Evaluating the Semantic Web: A Task-based Approach

Marta Sabou[1], Jorge Gracia[2], Sofia Angeletou[1], Mathieu d'Aquin[1], and Enrico Motta[1]

[1] Knowledge Media Institute (KMi), The Open University, United Kingdom
{r.m.sabou, s.angeletou, m.daquin, e.motta}@open.ac.uk
[2] IIS Department,University of Zaragoza, Spain, jogracia@unizar.es

Abstract. The increased availability of online knowledge has led to the design of several algorithms that solve a variety of tasks by harvesting the Semantic Web, i.e., by dynamically selecting and exploring a multitude of online ontologies. Our hypothesis is that the performance of such novel algorithms implicitly provides an insight into the quality of the used ontologies and thus opens the way to a task-based evaluation of the Semantic Web. We have investigated this hypothesis by studying the lessons learnt about online ontologies when used to solve three tasks: ontology matching, folksonomy enrichment, and word sense disambiguation. Our analysis leads to a suit of conclusions about the status of the Semantic Web, which highlight a number of strengths and weaknesses of the semantic information available online and complement the findings of other analysis of the Semantic Web landscape.

1 Introduction

The recent growth of the Semantic Web [19] and the appearance of semantic search engines such as Swoogle [11] and WATSON [9] that allow quick access to online knowledge has had a considerable impact on the design of Semantic Web applications. Indeed, there is a trend to move away from applications relying on a small amount of manually selected semantic sources towards a *new generation of Semantic Web tools* which dynamically select, reuse and combine a multitude of heterogeneous, online available ontologies [20, 21]. This paradigm of *harvesting the Semantic Web* has also inspired novel ways of performing a variety of tasks. For example, Alani [1] proposes a method for *ontology learning* that relies on cutting and pasting modules from online ontologies relevant to a set of keywords. In [14] the authors rely on online ontologies to *disambiguate the senses* of keywords used in a search engine query. Dynamically selected online ontologies play the role of background knowledge in *ontology matching* [26, 27] or can be used to *semantically enrich folksonomy tag spaces* [4, 31]. The experimental evaluations of these algorithms [4, 14, 27] are still at an early stage, but, nevertheless, they provide strong evidence that the Semantic Web has reached a critical point where it can be used as a valuable source of knowledge to perform a variety of tasks.

Our hypothesis is that an important benefit of such novel algorithms lies in their potential use for *evaluating the Semantic Web*. Indeed, because they reuse a multitude of online ontologies, they can provide valuable insights into the qualitative aspects of these ontologies such as their suitability for a task, the properties of their vocabularies or the quality of their conceptual structure.

Such a task-based evaluation of the Semantic Web complements current efforts for evaluating (online) ontologies. Ontology evaluation has been a core research topic from the early stages of the Semantic Web leading to a set of approaches [6, 16] distributed in two major categories. On the one side, a few approaches exist, which are based on the manual assessment of a set of ontology design criteria (e.g., OntoClean [15]). On the other side, there are many automatic approaches, which evaluate different aspects of an ontology (e.g., vocabulary, conceptual structure) by relying on different views of what constitutes a good "quality" ontology [28]. For example, the quality of an automatically learnt ontology can be judged in terms of its similarity to a manually constructed ontology or to the corpus from which it was extracted. Or, adopting a task-based view, the quality of an ontology can also be judged with respect to the performance of a task that uses it [25].

With the growth of the Semantic Web, the focus of ontology evaluation efforts has shifted towards *online ontologies*. Ontology selection methods [28] rely on evaluating ontology aspects such as popularity [7, 11, 23], similarity to a domain or set of keywords [2, 7, 11, 23] and the richness of the internal structure [2, 7]. Furthermore, several overviews of the Semantic Web as a whole focus on the totality of online ontologies. Existing studies assess the size and growth rate of online knowledge [10, 19], as well as emerging trends in the adoption and use of representation languages and their primitives [5, 8, 34]. While these findings are important, they do not give an insight into the suitability of online ontologies to be used for certain tasks. Hence, inspired by the paradigm introduced in [25], we propose to perform a task-based evaluation of the Semantic Web by analyzing the performance of novel algorithms that harvest it.

We test the feasibility and usefulness of such a task based evaluation approach by detailing the lessons we have learnt about the quality of online ontologies when they were employed to solve three different tasks: ontology matching (Section 2), folksonomy tagspace enrichment (Section 3) and query disambiguation (Section 4). We conclude in Section 5 with a number of observations about the status of the Semantic Web that support our hypothesis and are complementary to findings provided by similar studies of online ontologies [5, 8, 10, 19, 34].

