What makes a good ontology? A case-study in fine-grained knowledge reuse

Miriam Fernández, Chwhynny Overbeeke, Marta Sabou, Enrico Motta

[1] Knowledge Media Institute
The Open University, Milton Keynes, United Kingdom

{M.Fernandez, C.Overbeeke, R.M.Sabou, E.Motta}@open.ac.uk

Abstract. Understanding which ontology characteristics can predict a "good" quality ontology, is a core and ongoing task in the Semantic Web. In this paper, we provide our findings on which structural ontology characteristics are usually observed in high-quality ontologies. We obtain these findings through a task-based evaluation, where the task is the assessment of the correctness of semantic relations. This task is of increasing importance for a set of novel Semantic Web tools, which perform fine-grained knowledge reuse (i.e., they reuse only appropriate parts of a given ontology instead of the entire ontology). We conclude that, while structural ontology characteristics do not provide statistically significant information to ensure that an ontology is reliable ("good"), in general, richly populated ontologies, with higher depth and breadth variance are more likely to provide reliable semantic content.

Keywords: semantic relations, knowledge reuse, Semantic Web.

1 Introduction

Ontologies are fundamental Semantic Web (SW) technologies, and as such, the problem of their evaluation has received much attention from areas such as ontology ranking [8], selection [16][21], evaluation [11] and reuse [22]. Various approaches have been proposed in these fields, ranging from manual evaluation to (semi-) automatic evaluation of a single ontology to benchmark evaluation of the entire Semantic Web, and, finally, to task-based evaluations of a single ontology or a collection of ontologies. These studies have explored a variety of ontology characteristics that could predict ontology quality, including characteristics such as the modeling style of the ontologies, their vocabulary, structure, or performance within a given task. In this paper we continue the investigation of what makes a "good" ontology by using a *task-based approach* to evaluate the *collection of ontologies* available on the SW in terms of measures relating to their *structure*.

The context of our work is that of fine-grained knowledge reuse, i.e., the reuse of ontology parts rather than the ontology as a whole. This kind of knowledge reuse is increasingly frequent, particularly for the new family of applications that take advantage of the large scale of the Semantic Web and the set of mature technologies

for accessing its content[1] in order to reuse online knowledge. In the case of these applications, knowledge reuse happens at run-time, and therefore it primarily focuses on the reuse of small parts of ontologies, typically at the level of a semantic relation [17]. This is why it is essential to automatically detect the quality of such relations.

The task we focus on in this paper is the evaluation of a single semantic relation (and not that of an entire ontology). We have built an algorithm that explores online ontologies in order to perform this task [18]. The performance of the task depends on the selection of these ontologies. We experiment with a set of structure-based ontology characteristics to select appropriate ontologies and decide which characteristics are more important by measuring their influence on the performance achieved when predicting the quality of relations. The correlation between structure-based ontology characteristics and ontology correctness arises from our own experience in previous works [10][18], and other ontology evaluation studies where this distinction seems to be natural, useful and recurrent (see e.g. [15]).

Our findings show that while structural ontology characteristics do not provide statistically significant information to identify a correct ontology, some of them point to valuable information that can help enhance ontology selection techniques. In particular, we conclude that richly populated ontologies with a high breadth and depth variance are more likely to be correct, and should be ranked higher by ontology selection algorithms.

The contribution of our paper is two-fold. On the one hand, we further advance work on automatic relation evaluation by providing our findings on the ontology characteristics which could predict which ontologies are most likely to provide correct relations. On the other hand, a side-effect of this work is a large-scale investigation of what are the core structural characteristics that can predict a good-quality ontology.

The rest of the paper is structured as follows. We present related work in Section 2 and describe some motivating scenarios in the context of fine-grained knowledge reuse in Section 3. Section 4 introduces the task we focus on, the evaluation of a single semantic relation, and its implementation. We present the evaluation setup in Section 5 and detail experimental results in Section 6. We conclude in Section 7.

2 Related Work

As the number of ontologies on the Web increases, the need arises to determine which ontologies are of the highest quality or are the most appropriate for a certain task. There are several conceptions of what makes a “good” ontology, which will be discussed in this section.

Significant work has been done in the area of ontology quality assessment [6][14]. Most of these attempts try to define a generic quality evaluation framework. As a result, specific applications of ontologies are not taken into account, and the ontology is considered as a whole during its quality evaluation.

---

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

Existing evaluation methods rely on rather simple ways of specifying an information need, such as (sets of) keywords or a corpus from which sets of keywords are abstracted and output their results as a ranked list of ontologies [21].

There are three major categories of ontology evaluation approaches:

- *Manual approaches* are those based on human interaction to measure ontology features not recognizable by machines [14].
- *Automatic approaches* are those that evaluate an ontology by comparing it to a Golden Standard, which may itself be an ontology [15] or some other kind of representation of the problem domain [4].
- *Task-based approaches* are those that evaluate the ontologies by plugging them in an application, and measuring the quality of the results that the application returns [19].

