Evaluating semantic relations by exploring ontologies on the Semantic Web

Conference or Workshop Item

How to cite:

Sabou, Marta; Fernandez, Miriam and Motta, Enrico (2009). Evaluating semantic relations by exploring ontologies on the Semantic Web. In: 14th International Conference on Applications of Natural Language to Information Systems (NLDB), 23-26 Jun 2009, Saarland, Germany.

Version: Accepted Manuscript

Link(s) to article on publisher's website:
http://dx.doi.org/doi:10.1007/978-3-642-12550-8$_2$2

Evaluating Semantic Relations by Exploring Ontologies on the Semantic Web

Marta Sabou, Miriam Fernandez, and Enrico Motta

Knowledge Media Institute (KMi)
The Open University, Milton Keynes, United Kingdom
{R.M.Sabou,M.Fernandez,E.Motta}@open.ac.uk

Abstract. We investigate the problem of evaluating the correctness of a semantic relation and propose two methods which explore the increasing number of online ontologies as a source of evidence for predicting correctness. We obtain encouraging results, with some of our measures reaching average precision values of 75%.

1 Introduction

The problem of understanding how two concepts relate to each other has been investigated in various fields and from different points of view. Firstly, the level of relatedness between two terms is a core input for several Natural Language Processing (NLP) tasks, such as word sense disambiguation, text summarization, annotation or correction of spelling errors in text. As a result, a wide range of approaches to this problem have been proposed which mainly explore two paradigms. On the one hand, corpora-based methods measure co-occurrence in a given context (usually characterized by means of linguistic patterns) across large-scale text collections [4,14]. On the other hand, knowledge rich methods use world knowledge explicitly declared in ontologies or thesauri (usually, WordNet) as a source of evidence for relatedness [3].

Secondly, from the beginnings of the Semantic Web (SW), where semantic relations are the core components of ontologies, the task of identifying the actual semantic relation that holds between two concepts has received attention in the context of the ontology learning field [5]. Finally, recent years have seen an evolution of Semantic Web technologies, which lead both to an increased number of online ontologies and to a set of mature technologies for accessing them[1]. These changes have facilitated the appearance of a new generation of applications which are based on the paradigm of reusing this online knowledge [6]. These applications differ substantially from the typical knowledge-based AI applications (as well as some of the early SW applications) whose knowledge base is provided a-priory rather than being acquired through re-use during runtime. They also reform the notion of knowledge reuse, from an ontology-centered view,

[1] http://esw.w3.org/topic/TaskForces/CommunityProjects/LinkingOpenData/SemanticWebSearchEngines

H. Horacek et al. (Eds.): NLDB 2009, LNCS 5723, pp. 269–280, 2010.

to a more fine-grained perspective where individual knowledge statements (i.e., semantic relations) are reused rather than entire ontologies. In the case of these applications, it is therefore important to estimate the correctness of a relation, especially when it originates from a pool of ontologies with varying quality.

The problem we investigate in this paper is evaluating the correctness of a semantic relation. Our hypothesis is that the Semantic Web is not just a motivation for investigating this problem, but can actually be used as part of the solution. We base this hypothesis on the observation that the Semantic Web is a large collection of knowledge-rich resources, and, as such it exhibits core characteristics of both data source types used in NLP for investigating relatedness: knowledge resources (structured knowledge) and corpora (large scale, federated). Earlier research has showed that although contributed by heterogeneous sources, online ontologies provide a good enough quality to support a variety of tasks [17]. It is therefore potentially promising to explore this novel source and to investigate how NLP paradigms can be adapted to a source with hybrid characteristics such as the SW. We phrase the above considerations into two research questions:

1. *Can the SW be used as a source for predicting the correctness of a relation?*
2. *Can we adapt existing NLP paradigms to the SW?*

To answer these questions we present two methods that explore online ontologies to estimate the correctness of a relation and which are inspired from two core paradigms used for assessing semantic relatedness. We perform an extensive experimental evaluation involving 5 datasets from two topic domains and covering more than 1400 relations of various types. We obtain encouraging results, with one of our measures reaching average precision values of 75%.