2 Case Study 1 - Ontology Matching

Ontology matching is the task of determining the relations that hold between the entities of two ontologies [30]. In [26] we proposed *a new paradigm to ontology matching* which relies on harvesting the Semantic Web: it derives semantic mappings by dynamically selecting, exploiting, and combining multiple and heterogeneous online ontologies. For example, when matching two concepts labeled

Researcher and *AcademicStaff*, the matcher would 1) identify (at run-time, during matching) online ontologies that can provide information about how these two concepts inter-relate and then 2) combine this information to infer the mapping. We distinguish two different strategies for deriving mappings [26]. In strategy S1 the mapping can be provided by a *single* ontology (e.g., stating that $Researcher \sqsubseteq AcademicStaff$). In strategy S2 a mapping can be derived by reasoning with information spread over several ontologies (e.g., that $Researcher \sqsubseteq ResearchStaff$ in one ontology and that $ResearchStaff \sqsubseteq AcademicStaff$ in another). We performed a large scale investigation and evaluation of this matching paradigm in [27] which provided a variety of insights into the quality of online ontologies as described next.

2.1 Experimental Data and Results

For experimental purposes we used two large, real life thesauri[3]. The United Nations Food and Agriculture Organization (FAO)'s **AGROVOC** thesaurus, version May 2006, consists of 28.174 descriptor terms (i.e., preferred terms) and 10.028 non-descriptor terms (i.e., alternative terms). The United States National Agricultural Library (NAL) Agricultural thesaurus **NALT**, version 2006, consists of 41.577 descriptor terms and 24.525 non-descriptor terms. We used both alternative and preferred terms in our experiments.

The matching process performed by using strategy S1 (see implementation details in [26]) resulted in a total of 6687 mappings (2330 subclass, 3710 superclass and 647 disjoint relations) obtained by dynamically selecting, exploring and combining 226 online ontologies. Fig. 1 shows the contribution of each of these ontologies to the alignment in terms of the number of mappings to which each ontology contributed and the percentage that this number represents.

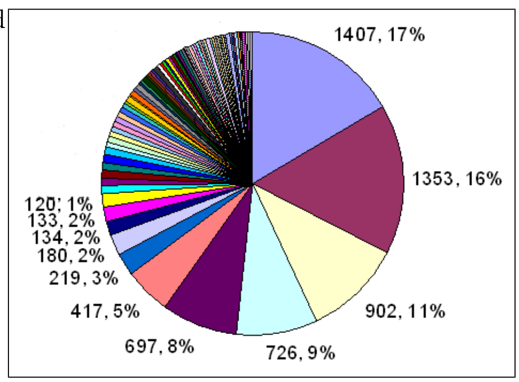

Fig. 1: Contribution of the online ontologies used by S1 to the alignment.

Conclusion C1: Online ontologies are useful for the matching task. Based on these results, we can already conclude that online ontologies are useful to solve real life matching tasks. Indeed, if combined appropriately, they can provide a large amount of mappings between the matched ontologies. Moreover, in the next section we will show that the quality of the knowledge provided by the Semantic Web allows us to produce a performance comparable with the best performers among alternative approaches to ontology matching.

[3] This data set was used in the "OAEI'06 *food* Thesaurus Mapping Task", `http://www.few.vu.nl/~wrvhage/oaei2006/`

2.2 Quality of Online Ontologies

According to [25], the essence of a task based evaluation is that the quality of an ontology correlates with the performance of the task in which it is employed. In the case of matching this means that the precision of the alignment is an indication of the quality of online ontologies explored to derive it.

To assess the quality of the knowledge provided by online ontologies we performed a manual assessment of 1000 mappings (i.e., 15% of the alignment). We relied on six members of our lab working in the area of the Semantic Web, and thus familiar with ontologies and ontology modeling. We performed two parallel evaluations of the sample mappings (i.e., each mapping has been evaluated by two different evaluators). The participants evaluated each mapping as *Correct*, *False* or *"Don't know"* when they could not judge the correctness of the statement. We computed the precision of the obtained alignment as the ratio of *Correct* mappings over all the evaluated mappings (i.e., those evaluated either as *Correct* or *False*) and obtained precision values of 63% and 69% for the two groups (see Table 1). To level out any differences, we also computed the precision of the fraction of the alignment on which both groups agreed (i.e., 742 mappings, 74%). We consider the so obtained precision value of 70% as a typical baseline performance that can be achieved by harvesting online knowledge.