The different existing methods of evaluation also vary with regard to their selection criteria and evaluation metrics. Aspects that are generally considered to be useful for the evaluation of the quality of an online ontology, shown in Table 1 are:

- Evaluation of *syntax* checks if an ontology is syntactically correct. This is most important for ontology-based applications as the correctness reflects on the application [14].
- *Cohesion to domain and vocabulary* measures the congruence between an ontology and a domain [4][6][16].
- *Structural* evaluation deals with the assessment of taxonomical relations versus other semantic relations, i.e., the ratio of Is-A relationships and other semantic relationships in an ontology is evaluated [5].
- *Population of classes* measures instance-related metrics such as how instances are distributed across classes or average population [23].
- *Usage statistics and metadata* evaluate those aspects that focus on the level of annotation of ontologies, i.e., the metadata of an ontology and its elements [6][9][8].

Table 1. Summary of existing approaches to ontology evaluation and the evaluation criteria they explore (adapted from [22]).

Quality Framework	**Syntax Evaluation**	**Domain cohesion**	**Structural evaluation**	**Population of classes**	**Usage statistics**
AKTiveRank [2]		X	X		
OntoClean [11]			X		
OntoKhoj [16]		X			X
Ontometric [14]	X				
OntoQA [23]			X	X	
OntoSelect [5]			X		
Semiotic metrics [6]	X	X			X
Swoogle [8]					X

In this work we report on a task-based evaluation of online available ontologies, where we investigate which structural and popularity characteristics of these ontologies are good indicators to measure their quality.

3 Use cases in the context of fine-grained knowledge reuse

In this section, we describe two motivating scenarios where fine-grained knowledge reuse is performed rather than reuse of ontologies as a whole.

Embedded in the NeOn Toolkit's ontology editor, the Watson plugin[2] allows the user to reuse a set of relevant ontology statements (equivalent to semantic relations) drawn from online ontologies in order to construct new knowledge. Concretely, for a given concept selected by the user, the plug-in retrieves all the relations in online ontologies that contain this concept (i.e., concepts that have 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: Book ⊆ Publication, Chapter ⊆ Book or Book -containsChapter- Chapter. These semantic statements are presented in an arbitrary order. Because of the typically large number of retrieved semantic statements it would be desirable to rank them according to their correctness.

Our second scenario is provided by PowerAqua [13], an ontology-based Question Answering (QA) system which receives questions in natural language and is capable of deriving an answer by combining knowledge gathered from multiple online ontologies. In a nutshell, the system breaks up the user query in several triple-like structures, which are then matched to appropriate triples (or relations) within online ontologies. PowerAqua derives the final answer by combining these ontology triples. As in the case of the Watson plug-in, PowerAqua does not evaluate the quality of these relations. Our work on establishing a correlation between certain ontology characteristics and the quality of the relations they provide would improve PowerAqua's ability to discard noise or irrelevant semantic information.

4 The task: Evaluating the quality of semantic statements

The task we use as a means to get an insight into the quality of online ontologies is that of evaluating the quality of a semantic relation. We define a semantic relation <*s, R, t*> as a triple where *s* represents the source term, *t* represents the target term, and *R* represents the relation between those terms, e.g., <*Helicopter,* ⊆ *, Aircraft*>. *R* can represent a wide range of relation types, such as hyponymy, disjointness, or simply any associative relation.

In our work, for any given relation we want to evaluate, we are capable to identify all online ontologies that directly or indirectly link *s* and *t*. Fig. 1 shows the example of three ontologies (O_1, O_2, O_3) that can lead to a relation between Aircraft and Helicopter. O_1[3] contains a direct subclass relation while O_2[4] contains a direct disjoint relation between Aircraft and Helicopter. O_3[5] provides an implicit subclass relation

[2] http://watson.kmi.open.ac.uk/WatsonWUI/
[3] http://reliant.teknowledge.com/DAML/Transportation.owl
[4] http://reliant.teknowledge.com/DAML/Mid-level-ontology.owl
[5] http://www.interq.or.jp/japan/koi_san/trash/aircraft3.rdf

between these two concepts, which can be inferred from the following derivation path: *Helicopter* ⊆ *Rotorcraft* ⊆ *HeavierThanAirCraft* ⊆ *Aircraft*

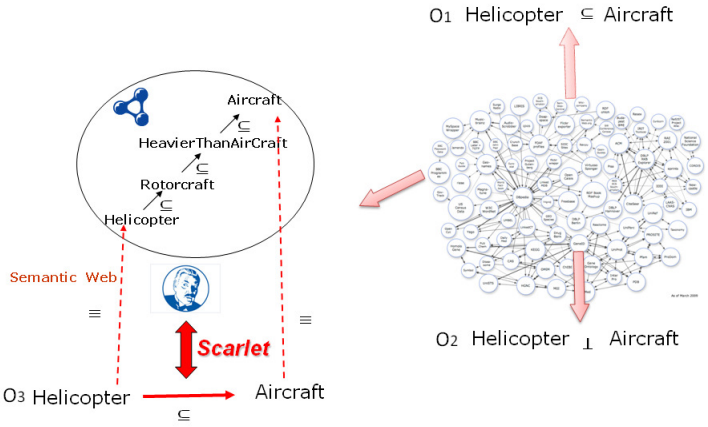


Fig. 1. Example of finding relations between Helicopter and Aircraft on the SW.