We start by describing some motivating scenarios where the evaluation of semantic relations is needed (Section 2). Then, we describe two measures designed for this purpose and give details over their implementation (Sections 3 and 4). In Section 5 we detail and discuss our experimental investigation and results. An overview of related work and our conclusions finalize the paper.

2 Motivating Scenarios

In this section we describe two motivating scenarios that would benefit from measures to evaluate the correctness of a semantic relation.

Embedded into the NeOn Toolkit's ontology editor, the Watson plugin[2] supports the ontology editing process by allowing the user to reuse a set of relevant ontology statements (equivalent to semantic relations) drawn from online ontologies. Concretely, for a given concept selected by the user, the plugin retrieves all the relations in online ontologies that contain this concept (i.e., concepts having the same label). The user can then integrate any of these relations into his ontology through a mouse click. For example, for the concept *Book* the plugin would suggest relations such as:

[2] http://watson.kmi.open.ac.uk/editor_plugins.html

- $Book \subseteq Publication$
- $Chapter \subseteq Book$
- $Book - containsChapter - Chapter$

The relations are presented in an arbitrary order. Because of the typically large number of retrieved relations it would be desirable to rank them according to their correctness. To date, however, no such methods exist thus hampering the user in finding the correct relations first, or indeed preventing him from reusing incorrect ones (e.g., $Chapter \subseteq Book$ where subsumption has been used incorrectly to model a meronymy relation).

As a second scenario we consider ontology matching [7], a core Semantic Web task. This task leads to establishing a set of mappings between the concepts of two input ontologies (i.e., an alignment). While these mappings take the form of semantic relations of various types, the focus of the community has primarily been in deriving and evaluating equivalence relations by comparing them against a-priory, manually-built, gold-standard alignments. However, as more and more matchers are capable of identifying other mappings than equivalence, the current gold-standard based evaluations need to be revised as it is impossible to manually predict all types of relations that would hold between the elements of two ontologies [16]. We hope that the methods described in this paper could be used as a way to automatically assess the correctness of alignments containing more than just equivalence mappings.

3 Evaluating the Correctness of Semantic Relations

To formally define our problem, let us denote a semantic relation as a triple $< s, R, t >$, where s is the source concept (or domain), t is the target concept (or range) and R denotes the relation that holds between the two concepts. R can define a wide range of relation types, such as hyponymy, disjointness, meronymy or simply any associative relation. Our aim is to derive a set of methods that can predict the level of correctness of such a relation, i.e., whether it is likely to be correct or incorrect. For the purposes of this work, we distinguish between a relation being *generically correct* and *correct or relevant in a given context*. When we decide on generic correctness we estimate the generic consensus on a relation independently of an interpretation context, while contextual-correctness or relevance should also take into account a given interpretation context. In this work we focus on generic correctness and leave contextual issues for future work.

In this section we propose two measures that exploit the large amount of online ontologies for predicting the correctness of a semantic relation. The measures are based on two different paradigms. The first measure explores the knowledge declared in online ontologies to predict correctness and as such it resembles the knowledge-rich methods reported in [3]. The second measure treats the Semantic Web as a corpus of ontologies for measuring the likely relatedness of the concepts involved in the relation and the popularity of that relation. As such, it is inspired from corpora-based methods similar to those described in [4,14].

3.1 Exploring Ontologies as Knowledge Artifacts

The measures in this section explore ontologies as knowledge artifacts and are based on the intuition that explicitly declared relations are more likely to be correct than implicit ones (i.e., those which are derived through reasoning).

Let $< s, R, t >$ be a relation which we wish to evaluate. Let n be the number of online ontologies such that each ontology O_i contains concepts similar to s and t ($s'_i = s$ and $t'_i = t$) and that a relation equivalent to R ($R'_i =$ R) is declared explicitly (or can be inferred) between s'_i and t'_i. For example, for the statement *aircraft* $\supseteq$ *helicopter* there are three ontologies (shown in Table 1) that explicitly (or implicitly) declare such a relation.