	Group 1	**Group 2**	**Agreed by All**
Correct	586	666	525
False	346	299	217
Don't know	68	35	10
Precision	**63%**	**69%**	**70%**

Table 1. Evaluation results.

A manual inspection of the 217 false mappings on which both groups agreed revealed that 114 (i.e., 53%) are due to our simplistic anchoring mechanism (i.e., finding concepts in online ontologies that correspond to the matched concepts). For example, in Table 2, *c_6443* labeled with *Rams* and referring to an *uncastrated adult male sheep*[4] is put in correspondence with a similarly labeled concept (*ram*), but which means *Random Access Memory* in the context of the online ontology. An anchoring mechanism that could prevent us from deriving these false mappings (thus reducing their number to 103) will imply an increase in precision from 70% to 87%.

To fully understand the significance of these values, it is important to compare them to the performance of other background knowledge based techniques. However, the precision values we found in the literature were reported on different data sets, therefore we can consider them only as indicative, and in addition only S-Match reports on recall values [13]. The technique of Aleksovski et al. was evaluated on a Gold Standard of mappings for 200 concepts and produced a precision of 76% [3]. The matching techniques proposed by van Hage et al.

[4] Definition from WordNet2.1.

reach precision values of 53% - 75% when exploring a domain specific textual resource [32]. Therefore, the 70% precision value (which could be potentially increased to 87%) obtained by *dynamically selecting and combining multiple, heterogeneous and generic ontologies* correlates with the precision of the other two techniques (75% - 76%) when exploring a *single, high-quality, domain specific resource* (i.e., DICE [3], CooksRecipes.coms Cooking Dictionary [32]).

Conclusion C2: Online ontologies have a good quality and lead to high precision alignments. We conclude that online ontologies have a good enough quality to lead to alignments with a high precision value which can rival alignments obtained with manually selected, domain specific resources (e.g., ontologies, texts). Even more, our findings show that online ontologies don't only have a high quality when taken stand alone, but most importantly their *combined use* also results in high quality alignments.

Error Type	**Nr./ %**	**Examples**				
		AGROVOC Concept	**Labels**	**Rel.**	**NALT Concept**	**Labels**
Anchor	114, 53%	*c_6443*	Rams, Tups	$\sqsubseteq$	*memory*	memory
		O_1:ram $\sqsubseteq$ O_1:memory				
		O_1 = `http://www.arches.uga.edu/~gonen/qos_bilal.owl`				
Subsumption as generic relation	40, 18%	*c_3954*	Irrigation	$\sqsubseteq$	*agriculture*	agriculture
		O_1:Irrigation $\sqsubseteq$ O_1:SoilCultivation $\sqsubseteq$ O_1:Agriculture				
		O_1 = `http://sweet.jpl.nasa.gov/ontology/human_activities.owl`				
Subsumption as part-whole	16, 7%	*c_23995*	Branches	$\sqsubseteq$	*trees*	trees
		O_1:Branch $\sqsubseteq$ O_1:Tree				
		O_1 = `http://site.uottawa.ca/~mkhedr/FuzzyOnto`				
Subsumption as role	11, 5%	*c_11091*	Garlic	$\sqsubseteq$	*ingredients*	ingredients
		O_1:garlic $\sqsubseteq$ O_1:vegetable $\sqsubseteq$ O_1:ingredient				
		O_1 = `http://cvs.sourceforge.net/viewcvs.py/instancestore/instancestore/ontologies/Attic/pizza9.daml?rev=1.2`				
Inaccurate labeling	12, 5%	*c_1693*	Coal	$\sqsubseteq$	*industry*	industry
		O_1:coal $\sqsubseteq$ O_1:industry				
		O_1 = `http://www.aifb.uni-karlsruhe.de/WBS/meh/mapping/data/russia1a.rdf`				
Different View	12, 5%	*c_2943*	Fishes	$\sqsupseteq$	*lobsters*	lobsters
		O_1:Fish $\sqsupseteq$ O_1:MarineInvertebrate $\sqsupseteq$ O_1:Crustacean $\sqsupseteq$ O_1:Lobster				
		O_1 = `http://139.91.183.30:9090/RDF/VRP/Examples/tap.rdf`				

Table 2. Examples of several types of false mappings. For each mapping we show the names and labels of the matched concepts, the reasoning which lead to deriving the mapping and the online ontology from which the mapping was derived.