In this example we can see that different ontologies provide information of a different quality. While O_1 and O_3 provide a correct relation in terms of domain modeling, this is not the case for O_2. Further, note that even if they agree on the relation between Helicopter and Aircraft, ontologies O_1 and O_3 have different ways to declare this relation: explicitly (derivation path length = 1) or implicitly (derivation path length = 3). In this work we make use of the fact that different ontologies provide relations between the same terms in order to investigate which ontology characteristics can predict the ontology that is most likely to provide a correct relation.

To perform our task we use a software package that, given two terms, can identify all online ontologies that lead to a relation between these terms, as well as the actual relation and its derivation path. We implemented this package using the services of the Watson[6] SW gateway. Watson crawls and indexes a large number of online ontologies[7] and provides a comprehensive API which allows us to explore these ontologies.

The relation extraction algorithm is highly parameterized[8]. For the purposes of this study we have configured it such 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 semantic 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 $A \equiv B$.
- If $A'_i \subseteq B'_i$ then derive $A \subseteq B$.

[6] http://watson.kmi.open.ac.uk
[7] Estimated to 250,000 during the writing of this paper.
[8] A demo of some of these parameters and an earlier version of the algorithm are available at http://scarlet.open.ac.uk/

- If $A'_i \supseteq B'_i$ then derive $A \supseteq B$.
- If $A'_i \perp B'_i$ then derive $A \perp B$.
- If $R(A'_i, B'_i)$ then derive $A \ _R \ B$.
- If P_i such that $A'_i \subseteq P_i$ and $B'_i \subseteq P_i$ then derive A sibling 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* $\subseteq$ *tap_water* through the following subsumption chain in the TAP[9] ontology: *DrinkingWater* $\subseteq$ *FlatDrinkingWater* $\subseteq$ *TapWater*.

5 Evaluation setup

This section describes the evaluation setup. Here we explain the set of measures and datasets that we have selected to perform the evaluation.

5.1 Measures

Twelve different measures have been considered to evaluate the quality of the ontologies. Because these measures are investigated in the context of applications that need to select semantic knowledge at runtime, they must accomplish two main requirements: *generality* and *performance*. *Generality* refers to the applicability of the measures to any potential ontology available in the Web, independent of its language, size, or any other characteristic. *Performance* refers to the availability of these measures at runtime. This requirement generally implies that the measures are either lightweight in terms of computational requirements or pre-computed. The list of selected measures has been divided in two main groups:

a) Knowledge coverage and popularity measures
- Number of classes: number of classes in a given ontology.
- Number of properties: number of properties in a given ontology.
- Number of individuals: number of individuals in a given ontology.
- Direct popularity: number of ontologies importing a given ontology.

b) Structural ontology measures
- Maximum depth: size of the longest branch in the given ontology.
- Minimum depth: size of the shortest branch in the given ontology.
- Average depth: average size of the branches of the given ontology.
- Depth variance: variance of the size of the branches in the ontology.
- Maximum breadth: size of the largest level of the ontology.
- Minimum breadth: size of the narrowest level of the ontology.
- Average breadth: average size of the levels of the ontology.
- Breadth variance: variance of the size of the levels in the ontology.

[9] http://139.91.183.30:9090/RDF/VRP/Examples/tap.rdf

5.2 Datasets

As experimental data we 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[10] (OAEI), an international body that coordinates evaluation campaigns for this task.

The AGROVOC/NALT dataset 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: ⊆, ⊇, and ⊥. Each relation has been evaluated by experts, as described in more detail in [17].

The OAEI'08 dataset represents the alignments obtained by the Spider system on the 3** benchmark datasets and their evaluation [20]. 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 ⊆, ⊇, and ⊥ relation types, this dataset also contains named relations, e.g. <*Article, inJournal, Journal*>. Table 2 provides a summary of these datasets and their characteristics.

Table 2. Overview of the experimental datasets and their characteristics.

Data Set	Nr. Of Relations	Type of Relations	Domain
AGROVOC/NALT	380	⊆, ⊇, ⊥	Agriculture
OAEI'O8 301	112	⊆, ⊇, ⊥, named relations	Academia
OAEI'O8 302	116	⊆, ⊇, ⊥, named relations	Academia
OAEI'O8 303	458	⊆, ⊇, ⊥, named relations	Academia
OAEI'O8 304	386	⊆, ⊇, ⊥, named relations	Academia
Total	**1452**		

6 Evaluation Results

In this section we describe the study we conducted to evaluate the discriminative effect of the proposed measures when selecting the ontologies that are most likely to provide correct relations. For this purpose we have used the datasets presented in Section 5.2 and the implementation described in Section 4.

6.1 Evaluating the quality of semantic statements: types of SW matches

As we can see in Section 5.2, the datasets selected for the study contain four different types of relations *R*: ⊆, ⊇, ⊥ and named. For each individual triple <*s, R, t*> in the dataset a user evaluation is available, stating whether the relation *R* between *s* and *t* is correct.