Table 1. Examples of derivation paths for $aircraft$ $\supseteq$ *helicopter*

Derivation Path and Ontology	**Path Length**
$O_1 : Aircraft \supseteq O_1 : Helicopter$ O_1 =`http://reliant.teknowledge.com/DAML/Mid-level-ontology.owl`	1
$O_2 : Aircraft \supseteq O_2 : Helicopter$ O_2 =`http://reliant.teknowledge.com/DAML/Transportation.owl`	1
$O_3 : Aircraft \supseteq O_3 : HeavierThanAirCraft \supseteq O_3 : Rotorcraft$ $\supseteq O_3 : Helicopter$ O_3 =`http://www.interq.or.jp/japan/koi_san/trash/aircraft3.rdf`	3

Our measure relies on the hypothesis that there is a correlation between the length of the derivation path and the correctness of the relation. In particular, we think that longer paths probably lead to the derivation of less obvious relations, which are therefore less likely to be correct. To verify this hypothesis we compute three values: $AveragePathLength_R$ is the average of the lengths of all derivation paths for relation R (e.g., in our case $(1 + 1 + 3)/3 = 1.66$), $minLength_R$ is the length of the shortest derivation path that lead to R (in our case, 1), and $maxLength_R$ is the length of the longest derivation path associated to R (in our case, 3). Formally:

$$AveragePathLength_R = \frac{\sum_i PathLength_{R,O_i}}{n}$$

$$minLength_R = min_i(PathLength_{R,O_i}); maxLength_R = max_i(PathLength_{R,O_i})$$

3.2 Exploring Online Ontologies as a Corpus

Unlike in the previous section, the focus of the measures presented here is on exploring the Semantic Web as a corpus of ontologies for computing concept relatedness and relation popularity.

For a relation $< s, R, t >$ to be evaluated, we define $RelatednessStrength_{s,t}$ as the ratio between the number of ontologies from which a relation can be deduced between s and t (i.e., $|O_{s,r,t}|$) and the number of all ontologies where these concepts are mentioned but not necessarily related (i.e., $|O_{s,t}|$). This measure is an indication of how likely it is that the two concepts are related. Indeed, if all the ontologies that mention s and t also lead to deriving a relation between them, then s and t are likely to be related. This measure takes its values in the interval (0,1], with low values corresponding to terms that are weakly related, and 1 to those that are related in all ontologies that they are mentioned. For example, *Rodents* and *Animals* appear in 5 ontologies and each of these ontologies leads to a relation between them. While this measure does not inform about the correctness of a particular relation R, we assume that a relation established between terms that are not likely to be related is less likely to be correct than a relation established between closely related terms. Formally:

$$RelatednessStrength_{s,t} = \frac{|O_{s,r,t}|}{|O_{s,t}|}$$

Table 2. Examples of relations between *honey* and *food*

Relation	Derivation Path and Ontology
sibling	$O_1 : Honey \subseteq O_1 : Sweetener \subseteq O_1 : SweetTaste \subseteq O_1 : PartiallyTangible$ $O_1 : Food \subseteq O_1 : FoodOrDrink \subseteq O_1 : HumanScaleObject \subseteq O_1 : PartiallyTangible$ $O_1 =$`http://secse.atosorigin.es:10000/ontologies/cyc.owl`
$\subseteq$	$O_2 : Honey \subseteq O_2 : Food$ $O_2 =$`http://sweet.jpl.nasa.gov/ontology/substance.owl`
$\subseteq$	$O_3 : honey \subseteq O_3 : sweetener \subseteq O_3 : flavoring \subseteq O_3 : plant-derived-foodstuff \subseteq O_3 : foodstuff \subseteq O_3 : food$ $O_3 =$`http://morpheus.cs.umbc.edu/aks1/ontosem.owl`