2.3 Frequent Errors in Online Ontologies

While the use of online ontologies generally leads to correct mappings there are also cases when false mappings are derived. Understanding the causes of false mappings can provide another interesting insight into the quality of online ontologies, namely, *the typical errors that lead to false mappings.* Our inspection of the false mappings revealed that 91 (i.e., 42%) are a direct consequence of the following types of errors in online ontologies (see Table 2).

Subsumption used to model generic relations. One of the most common errors was the use of subsumption as a way to model the fact that there exists some type of relation between two concepts, e.g., $Survey \sqsubseteq Marketing$, $Irrigation \sqsubseteq Agriculture$, $Biographies \sqsubseteq People$. This case leads to 40 false mappings (i.e., 18%).

Subsumption used to model part-whole relations. Subsumption is also used in several ontologies to model part-whole relations. These ontologies resulted in 16 (7%) incorrect mappings, e.g., $Branch \sqsubseteq Tree$, $Leaf \sqsubseteq Plant$.

Subsumption used to model roles. We found 11 false mappings (5%) derived because roles were incorrectly modeled as subclass relations, for example, that $Garlic, Leek \sqsubseteq Ingredient$ (in fact, $Leek$ is a $Vegetable$ but in some contexts it plays the the role of an ingredient).

Inaccurate labeling. We also found 12 cases (5%) when a correct subclass relation introduced errors due to the inaccurate labeling of its concepts. For example, O_1[5] states that $coal \sqsubseteq industry$, where *coal* refers to *coal industry* rather than the concept of *Coal* itself. Similarly, for $Database \sqsupseteq Enzime$ in O_1[6], *Enzyme* refers to an *enzyme database* rather than describing the class of all enzymes.

Different Views. Finally, some of the explored ontologies adopted a certain view on the relation of two concepts that was not in concordance with the context of the mapping and/or the perspective of the evaluators. For example, TAP considers *lobsters* kinds of *Fishes*, a perspective with which none of the evaluators agreed.

Conclusion C3: Online ontologies contain modeling errors which hamper the quality of the alignment. Most errors are due to the incorrect use of subsumption to model generic relations, roles and meronymy.

2.4 Contradictory Statements in Online Ontologies

The novelty of techniques that harvest the Semantic Web lies in their ability to combine information from multiple, different ontologies. As such, they need to deal with potentially contradictory information supplied by different sources. For example, in the case of ontology matching, contradictory mapping relations could be derived between two concepts by relying on different ontologies. The question is how frequent this phenomenon is, i.e., *do different online ontologies lead to contradictory mappings between two given terms?*

To answer this question, we ran a modified variant of S1: for every pair of concept labels we derive mappings from *all* the online ontologies that mention

[5] `http://www.aifb.uni-karlsruhe.de/WBS/meh/mapping/data/russia1a.rdf`
[6] `http://mensa.sl.iupui.edu/ontology/Database.owl`

AGROVOC label	NALT label	Nr. Subclass relations ($\xrightarrow{\sqsubseteq}$)	Nr. Superclass relations ($\xrightarrow{\sqsupseteq}$)	Nr. Disjunct relations($\xrightarrow{\perp}$)
fruit	tomato	0	3	1
sea	river	0	1	2
energy	light	0	1	1
meat	seafood	0	2	12
mushroom	pizza	1	0	1
sea	ocean	1	1	0

Table 3. Contradictory statements in online ontologies.

them. While we have discovered mappings between a high number of label pairs (6425), the number of cases when contradictory mappings are derived is surprisingly low and accounts to only six pairs (see Table 3).

Conclusion C4: Only few online ontologies contain contradictory relations between two given concepts. Our preliminary observations indicate that the correct mapping can normally be filtered out with simple statistical means: the most frequently derived relation is likely to be correct.

2.5 Inconsistencies in Multiple Mappings Drawn from Different Ontologies

In the previous section we have only looked at a rather basic form of contradiction, where contradictory relations have been explicitly stated between two items. As pointed out, these cases appear to be very infrequent. However, if we go beyond relations between two items and look at a number of mappings as a whole, then inconsistencies arise more frequently. Fig. 2 provides such an example, where $Vegetable_i$ is discovered to be disjoint with $Fruit\_(Food)_j$, $Tomatoes_j$ is a subclass of both concepts and thus *unsatisfiable*: there cannot be any instance of $Tomatoes_j$, since it would have to belong to two disjoint classes at the same time.

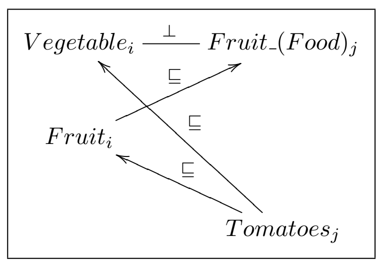

Fig. 2: Example of incoherence in mappings.