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

Each triple <*s, R, t*> is then searched in the SW using the methodology described in Section 4. As a result, all online ontologies that directly or indirectly link *s* and *t* are identified. For each relation *R* to be evaluated we consider five different potential matches within online ontologies: ⊆, ⊇, ⊥, named and sibling.

For example, for the semantic relation <*fog*, ⊆, *weather*>, which users have evaluated as a correct relation, we found two different matches in the SW: The ontology *http://morpheus.cs.umbc.edu/aks1/ontosem.owl* with the match <*fog*, ⊆, *weather*> and the ontology http://sweet.jpl.nasa.gov/ontology/phenomena.owl with the match <*fog, hasAssociatedPhenomena, weather*>.

Considering the semantic relation, its corresponding user evaluation and the relation matched in the ontology, we distinguish three different types of matches:

- *Correct matches*: they provide exactly the same relation that the users are considering true.
- *Incorrect matches*: they provide exactly the same relation that the users are considering false or a different relation to the one the users are considering true.
- *Unknown matches*: the rest of the cases in which we cannot determine if the ontologies are providing correct or incorrect information without a manual evaluation.

Table 3 summarizes the rules that we use to automatically judge the correctness of a match in online ontologies based on the value of the original relation (column 1) and the user evaluation of the original relation (column 2).

Table 3. Quality of identified matches

		Match quality		
Original relation	**User evaluation**	**Correct**	**Unknown**	**Incorrect**
⊆	True	⊆	Named, sibling	⊇, ⊥
⊇	True	⊇	Named, sibling	⊆, ⊥
⊥	True	⊥	Named	⊆, ⊇, sibling
named	True		named, sibling, ⊆, ⊇, ⊥	
⊆	False		⊇, ⊥, named, sibling	⊆
⊇	False		⊆, ⊥, named, sibling	⊇
⊥	False		⊆, ⊇, sibling, named	⊥
named	False		⊆, ⊇, ⊥, named, sibling	

For the 1452 semantic relations described in the five different datasets we have found 53726 matches in 283 online ontologies using the services provided by Watson. Following the classification mechanism described above, we have extracted 1498 correct matches from 140 different ontologies (O_{cm}), 2279 incorrect matches from 148 different ontologies (O_{im}) and 49949 unknown matches from 275 different ontologies(O_{um}). Note that the same ontology can fall within the three different subsets if it provides correct, incorrect and unknown mappings for the various semantic relations of the dataset.

6.2 Selecting correct and incorrect ontologies

The identified correct and incorrect matches will help us distinguish between two different subsets of ontologies: O_r, reliable ontologies when assessing the quality of a semantic relation and $O_{nr,}$ unreliable ontologies. In order to select these subsets of ontologies we try to maximize two different criteria: a) the number of matches generated by the ontology and b) over those matches, the percentage of correct ones in the case of O_r and incorrect ones in the case of $O_{nr.}$ Fig. 2 and 3 show the distribution of the ontologies meeting these two criteria. Note that in these figures the percentages are expressed on a per unit basis.

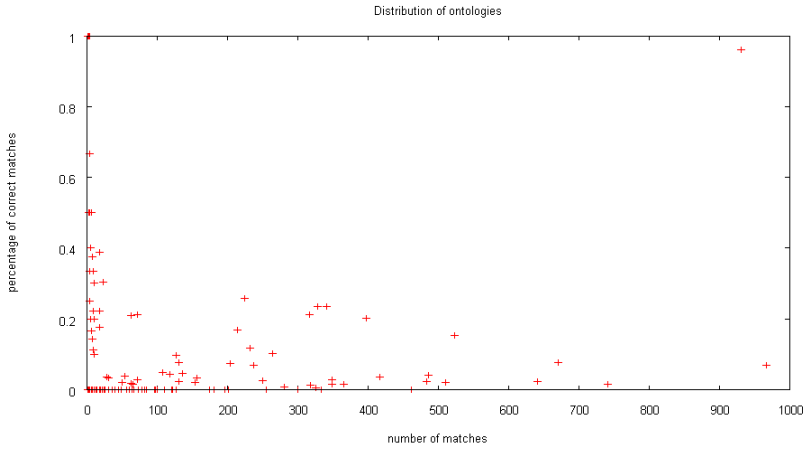


Fig. 2. Distribution of ontologies according to the number of matches and percentage of correct matches

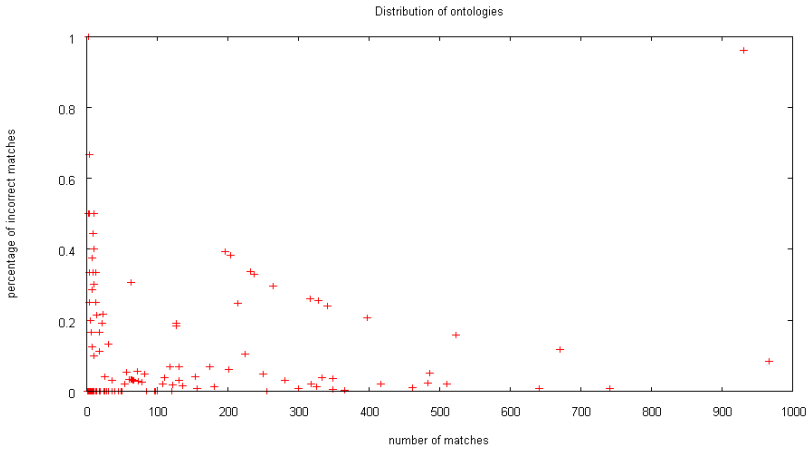


Fig. 3. Distribution of ontologies according to the number of matches and percentage of incorrect matches