We then define $StrengthRelation_R$ for measuring the popularity of a relation R over any type of relations that can be derived between s and t. This measure also takes its values from (0,1], with the lowest values indicating that R has a low popularity (and therefore it is likely to be incorrect) and a value of 1 being obtained when R is the only relation derivable between these concepts (and therefore it is likely to be correct). For example, as shown in Table 2, because it is more popular amongst online ontologies, the $\subseteq$ relation between *honey* and *food* will have a higher value for this measure (i.e., 0.66) than the *sibling* relation between the same concepts (i.e., 0.33). Formally:

$$StrengthRelation_R = \frac{freq(R)}{allRels_{s,t}}$$

Note that we have also experimented with various ways of normalizing these measures, however, we do not present them because experimental evaluation has showed a less optimal behavior than for the original measures.

4 Implementation

We implemented our measures using the services of the Watson[3] semantic web gateway. Watson crawls and indexes a large number of online ontologies[4] and provides a comprehensive API which allows exploring these ontologies.

We have also built an algorithm that, using Watson, extracts relations between two given terms from online ontologies. The algorithm is highly parameterized[5]. For the purposes of this study we have configured it so that for each pair (A,B) of terms it identifies all ontologies containing the concepts `A'` and `B'` corresponding to `A` and `B` from which a relation can be derived between these terms. Correspondence is established if the labels of the concepts are lexical variations of the same term. For a given ontology (O_i) the following derivation rules are used:

- if $A'_i \equiv B'_i$ then derive $A \xrightarrow{\equiv} B$;
- if $A'_i \sqsubseteq B'_i$ then derive $A \xrightarrow{\sqsubseteq} B$;
- if $A'_i \sqsupseteq B'_i$ then derive $A \xrightarrow{\sqsupseteq} B$;
- if $A'_i \perp B'_i$ then derive $A \xrightarrow{\perp} B$;
- if $R(A'_i, B'_i)$ then derive $A \xrightarrow{R} B$;
- if $\exists\ P_i$ such that $A'_i \sqsubseteq P_i$ and $B'_i \sqsubseteq P_i$ then derive $A \overset{sibling}{\longrightarrow} B$.

Note that in the above rules, the relations between A'_i and B'_i represent both explicit and implicit relations (i.e., relations inherited through reasoning) in O_i. For example, in the case of two concepts labeled $DrinkingWater$ and *tap_water*, the algorithm deduces the relation $DrinkingWater \xrightarrow{\sqsubseteq}$ *tap_water* by virtue of the following subsumption chain in the TAP ontology: *DrinkingWater* ⊑ *FlatDrinkingWater* ⊑ *TapWater*.

5 Experimental Evaluation

In this section we describe the experimental evaluation of the measures detailed in Section 3. We have used the implementation presented in Section 4 over the datasets described in Section 5.1. We then further explore and analyze the results for both measure types (Sections 5.2 and 5.3).

5.1 Data Sets

As experimental data we have used datasets from the domain of ontology matching, in the form of alignments obtained in two different test-cases put forward by the Ontology Alignment Evaluation Initiative[6](OAEI), an international body that coordinates evaluation campaigns for this task.

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

[4] Estimated to 250.000 during the writing of this paper.

[5] A demo of some of these parameters and an earlier version of the algorithm are available at `http://scarlet.open.ac.uk/`

[6] `http://oaei.ontologymatching.org/`

Table 3. Overview of the experimental data sets and their characteristics

Data Set	Nr. of Relations	Type of Relations	Domain
AGROVOC/NALT	380	$\subseteq$, $\supseteq$, $\perp$	Agriculture
OAEI'08 301	112	$\subseteq$, $\supseteq$, $\perp$, named relations	Academia
OAEI'08 302	116	$\subseteq$, $\supseteq$, $\perp$, named relations	Academia
OAEI'08 303	458	$\subseteq$, $\supseteq$, $\perp$, named relations	Academia
OAEI'08 304	386	$\subseteq$, $\supseteq$, $\perp$, named relations	Academia
Total	**1452**		

The AGROVOC/NALT data set has been obtained by performing an alignment between the United Nations' Food and Agriculture Organization (FAO)'s AGROVOC ontology and its US equivalent, NALT. The relations established between the concepts of the two ontologies are of three types: $\subseteq$, $\supseteq$ and $\perp$. Each relation has been evaluated by two experts, as described in more detail in [15].