As already said, this phenomenon of generating sets of inconsistent mappings was more pronounced than expected. Indeed, our automatic incoherence detection mechanism has identified 306 base incoherences that corrupted the entire alignment. A few mappings between very generic concepts with many subclasses (e.g., $Foods \sqsubseteq Products$, $People \sqsubseteq Agents$) caused the majority of the incoherences. Fortunately, these can be isolated and disposed of automatically using reasoning mechanisms, thus leading to the improvement of the entire alignment.

Conclusion C5: Because different ontologies rely on different views or different contexts, they may contain contradictory information, leading to inconsistent sets of mappings.

3 Case Study 2 - Folksonomy Tagspace Enrichment

Social tagging systems[7] are highly successful due to the ease of the tagging process: users need neither to have prior knowledge or specific skills to use them [18, 35], nor to rely on a priori agreed structure or shared vocabulary. While folksonomies (i.e., lightweight structures that emerge from the tag space) are easy to create, they only weakly support content retrieval since they are agnostic to the relations between their tags: a search for `mammal` ignores all resources that are not tagged with this specific word, even if they are tagged with semantically related terms such as `lion, cow, cat`. Most approaches which address this problem [12, 29, 35] identify clusters of *implicitly* related tags (e.g., that `mammal` and `lion` are related). Specia and Motta [31] go one step further by proposing to make the semantic relations between tags *explicit* (e.g., that `mammal` is more generic than `lion`). They envision a semantic enrichment algorithm which complies with the paradigm of harvesting the Semantic Web by dynamically exploring and combining multiple online ontologies to derive explicit relations among implicitly interrelated tags.

A simplified version of the enrichment algorithm has been experimentally investigated in [4] by relying on the same implementation of relation discovery as used for ontology matching in [26] (i.e., strategy S1). Given a set of implicitly related tags, the prototype identifies subsumption and disjointness relations between them and constructs a semantic structure based on these relations. The first experiments on tag sets identified in [31] led to suboptimal results due to (1) the small size of the clusters (3-5 tags), (2) the low coverage of certain tag types in online ontologies and (3) the limitation of the software (it only identifies subsumption relations while most tags were related through generic relations). Therefore, we ran a second set of experiments on larger tag clusters identified with the Flickr API[8] around a handful of terms from domains that are well-covered by online ontologies. In these cases the process resulted in rich knowledge structures, such as for the tag cluster of *Fruit* in Fig. 3 (dotted lines denote disjointness). The general conclusion of the study is that while online ontologies can indeed be used to semantically enrich folksonomies, some of their characteristics hamper the process, as described next.

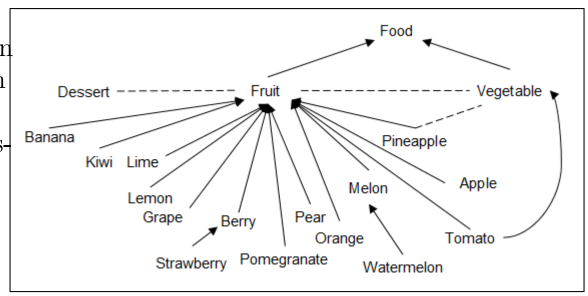

Fig. 3: Semantically enriched tag cluster for *Fruit*.

Conclusion C6: Online ontologies only weakly cover certain tag categories, as follows:

[7] E.g., Flickr (`http://flickr.com/`), del.icio.us (`http://del.icio.us/`).
[8] `http://www.flickr.com/services/api/flickr.tags.getRelated.html`

Novel terminology. Folksonomies are social artifacts built by large masses of people. They dynamically change to reflect the latest terminology in several domains and therefore greatly differ from ontologies which are usually developed by small groups of people and evolve much slower. As such, it is not surprising that many of the tags used in folksonomies, e.g., {`ajax, css`}, have not yet been integrated into ontologies[9]. Identifying such novel terminology has a great potential for the Semantic Web as it could represent a first step towards updating existing ontologies.

Scientific terminology (e.g., plant taxonomy) can only rarely be found in online ontologies. This could change however if large ontologies containing such information (e.g., AGROVOC) would be made available online.

Multilingual tags. Both Flickr and Del.icio.us (but especially Flickr) contain tags from a variety of languages and not only English. These tags are usually hard to find on the Semantic Web because the language coverage of the existing ontologies is rather low. Indeed, statistics[10] performed on a large collection of online ontologies (1177) in the context of the OntoSelect library [7] indicate that 63% of these ontolgies contain English labels, while a much smaller percentage contains labels in other languages (German 13.25%, French 6.02%, Portuguese 3.61%, Spanish 3.01%).