As we can see in both figures, the percentage of correct and incorrect matches decreases in correlation with the increase in the number of matches. This is due to the fact that, for those ontologies that are able to provide a higher number of matches, the majority of identified matches have an unknown quality, i.e., we cannot determine if they are correct without a manual evaluation. This effect partially invalidates the criterion of maximizing the number of matches in order to select O_r and $O_{nr.}$ To avoid this effect we consider that: a) those ontologies that provide a number of matches greater than or equal to the average obtain the maximum score for this criterion and b) the criterion of maximizing the percentage of correct and incorrect matches should have slightly more relevance than the criterion of maximizing the number of matches. Considering these constraints we define O_r and O_{nr} as:

$$O_r = \{o_i \in O_{cm}, where, \alpha * \min(1, \frac{m_{oi}}{Avg_n(m_{oi})}) + (1-\alpha) * \frac{mc_{oi}}{m_{oi}} > \lambda\}$$

$$O_{nr} = \{o_i \in O_{im}, where, \alpha * \min(1, \frac{m_{oi}}{Avg_n(m_{oi})}) + (1-\alpha) * \frac{mi_{oi}}{m_{oi}} > \lambda\}$$

Where: m_{oi} is the set of matches found for the ontology o_i, mc_{oi} is the subset of correct matches found for the ontology o_i, mi_{oi} is the subset of incorrect matches found for the ontology o_i, n is the total number of ontologies that provided matches for the relations in our dataset (283), α is a constant parameter that determines the relevance for each criterion and λ is a certain threshold that discriminates the final subset of ontologies.

For our experiments α has been empirically set to 0.4, providing less relevance to the criterion of maximizing the number of matches. λ has been empirically set to 0.5 in the selection of O_r and to 0.6 in the selection of O_{nr} in order to obtain the top 40 ontologies for each subset ($|O_r| = |O_{nr}| = 40$).

A relevant aspect to consider in the selection of O_r is that we have discarded all the ontologies potentially involved in the generation of the experimental dataset in order to avoid biased information. An example of these ontologies is: http://oaei.ontologymatching.org/2004/Contest/228/onto.rdf.

6.3 Studying the discriminative effect of ontology quality measures

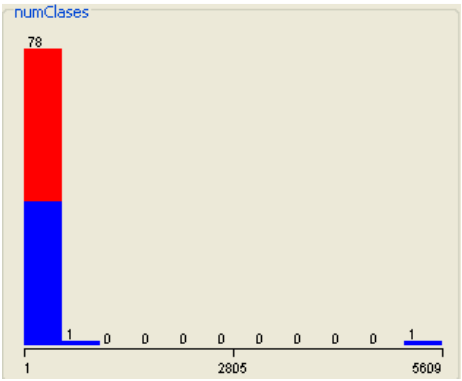
O_r and O_{nr} represent respectively reliable an unreliable online semantic content in the context of assessing the quality of semantic relations, i.e., O_r represents the subset of correct ontologies and O_{nr} the subset of incorrect ontologies. In this section we study how well the previously introduced measures (Section 5.1) are able to discriminate between these two types of ontologies. We therefore compute the measures for the 80 ontologies selected (40 belonging to O_r and 40 belonging to O_{nr}).

The analysis has been performed using the preprocessing tools of the Weka[11] data mining software. For each measure we present a figure that contains the ranges of values for the measure on the x-axis and the number of reliable versus unreliable ontologies that fall in each of these ranges on the y-axis. Reliable ontologies are presented in blue, and unreliable ones are presented in red.

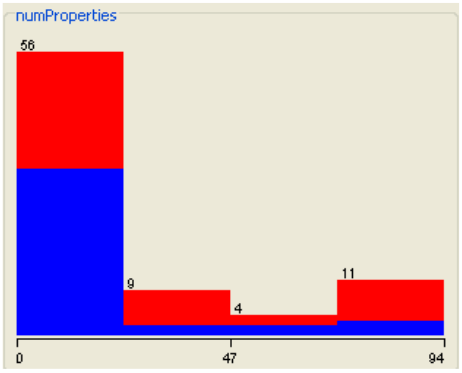
[11] http://www.cs.waikato.ac.nz/ml/weka/