The OAEI'08 dataset represents the alignments obtained by the Spider system on the 3** benchmark datasets and their evaluation [16]. This dataset contains four distinct datasets representing the alignment between the benchmark ontology and the MIT (301), UMBC(302), KARLSRUHE(303) and INRIA(304) ontologies respectively. Besides the $\subseteq$, $\supseteq$ and $\perp$ relation types, this data set also contains named relations (e.g., $inJournal(Article, Journal)$). Table 3 provides a summary of these datasets and their characteristics.

5.2 Results for the Derivation Path Based Measures

To investigate the correlation between the characteristics of the derivation path and the correctness of a relation, we computed the $AveragePathLength_R$, $minLength_R$ and $maxLength_R$ values for all relations in our five datasets. Then, for each dataset we computed the mean value for $AveragePathLength_R$ for relations judged to be false (F-Mean) and those judged to be true (T-Mean). We also repeated these calculations for the dataset obtained by merging the relations in all datasets. The values of these computations are shown in columns two and three of Table 4. We notice that for all datasets there is a clear difference between the mean path length of true and false relations, where false relations, on average, have a longer derivation path (always over 2) than the true ones (always under 2). This is already a good indication that this measure captures a valid hypothesis.

We continued our investigations by computing a threshold value for which the assignment of correctness values correlates best with that of the human judgement. This was measured in terms of a precision value computed as the ratio of correctly assessed relations with that threshold over all relations in the dataset. Columns four and five of Table 4 show our results. We note that there is considerable variation in the values of the optimal threshold between datasets and that some are very close to the extreme values (e.g., in the case of AGROVOC/NALT, and OAEI'08 304 the best threshold is close to F-Mean, while for OAEI'08 302 the threshold is almost identical with T-Mean). Given this situation we tried to

Table 4. Correlation between the derivation path characteristics and correctness

Data Set	$AveragePathLength_R$		Best Threshold	Prec.	Best Threshold'	Prec.'
	F-Mean	T-Mean				
AGROVOC/NALT	2.07	1.77	2.08	65%	2.00	71%
OAEI'08 301	2.58	1.29	1.66	86%	1.29	**94%**
OAEI'08 302	2.50	1.70	1.71	74%	1.71	80%
OAEI'08 303	2.83	1.76	2.60	76%	2.00	78%
OAEI'08 304	2.31	1.81	2.25	69%	2.00	73%
Merged Datasets	2.46	1.73	2.33	71%	**2.00**	**75%**