Photographic jargon. Because Flickr is a photo annotation and sharing site, many tags reflect terms used in photography, such as {`nikon, canon, closeup`}. Unfortunately, this domain is weakly covered in the Semantic Web.

Our study also found that, like in the case of ontology matching (C5), online ontologies can reflect different views and when used in combination can lead to inconsistencies in the derived structures. For example, the structure in Fig. 3 states that *Fruit* is disjoint with *Dessert*. The validity of this statement depends on the point of view we adopt since some would argue that fruits are desserts. Such different views can have more serious consequences. For example, *Tomato* is considered to be both a *Fruit* and a *Vegetable*. The first statement is valid in a biological context, since a tomato is the fruit of a tomato plant, however, normally one would classify tomatoes as types of vegetables. While such different views can co-exist, the fact that *Fruit* and *Vegetable* are disjoint makes the derived semantic structure inconsistent.

4 Case Study 3 - Word Sense Disambiguation

The goal of the Word Sense Disambiguation (WSD) task is to identify the appropriate sense of a word in a given context. Usually this task involves identifying a set of possible senses and then filtering out the right one based on some similarity algorithms. Existing approaches [17, 22, 24] exploit a given lexical resource (mainly WordNet) or ontology (small set of ontologies) as sources of word senses

[9] At the time of our experiments, March 2007.

[10] `http://olp.dfki.de/OntoSelect/w/index.php?mode=stats`

and then rely on one or more particular ontologies or corpora to compute semantic measures. Because they require the selection of appropriate knowledge sources a priori, these approaches are not suitable in cases when the domain of the words to be disambiguated is determined at run time. This limitation is addressed by a novel, unsupervised, multi-ontology WSD method [14] which 1) relies on dynamically identified online ontologies as sources for candidate word senses and 2) employs algorithms that combine information available both on the Semantic Web and the Web in order to compute semantic measures.

For example, suppose that we want to disambiguate *Java* in the context of *"Indonesia Java"*. In a first step, the algorithm identifies a set of possible senses[11] for each keyword by exploring online ontologies [12] and combines highly similar senses to avoid redundancies. Table 4 shows the candidate senses for *Java* and their characterization by their synonyms (i.e., Level 0) and superterms (Level 1, in this case direct hypernyms). A second, disambiguation step consists in computing a *Google based semantic relatedness* between *Indonesia* and each involved term (e.g., $Indonesia \leftrightarrow Java$, $Indonesia \leftrightarrow Island$) and combining the obtained values into a final [0,1] range score. The highest score indicates the most appropriate sense, i.e., $Java \sqsubseteq Island$ in our case. While the large scale evaluation of this method is still in progress, we can already summarize some qualitative conclusions about the characteristics of online ontologies.

#Sense	Type	Level 0	Level 1	Score
1	concept	Java	island	**0.387**
2	concept	java, coffee	drink	0.251
3	concept	java	programming language	0.116

Table 4. Disambiguation of *Java* in the context of *"Indonesia Java"*

Conclusion C7: Online ontologies provide a good source for word sense definitions. A major benefit of relying on multiple, online ontologies is that a much larger set of keyword senses can be discovered than in cases when few, predefined resources are used. For example, many traditional methods fail to disambiguate *developer* in *"UML handbook for developers"* because WordNet2.1 does not contain the word *UML* (acronym of Unified Modeling Language), neither the intended meaning of *developer* as someone who develops software. This information is however available in online ontologies: *UML* is a concept in the Book[13] ontology (subsumed by *SoftwareDesigns*), and *developer* is a property described as *"Developer of software"* in the DOAP[14] ontology. As evident from Table 5[15] this extra information discovered at runtime in DOAP is crucial for identifying the appropriate sense for *developer*.

---

[11] Defined by the ontological context of the term: synonyms, hypernyms etc.
[12] In addition to WordNet or any other local resource.
[13] `http://islab.hanyang.ac.kr/damls/Book.daml`
[14] `http://usefulinc.com/ns/doap`
[15] Level 1 here contains direct hypernyms for concepts and domains for properties.

#Sense	Type	Source	Description	Level 0	Level 1	Score
1	property	DOAP	"Developer of software for the project"	developer	project	**0.293**
2	concept	WordNet	"photographic equipment ..."	developer	photographic equipment	0.239
3	concept	WordNet	"someone who develops real estate"	developer	creator	0.230

Table 5. Disambiguation of *developer* in the context *"UML handbook developer"*.