6.3.1 Knowledge coverage and popularity measures

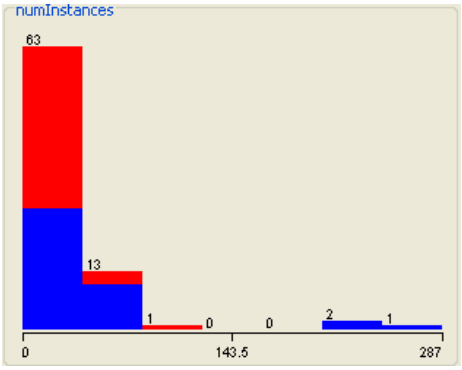
Number of classes



Number of properties



Number of instances



Ontology Direct Popularity

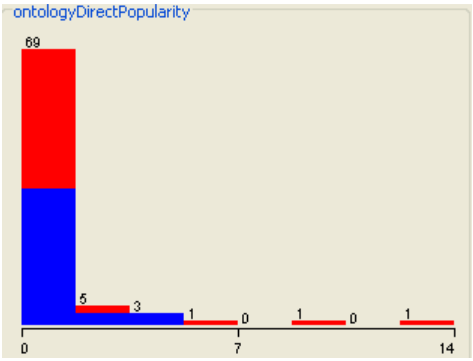


Fig. 4. Discriminative effect of the knowledge coverage and popularity measures

Fig. 4 shows the results obtained for the knowledge coverage and the population measures. As we can see in the figure, the number of classes in the ontologies varies between 1 and 5609. The higher percentage of ontologies contains between 1 and 1000 classes and this includes reliable and unreliable ones. Only two reliable ontologies present a number of classes higher than 1000 but this number of ontologies is not statistically significant to claim that ontologies with a higher number of classes provide more reliable semantic relations.

The number of properties varies from 0 to 94 in the selected subset of ontologies. We can see that reliable ontologies tend to have fewer properties than the unreliable ones on average. However, this measure does not draw a clear line between the two subsets of ontologies either.

The number of individuals varies between 0 and 287. While there is a small subset of reliable ontologies able to provide a higher number of individuals, again this is not

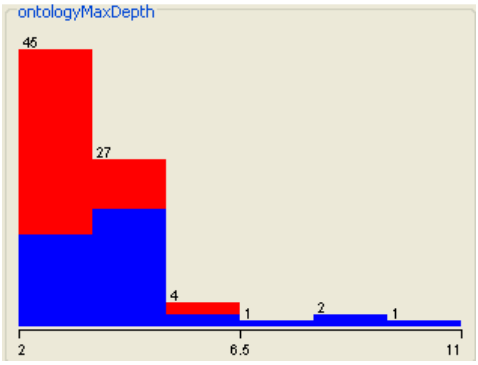
discriminative enough to consider that, in general, more populated ontologies provide better semantic relations.

The popularity measure varies between 0 and 14 imports per ontology. All ontologies with a popularity value higher than 6 are considered unreliable. However, there are only three ontologies in the dataset showing this effect, and therefore this measure cannot be considered discriminative either.

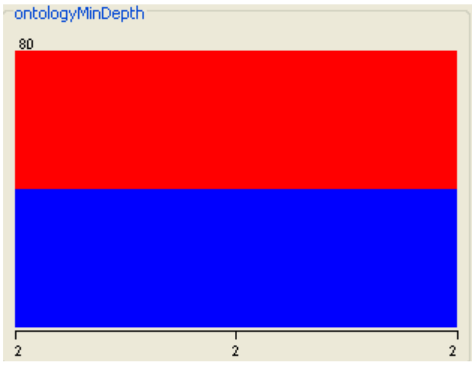
6.3.2 Structural ontology measures

Structural measures aim to study the topology of the ontologies, and more concretely their *depth* and *breadth*. We hypothesize that these measures can help us to better understand how conceptual relations are spread within the ontologies and therefore to determine which ontologies are better when assessing the quality of the relations.

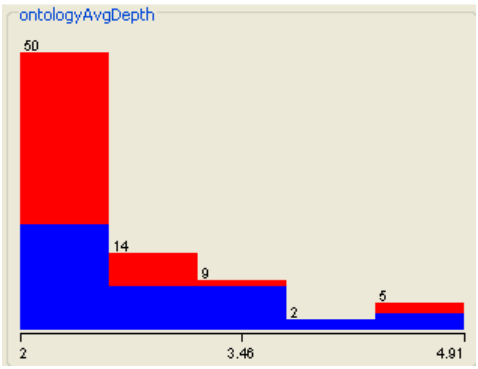
Maximum depth



Minimum depth



Average depth



Depth variance

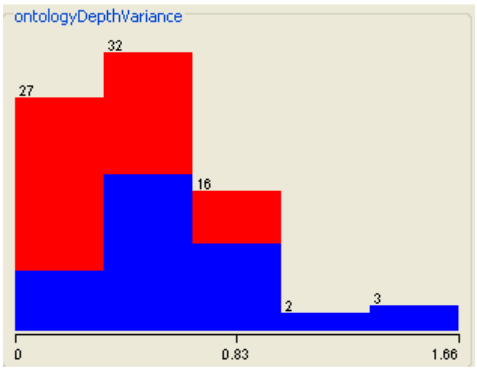
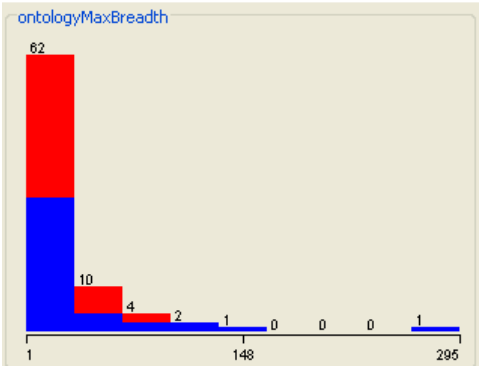


