are exploitable correlations between our measures and the 'correctness' of the considered relations.

For this experiment, we exploit the set of ontology map-

Statement	Nb1	Nb2	a	d	cs	ct
$< SeaFood, disjointWith, Dessert >$	12	12	1.0	0.0	1.0	0.0
$< Fowl, disjointWith, SeaFood >$	11	11	1.0	0.0	1.0	0.0
$< Pasta, disjointWith, SeaFood >$	12	12	1.0	0.0	1.0	0.0
$< SeaFood, subClassOf, EdibleThing >$	11	11	1.0	0.0	1.0	0.0
$< ShellFish, subClassOf, SeaFood >$	14	16	0.937	0.047	0.89	0.109
$< Fish, subClassOf, SeaFood >$	12	14	0.928	0.053	0.875	0.125
$< SeaFood, disjointWith, Fruit >$	11	14	0.85	0.1	0.75	0.25
$< Meat, disjointWith, SeaFood >$	8	16	0.75	0.22	0.53	0.46
$< SeaFood, subClassOf, Meat >$	4	16	0.125	0.844	-0.719	0.281

Table 2: Dis/agreement measures on statements from the Watson Plugin with the class $SeaFood$. $Nb1$ is a number of ontologies in which the statement appears and $Nb2$ the number of ontologies containing entities matching the subject and object of the statement. Measures of agreement (a), disagreement (d), consensus (cs) and controversy (ct) are computed on the Watson collection.

pings between 2 large thesaurus in the agriculture domain described in [5], which has been partially evaluated by 3 persons. We filter out mappings for which the global measures of agreement, disagreement, consensus and controversy cannot be computed (because no ontology contains the corresponding entities). We obtain a set of 456 mappings, out of which 325 have been evaluated to be correct (71.3% precision).

To check the behavior of our measures, we computed the averages of the measures of global agreement, global disagreement, consensus and controversy for both the mappings that have been evaluated correct by evaluators and the ones marked as incorrect (see Table 3). As expected, these values show a significant difference, with agreement and consensus favoring correct mappings, while disagreement and controversy favor incorrect mappings. To confirm these results, we also computed the correlations between each measure and the evaluated correctness of the mappings[2]. The results (see Table 3) show clear positive correlations for agreement and consensus, and negative correlations for disagreement and controversy (though the correlation for controversy is less significant). This confirms that there is a link between our agreement-related measures and the correctness of statements and so that these measures can be used to facilitate the assessment of such relations. Indeed, we calculated that, using the adequate threshold, we could obtain a precision value of up to 80% on the tested mappings by filtering on the measure of consensus alone.

7.3 Facilitating the Selection of Ontologies

Watson and other Semantic Web search engines provide the elementary functionalities to search and explore ontologies, but to better support the task of selecting the right ontologies, there is a need for a complete overview of the set of candidate ontologies they return and on how they relate to each other. Here, we experiment on using the agreement and disagreement measures to provide one element of such an overview. We used the 21 ontologies returned by Watson when querying for semantic documents containing a class with the term *SeaFood* in its ID or label and computed the agreement and disagreement measures for all pairs of ontologies in this set. The results are shown in Figure 1 where ontologies are numbered according to their rank in Watson[3].

Analyzing these diagrams, it appears that there is a certain level of 'coherence' in the results. In particular, homogeneous clusters can be built from the agreement and disagreement values. Indeed, the ontologies O1, O2, O3, O4, O5, O6, O7, O11, O12, O13, O16, O17, O18, O19 and O20 all fully agree with each other and, at the same time, partially agree and disagree with O14 and O15. O14 and O15 also form a cluster since they agree with each other, and consistently disagree with the same set of ontologies (the reason being that O14 and O15 are the ontologies considering that SeaFood is a subclass of Meat, but agree on all the other related statements). O21 is also particular, since it disagrees with most of the ontologies of the first cluster, sometimes fully. Indeed, it also considers SeaFood to be a subclass of Meat, and additionally disagrees on several other statements with some of the other ontologies (for example, it considers that tuna is a subclass of fish while several other ontologies consider tuna as an instance of fish). O8, O9 and O10 are particular since there is only a very small overlap between them and the other ontologies. For example, O9 only agrees with O11 that Vegan is a subclass of Vegetarian.

As a result of this experiment, it appears that the agreement and disagreement measures provide interesting elements for analyzing a set of candidate ontologies to identify possible contradictions, exceptions, as well as

[2]Using the Pearson sample correlation measure.

[3]valid on the 20/04/2009.

	agreem^t	disagreem^t	cons.	contro.
avg. correct	0.60	0.30	0.30	0.32
avg. incorrect	0.45	0.41	0.05	0.40
correlation	0.166	-0.165	0.165	-0.109

Table 3: Average of each measures on the sets of correct and incorrect mappings, and Pearson correlation between the measures and the correctness of the mappings.