Conclusion C8: Disambiguation results are influenced by modeling errors in online ontologies. However, due to its nature, the algorithm, is only partially affected by the typical ontology errors described in Section 2.3 (see Table 6). Indeed, the disambiguation algorithm uses part of the ontological context that characterizes a sense (e.g., subsumption, generic relations) in order to restrict the semantic field of the sense and to distinguish it from other senses of the same word. Such ontological information is used as a basis for relatedness computation and not exploited through formal reasoning as in the case of ontology matching. Therefore, the algorithm is not affected by the quality of formal modeling. For example, to characterize *branch* in its biological sense, an incorrectly modeled part-whole relation ($Branch \sqsubseteq Tree$) could lead to the same disambiguation result as using a correct subsumption ($Branch \sqsubseteq Stalk$). Also, *agriculture* could be an acceptable context to distinguish *irrigation* as *supplying dry land with water* from its medical sense. We conclude that error types 1 and 2 do not affect intrinsically the algorithm. On the other hand, the last three types of errors which associate a given term with other terms that do not reflect its sense have a major influence on the algorithm. For example, in $Enzyme \sqsubseteq Database$, the inaccurate labeling could give unpredictable results in the computed semantic measures. Also, the user could have different views from some online ontologies (error 5) and thus obtain an undesired result.

	Error Type	Example	Effect on algorithm?
1	Subsumption as generic relation	$Irrigation \sqsubseteq Agriculture$	No
2	Subsumption as part-whole	$Branch \sqsubseteq Tree$	No
3	Subsumption as role	$garlic \sqsubseteq ingredient$	Yes
4	Inaccurate labeling	$enzyme \sqsubseteq database$	Yes
5	Different view	$lobster \sqsubseteq fish$	Yes

Table 6. Sensitivity of disambiguation algorithm to frequent ontology errors.

Conclusion C9: Many online ontologies have a weak internal structure and thus hamper the performance of the method. For example, few online ontologies contain synonyms or non-taxonomic relations. We even found ontologies containing no relations at all. As a result, our algorithm can identify richly (e.g., extracted from WordNet) as well as poorly defined senses for the same word. Such uneven semantic characterization has a negative effect on the algorithm (which was built to compare similarly rich descriptions of senses) and can lead to suboptimal results. This insight in the general quality of online ontologies lead us to envision two important future changes. First, our tool should only rely on senses extracted from semantically rich ontologies which could be identified using a ranking mechanism such as AKTiveRank [2]. Second, the semantic measures we use should adapt to ontological contexts of variable richness (e.g., glosses should be given a high importance in ontologies with a poor taxonomy but rich in descriptions).

Conclusion C10: Parsing errors and broken links further hamper the functioning of the method. For example, from the 602 online ontologies identified for describing 25 terms randomly extracted from a list of frequently used keywords[16], 252 (42%) could not be correctly parsed into Jena[17] models due to parsing errors or broken links. Without being conclusive, this limited example illustrates the proportion of the problem.

5 Conclusions and Future Work

The hypothesis put forward in this paper is that novel algorithms which harvest online knowledge can facilitate a task based evaluation of the Semantic Web. Accordingly, we report on quality characteristics of online ontologies determined by analyzing the experimental results of three algorithms which solve divers tasks: ontology matching, folksonomy enrichment and WSD.

The major conclusion that we derive based on the content of our observations is that *online ontologies have a great potential for being used in combination to solve a variety of real life tasks*. Indeed, combining knowledge from multiple ontologies lead to a broad range of high quality mappings (C1) and to more word sense definitions during WSD (C7). In the case of ontology matching, we could also experimentally prove that the obtained alignment had a high precision, despite relying on more than 200 ontologies (C2). There are, however, some undesired effects caused by combining knowledge from multiple sources. Even if only in very few cases, contradicting statements can be obtained about two given concepts (C4). Then, the first two case studies were affected by the fact that online ontologies often reflect different views which can lead to incoherent knowledge structures when combined (C5). Overall, however, these findings deliver an important message: even at this early stage of development, the Semantic Web is a powerful source of background knowledge that can be exploited to successfully tackle real world tasks.