Fig. 5. Discriminative effect of the ontology depth measures

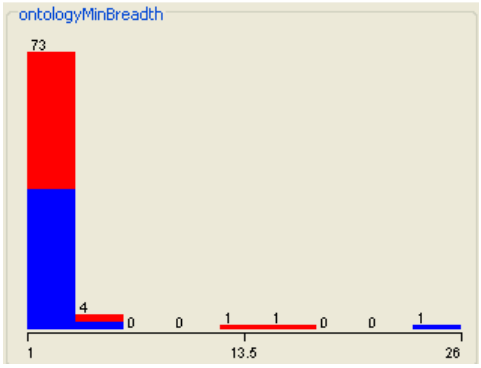
The first group of measures that we have considered for this evaluation is related to the depth of the ontology. As we can see in Fig. 5, the minimum depth is always 2 so this measure is not discriminative at all. However, the rest of the measures slightly

show that in general, those ontologies with higher levels of maximum depth, average depth and depth variance belong to O_r . Over the three measures we should highlight the ontology depth variance, since all ontologies with values higher than 0.9 are considered reliable. Even though these results are not statistically significant, there is a tendency that shows that those ontologies with higher depth variance can be considered "better" when assessing the quality of semantic relations.

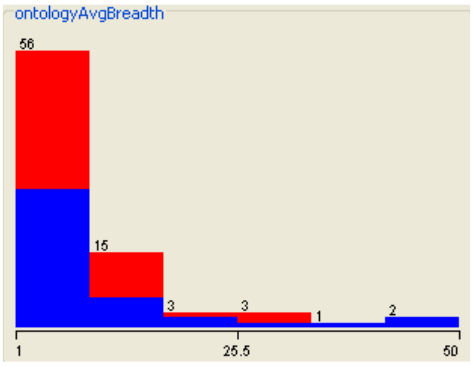
Maximum breadth



Minimum breadth



Average breadth



Breadth variance

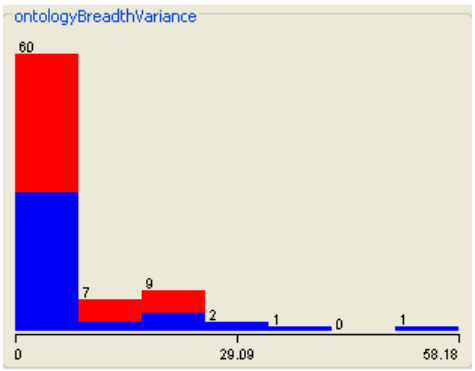


Fig 6. Discriminative effect of the ontology breadth measures

The second group of measures that we have considered in this study is related to the breadth of the ontology. Here the tendency is not as visible as in the case of the depth measures, but we can also see that all ontologies with maximum breadth values higher than 100 and breadth variance values higher than 20 always belong to O_r.

In summary we can conclude that, even though there is no statistically significant information to affirm that the topology characteristics of the ontologies are discriminative measures to distinguish reliable versus unreliable ontologies, there is a tendency showing that those ontologies that present higher values of depth and breadth variance are able to provide better semantic relations.

7 Conclusions

Understanding which ontology characteristics can predict "good quality ontologies" is a core and ongoing task in the SW. In this paper we studied the effect of several structural ontology measures to discriminate a "good ontology" in the context of a task-based evaluation, the assessment of correct semantic relations.

Our study shows that there is no statistically significant information to assure that these measures are able to identify the best semantic content in the context of this task. However, we have detected some tendencies which may show that the "best" ontologies are generally those that are more populated and have higher values of depth and breadth variance in their structure.

Several issues remain open nonetheless. On the one hand, the selection of O_r and O_{nr} (the correct and incorrect subsets of ontologies) can be biased by the high number of unknown matches (relations provided by the ontologies where we can only be sure if they are correct or not by means of manual evaluation). On the other hand, the datasets selected for this experiment only cover two domains: agriculture and academia. It would therefore be desirable to have more heterogeneous and completed evaluated datasets in order to discern with more accuracy if structural ontology measures can identify the best ontologies to assess the correctness of semantic relations.