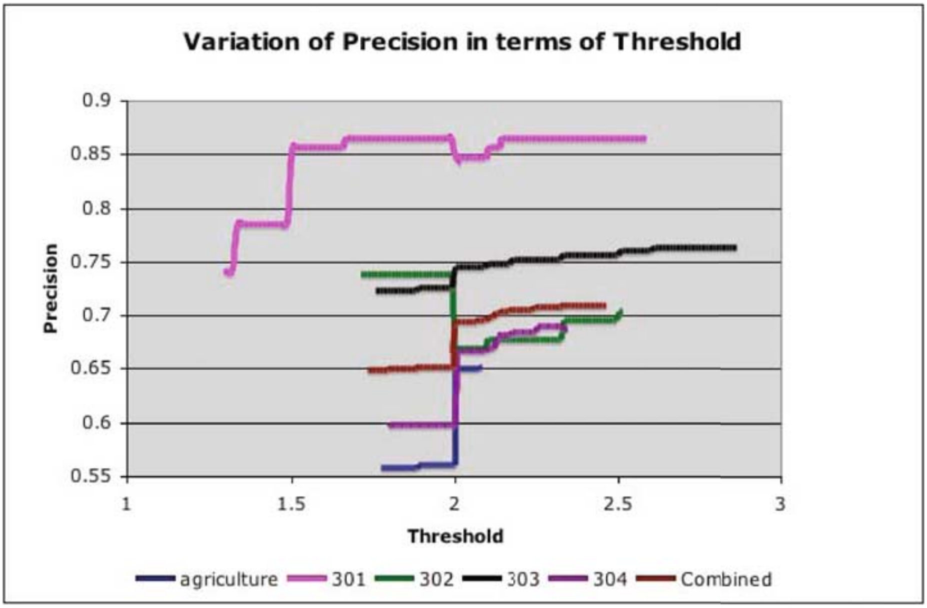

Fig. 1. Precision variation in terms of threshold values set for the length of the derivation path

approximate a global optimal threshold by computing it on the merged dataset. This yielded the value 2.33. The precision values per dataset vary from a minimum of 65% to a maximum of 86%, and we obtained an average precision for the merged dataset of 71%. Figure 1 graphically depicts the variation of precision in terms of threshold for all the five datasets and the merged datasets.

When examining the values of the $minLength_R$ and $maxLength_R$ measures, we observed that the overwhelming majority of relations that were deduced with paths of different lengths (i.e., their min and max path values were different) were correct relations. A good example is that of *aircraft* $\supseteq$ *helicopter* which is explicitly declared in two ontologies, while in another ontology this relation is defined in terms of a chain of more fine-grained relations (see Table 1). Another

Table 5. Average values for True and False relations, best threshold and precision values for *RelatednessStrength* and *StrengthRelation*

Data Set	Relatedness Strength		Best Thresh.	Prec.	Strength Relation		Best Thresh.	Prec.
	T	F			T	F		
AGROVOC/NALT	0.91	0.88	0.89	45%	0.34	0.34	0.34	36%
OAEI'08 301	0.81	0.75	0.75	41%	0.36	0.04	0.33	42%
OAEI'08 302	0.80	0.75	0.80	46%	0.38	0.11	0.11	38%
OAEI'08 303	0.58	0.50	0.55	43%	0.15	0.11	0.12	53%
OAEI'08 304	0.63	0.55	0.59	46%	0.23	0.15	0.16	56%

example relates to relations involving plants or animals, such as $goat \subseteq animal$. Some ontologies contain these relations explicitly (i.e., with a path length of 1), while others contain a more fine-grained path between these concepts, e.g., $goat \subseteq ungulate \subseteq mammal \subseteq vertebrate \subseteq animal$[7]. We have incorporated this observation in the calculation of the best threshold as follows: any relation which has the $AveragePathLength_R$ over the threshold but whose values for $minLength_R$ and $maxLength_R$ differ, is considered to be a True relation. The recomputed values for the best threshold and the corresponding precision are shown in the last two columns of Table 4. Remarkably, in the case of most datasets this observation has lowered the threshold and for all datasets it increased the precision to values ranging now from 71% to 94%. On the combined dataset this lead to a threshold of 2% and a precision value of 75%. We regard these values as illustrative for our derivation path based measures.

5.3 Results for the Corpora Inspired Measures

In columns two and three of Table 5 we present the average values of the *RelatednessStrength* measure for True and False relations respectively. Our hypothesis for this measure was that correct relations will most likely be declared between highly related terms (i.e., where the value of this measure is high), while the inverse will hold for false relations. Indeed, this hypothesis is verified by the obtained numbers as, for all datasets, on average, True relations are established between terms with higher *RelatednessStrength* than False ones. We note however, that the difference between the average values of this measure for True and False relations is rather small thus potentially decreasing its discriminative power. Indeed, this is verified when computing the best threshold and the corresponding precisions (columns four and five), as the precision values are quite low, not even reaching 50%.