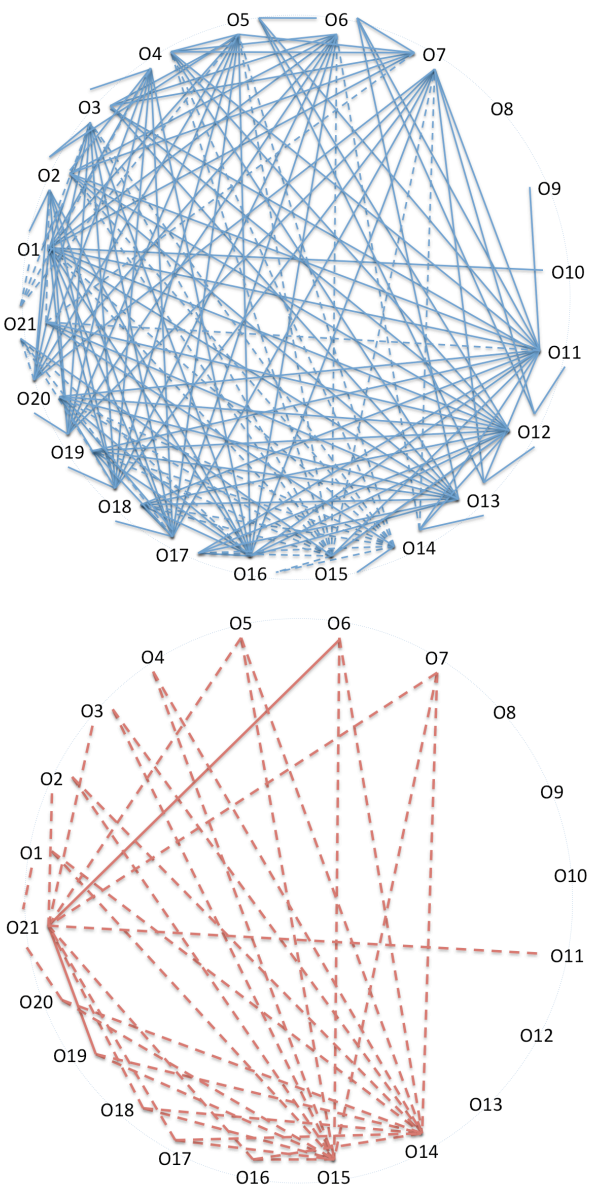

Figure 1: Agreement (top) and disagreement (bottom) relations among the 21 test ontologies. Plain lines represent full dis/agreement (measures' values = 1). Dashed lines represent partial dis/agreement (measures' values greater than 0).

sets of ontologies expressing compatible knowledge, hence supporting a more informed choice of ontologies to reuse.

8. CONCLUSION AND FUTURE WORK

In this paper, we presented a set of formally defined measures relying on the formal semantics of ontologies to assess the agreement and disagreement between a statement and an ontology or a set of ontologies, as well as between two ontologies. We have realized a straightforward implementation of all the measures described above and tested them on a number of examples.

At a short term, we intend to provide these measures as Web services on top of Watson, allowing external applications to check agreement related measures for particular statements. In particular, these services will be used in the Watson Plugin to provide useful indications to help the user in selecting good statements, and to rank them according to their level of dis/agreement.

Another interesting direction concerns the computation of 'explanations' for the measures, showing what statements conflict or enforce the one which is tested and, in addition to providing measures of agreements and disagreements between ontologies, also computing precisely 'on what' the considered ontologies dis/agree.

Finally, we believe that there are many possible usage scenarios for our measures which we have not explored yet. In particular, it would be interesting to investigate how meta information such as trust statements on particular ontologies could be propagated based on the dis/agreement values between these ontologies.

9. REFERENCES

[1] M. d'Aquin, C. Baldassarre, L. Gridinoc, M. Sabou, S. Angeletou, and E. Motta. Watson: Supporting next generation semantic web applications. In *WWW/Internet conference*, 2007.

[2] M. d'Aquin, M. Sabou, and E. Motta. Reusing Knowledge from the Semantic Web with the Watson Plugin. In *Demo session of International Semantic Web Conference, ISWC*, 2008.

[3] J. Euzenat and P. Shvaiko. *Ontology Matching*. Springer, 2007.

[4] S. Ghilardi, C. Lutz, and F. Wolter. Did I damage my ontology? a case for conservative extensions in description logics. In P. Doherty, J. Mylopoulos, and C. Welty, editors, *Proceedings of the Tenth International Conference on Principles of Knowledge Representation and Reasoning (KR'06)*, pages 187–197. AAAI Press, 2006.

[5] M. Sabou, M. d'Aquin, and E. Motta. Exploring the Semantic Web as Background Knowledge for Ontology Matching. *Journal of Data Semantics*, 2008.