References

[1] Alani, H., Brewster, C. Ontology Ranking Based on the Analysis of Concept Structures. In: Proceedings of the 3rd International Conference on Knowledge Capture (K-CAP 2005) (2005)

[2] Alani, H., Brewster, C., Shadbolt, N. Ranking Ontologies with AKTiveRank. In: Proceedings of the 5th International Semantic Web Conference (ISWC 2006) (2006)

[3] Brank, J., Grobelnik, M., Mladenić, D. A Survey of Ontology Evaluation Techniques. In: Proceedings of the Conference on Data Mining and Data Warehouses (2005)

[4] Brewster, C., Alani, H., Dasmahapatra, S., Wilks, Y. Data Driven Ontology Evaluation. In: Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004) (2004)

[5] Buitelaar, P., Eigner, T., Declerck, T. OntoSelect: A Dynamic Ontology Library with Support for Ontology Selection. In: Proceedings of the Demo Session at the 3rd International Semantic Web Conference (ISWC 2004) (2004)

[6] Burton-Jones, A., Storey, V., Sugumaran, V., Ahluwalia, P. A Semiotic Metrics Suite for Assessing the Quality of Ontologies. In: Data and Knowledge Engineering 55(1), pp. 84--102 (2005)

[7] d'Aquin, M., Baldassarre, C., Gridinoc, L., Angeletou, S., Sabou, M., Motta, E. Characterizing Knowledge on the Semantic Web with Watson. In: Proceedings of the 5th International EON Workshop (EON 2007) at the 6th International Semantic Web Conference (ISWC 2007) (2007)

[8] Ding, L., Finin, T., Joshi, A., Pan, R., Cost, R.S., Peng, Y., Reddivari, P, Doshi, V.C., Sachs, J. Swoogle: A Semantic Web Search and Metadata Engine. In: Proceedings of the 13th ACM Conference on Information and Knowledge Management (CIKM 2004) (2004)

[9] Ding, L., Pan, R., Finin, T., Joshi, A., Peng, Y., Kolari, P. Finding and Ranking Knowledge on the Semantic Web. In: Proceedings of the 4th International Semantic Web Conference (ISWC 2005) (2005)
[10] Fernández, M., Cantador, I., Castells, P. CORE: A Tool for Collaborative Ontology Reuse and Evaluation. In: Proceedings of the 4th International EON Workshop (EON 2006) at the 15th International World Wide Web Conference (WWW 2006) (2006)
[11] Guarino, N., Welty, C. An Overview of OntoClean. In: Staab, S., Studer, R. (eds.) Handbook on Ontologies, pp. 151--172. Springer, Heidelberg (2004)
[12] Hartmann, J., Sure, Y., Giboin, A., Maynard, D., Suarez-Figueroa, M.C., Cuel, R. Methods for Ontology Evaluation. In: Knowledge Web Deliverable D1.2.3 (2005)
[13] Lopez, V., Motta, E., Uren, V. PowerAqua: Fishing the Semantic Web. In: Proceedings of the 3rd European Semantic Web Conference (ESWC 2006) (2006)
[14] Lozano-Tello, A., Gómez-Pérez, A. ONTOMETRIC: A Method to Choose the Appropriate Ontology. Journal of Database Management 15(2), pp. 1--18 (2004)
[15] Maedche, A., Staab, S. Measuring Similarity between Ontologies. In: Proceedings of the 13th European Conference on Knowledge Acquisition and Management (EKAW 2002) (2002)
[16] Patel, C., Supekar, K., Lee, Y., Park, E. OntoKhoj: A Semantic Web Portal for Ontology Searching, Ranking, and Classification. In: Chiang, R.H.L., Laender, A.H.F., Lim, E.P. (eds.) Proceedings of the 5th ACM CIKM International Workshop on Web Information and Data Management (WIDM 2003), pp. 58--61. ACM Press, New York (2003)
[17] Sabou, M., d'Aquin, M., Motta, E. Exploring the Semantic Web as Background Knowledge for Ontology Matching. Journal of Data Semantics 11, pp. 156--190 (2008)
[18] Sabou, M., Fernández, M., Motta, E. Evaluating Semantic Relations by Exploring Ontologies on the Semantic Web. In: Proceedings of the 14th International Conference on Applications of Natural Language to Information Systems (NLDB 2009) (2009)
[19] Sabou, M., Garcia, J., Angeletou, S., d'Aquin, M., Motta, E. Evaluating the Semantic Web: A Task-Based Approach. In: Proceedings of the 6th International Semantic Web Conference (ISWC 2007) and the 2nd Asian Semantic Web Conference (ASWC 2007) (2007)
[20] Sabou, M., Gracia, J. Spider: Bringing Non-Equivalence Mappings to OAEI. In: Proceedings of the 3rd International Workshop on Ontology Matching (OM-2008) at the 7th International Semantic Web Conference (ISWC 2008) (2008)
[21] Sabou, M., Lopez, V., Motta, E., Uren, V. Ontology Selection: Ontology Evaluation on the Real Semantic Web. In: Proceedings of the 4th International EON Workshop (EON 2006) at the 15th International World Wide Web Conference (WWW 2006) (2006)
[22] Strasunskas, D., Tomassen, S. Empirical Insights on a Value of Ontology Quality in Ontology-Driven Web Search. In: Meersman, R., Tari, Z. (eds.) OTM 2008, Part II, LNCS 5332, pp. 1319--1337. Springer, Heidelberg (2008)
[23] Tartir, S., Arpinar, I., Moore, M., Sheth, A., Aleman-Meza, B. OntoQA: Metric-Based Ontology Quality Analysis. In: IEEE Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources, pp. 45--53. IEEE Computer Society, Los Alamitos (2005)