In the second half of Table 5 we present the results of our experiments for the *StrengthRelation* measure. Our hypothesis was that high values of this measure, corresponding to popular relations, will mostly characterize True relations, while False relations will be associated with lower values. This hypothesis has been

[7] `http://morpheus.cs.umbc.edu/aks1/ontosem.owl`

verified in four out of five datasets, where the average value of the measure is lower for False relations than for True relations. The AGROVOC/NALT dataset is an exception, where both values are the same. We also notice that the difference between these values is higher than for the previous measure. This has a positive effect on the discriminative value of the measure, and indeed, we obtain higher precision values than for *RelatednessStrength* (up to 56%).

We conclude that, overall, the *StrengthRelation* measure has a better behavior than *RelatednessStrength*, although both are clearly inferior to the derivation path based measures discussed before. We think this is primarily due to the fact that, despite its increasing size, the Semantic Web is still rather sparse and as such negatively affects any corpus based measures. These measures could potentially be strengthened when combined with path based measures.

6 Related Work

An overview of related work suggests that various approaches are used to evaluate relatedness or semantic relations. The output of measures that provide a relatedness (or similarity) coefficient [3,14] has been evaluated through theoretical examination of the desirable mathematical properties [10], by assessing their effect on the performance of other tasks [3], and mainly by comparison against human judgement by relying on gold-standards such as the Miller Charles data set [13] or WordSim353[8]. The field of ontology learning has focused on learning taxonomic structures (consisting of hyponymy relations) and other types of relations [5]. For example, Hearst pattern based techniques have been successfully scaled up to the Web in order to identify certain types of relations such as hyponymy, meronymy [18] or complex qualia structures [5]. The evaluation measures used to assess the correctness of the learned relations either rely on comparison to a conceptual structure that plays the role of a gold-standard (mostly using the measures described in [12]) or on expert evaluation. Note that the techniques that use Hearst patterns on the Web can implicitly be used to verify whether a relation is of a given type. As such, these techniques are the most similar to the presented work, with the difference that they explore the Web (a large body of *unstructured* knowledge) rather than the Semantic Web (a collection of *structured* knowledge).

Another important body of work exists in the context of ontology evaluation (see two recent surveys for an overview [2], [9]), where existing approaches are unevenly distributed in two major categories. On the one hand, a few principled approaches define a set of well-studied, high level ontology criteria to be manually assessed (e.g., OntoClean [8], Ontometric [11]). On the other hand, *automatic* approaches cover different evaluation perspectives (coverage of a corpus, similarity to a gold standard ontology) and levels (e.g., labels, conceptual structure). Common to these approaches is that they focus on evaluating an

[8] `http://www.cs.technion.ac.il/~gabr/resources/data/wordsim353/wordsim353.html`

ontology as a whole rather than on assessing the correctness of a given relation as we do in this work.

7 Conclusions and Future Work

In this paper we investigated the problem of predicting the correctness of a semantic relation. Our hypothesis was that the Semantic Web can be used as a source of knowledge for this task and that existing NLP paradigms can be adapted to explore online ontologies.

Based on our experimental results, we can conclude that the Semantic Web is a promising source of information for addressing the relation evaluation problem. Indeed, a combination of our measures which explore ontologies as knowledge artifacts lead to an average precision value of 75% (with an individual result of 94% for one of the datasets). Our results have also shown that the measures inspired from different paradigms had varying performance. The measures that explored the knowledge provided by ontologies outperformed those that regarded the Semantic Web as a corpus. A simple explanation could be the still sparse nature of the Semantic Web which hampers its meaningful use as a corpus. Our future work will focus in trying to enhance and combine the methods from these two paradigms, as well as complementing them with other sources than the SW.

Additionally to our conclusions, we observe a potential of using the proposed measures for evaluating ontology characteristics. For example, in the case of a relation that is derived from paths of different lengths, we can conclude that the ontology which leads to the shorter path is more concise (less detailed) than the one which leads to a longer derivation path. While valuable, such estimations of conceptual complexity have been difficult to capture with current ontology evaluation measures such as those described in [1].

In this work we have taken some simplifying assumptions which will be revisited during future work. Firstly, we gave a broad definition of correctness without distinguishing between different types of correct or incorrect relations. In future work we plan to identify and individually investigate different types of correct/incorrect relations. Secondly, when counting named relations we have assumed that a relation can have a single lexicalization. This assumption is however not verified in a minimal number of cases when a given semantic relation is present with different labels. Finally, in the case of path based measures we have given the same weight to each relation within a path, although, it is well-known from NLP, that even within the same ontology, different relations often cover different conceptual distances and should be weighted differently [3]. Our ongoing work explores ontology structure characteristics (e.g., depth, breadth) as a way to predict the granularity of the conceptual space covered by relations.

Acknowledgements

Work funded by the NeOn IST-FF6-027595 and X-Media IST-FP6-026978 projects.

References

1. Alani, H., Brewster, C.: Ontology Ranking based on the Analysis of Concept Structures. In: Proc. of the Third Int. Conf. on Knowledge Capture. ACM, New York (2005)
2. Brank, J., Grobelnik, M., Mladenic, D.: A survey of ontology evaluation techniques. In: Proc. of the Conf. on Data Mining and Data Warehouses (2005)
3. Budanitsky, A., Hirst, G.: Evaluating WordNet-based measures of semantic distance. Computational Linguistics 32(1), 13–47 (2006)
4. Calibrasi, R.L., Vitanyi, P.M.: The Google Similarity Distance. IEEE Transactions on Knowledge and Data Engineering 19(3), 370–383 (2007)
5. Cimiano, P.: Ontology Learning and Population from Text: Algorithms, Evaluation and Applications. Springer, Heidelberg (2006)
6. d'Aquin, M., Motta, E., Sabou, M., Angeletou, S., Gridinoc, L., Lopez, V., Guidi, D.: Towards a New Generation of Semantic Web Applications. IEEE Intelligent Systems 23(3), 20–28 (2008)
7. Euzenat, J., Shvaiko, P.: Ontology Matching. Springer, Heidelberg (2007)
8. Guarino, N., Welty, C.A.: An Overview of OntoClean. In: Staab, S., Studer, R. (eds.) Handbook on Ontologies. Springer, Heidelberg (2004)
9. Hartmann, J., Sure, Y., Giboin, A., Maynard, D., Suarez-Figueroa, M.C., Cuel, R.: Methods for ontology evaluation. Knowledge Web Deliverable D1.2.3 (2005)
10. Lin, D.: An information-theoretic definition of similarity. In: Proc. of the 15th Int. Conf. on Machine Learning (1998)
11. Lozano-Tello, A., Gomez-Perez, A.: ONTOMETRIC: A Method to Choose the Appropriate Ontology. Journal of Database Management 15(2), 1–18 (2004)
12. Madche, A., Staab, S.: Measuring similarity between ontologies. In: Proc. of the European Conf. on Knowledge Acquisition and Management (2002)
13. Miller, G.A., Charles, W.G.: Contextual Correlates of Semantic Similarity. Language and Cognitive Processes 6(1), 1–28 (1991)
14. Mohammad, S., Hirst, G.: Distributional Measures as Proxies for Semantic Relatedness. Submitted for peer review
15. Sabou, M., d'Aquin, M., Motta, E.: Exploring the Semantic Web as Background Knowledge for Ontology Matching. Journal on Data Semantics XI (2008)
16. Sabou, M., Gracia, J.: Spider: Bringing Non-Equivalence Mappings to OAEI. In: Proc. of the Third International Workshop on Ontology Matching (2008)
17. Sabou, M., Gracia, J., Angeletou, S., d'Aquin, M., Motta, E.: Evaluating the Semantic Web: A Task-based Approach. In: Proc. of ISWC/ASWC (2007)
18. van Hage, W., Kolb, H., Schreiber, G.: A Method for Learning Part-Whole Relations. In: Proc. of the 5th Int. Semantic Web Conf. (2006)