Besides providing task-centric conclusions, our approach also lead to observations about other aspects of online ontologies. At a syntactic level, several ontologies cannot be accessed due to parsing errors and broken links (C10). Second, regarding their vocabularies, online ontologies provide a weak coverage of certain types of folksonomy tags, such as novel terms, multilingual tags or scientific terms (C6). Third, we gained insight into major issues with the quality

[16] `http://www.google.com/press/zeitgeist_monthly.html`
[17] `http://jena.sourceforge.net/`

of the knowledge structures of online ontologies. We found that many have a weak (or no) structure and thus hampered the WSD method (C9). Even more worryingly, we identified a set of modeling errors, mostly related to the misuse of subsumption relations, which affected (to different degrees) both the formal reasoning based matching algorithm (C3) and the WSD process (C8).

While our conclusions provide a better understanding of the current state of the Semantic Web (complementary with the conclusions of other similar studies [5, 8, 10, 19, 34]), they could further benefit the research community as follows. First, we consider them as a proof that a task based evaluation is feasible and useful, thus supporting the hypothesis of the paper. Therefore, we wish to provide a more formal model for performing evaluations in this manner. Second, our findings have highlighted the need for novel evaluation methods that are capable to automatically identify more subtle characteristics such as the quality of the modeling [33]. Finally, these findings are valuable knowledge for those who wish to (re-)design algorithms that harvest the Semantic Web in a way that they maximally benefit from this rich and growing online knowledge repository.

Acknowledgements: This work was funded by the OpenKnowledge IST-FF6-027253 and NeOn IST-FF6-027595 projects, the CICYT TIN2004-07999-C02-02 project and by the Government of Aragon-CAI grant IT11/06.

References

1. H. Alani. Position Paper: Ontology Construction from Online Ontologies. In *Proc. of WWW*, 2006.
2. H. Alani, C. Brewster, and N. Shadbolt. Ranking Ontologies with AKTiveRank. In *Proc. of ISWC*, 2006.
3. Z. Aleksovski, M. Klein, W. ten Katen, and F. van Harmelen. Matching Unstructured Vocabularies using a Background Ontology. In *Proc. of EKAW*, 2006.
4. S. Angeletou, M. Sabou, L. Specia, and E. Motta. Bridging the Gap Between Folksonomies and the Semantic Web: An Experience Report. In *Proc. of the ESWC "Bridging the Gap between Semantic Web and Web 2.0" Wokshop*, 2007.
5. S. Bechhofer and R. Volz. Patching Syntax in OWL ontologies. In *Proc. of ISWC*, 2004.
6. J. Brank, M. Grobelnik, and D. Mladenic. A survey of ontology evaluation techniques. In *Proc. of the Conf. on Data Mining and Data Warehouses*, 2005.
7. P. Buitelaar, T. Eigner, and T. Declerck. OntoSelect: A Dynamic Ontology Library with Support for Ontology Selection. In *Proc. of the ISWC Demo Session*. 2004.
8. M. d'Aquin, C. Baldassarre, L. Gridinoc, S. Angeletou, M. Sabou, and E. Motta. Characterizig Knowledge on the Semantic Web with WATSON. Submitted for peer review, 2007.
9. M. d'Aquin, M. Sabou, M. Dzbor, C. Baldassarre, L. Gridinoc, S. Angeletou, and E. Motta. WATSON: A Gateway for the Semantic Web. ESWC Poster, 2007.
10. L. Ding and T. Finin. Characterizing the Semantic Web on the Web. In *Proc. of ISWC*, 2006.
11. L. Ding, R. Pan, T. Finin, A. Joshi, Y. Peng, and P. Kolari. Finding and Ranking Knowledge on the Semantic Web. In *Proc. of ISWC*, 2005.

12. G.Begelman, P. Keller, and F.Smadja. Automated Tag Clustering: Improving search and exploration in the tag space. In *Proc. of the Collaborative Web Tagging Workshop at WWW'06*, 2006.
13. F. Giunchiglia, P. Shvaiko, and M. Yatskevich. Discovering Missing Background Knowledge in Ontology Matching. In *Proc. of ECAI*, 2006.
14. J. Gracia, R. Trillo, M. Espinoza, and E. Mena. Querying the Web: A Multiontology Disambiguation Method. In *Proc. of ICWE*, 2006.
15. N. Guarino and C.A. Welty. An Overview of OntoClean. In S. Staab and R. Studer, editors, *Handbook on Ontologies*. Springer, 2004.
16. J. Hartmann, Y. Sure, A. Giboin, D. Maynard, M. C. Suarez-Figueroa, and R. Cuel. Methods for ontology evaluation. Knowledge Web Deliverable D1.2.3, 2005.
17. J. Hassell, B. Aleman-Meza, and I.B. Arpinar. Ontology-Driven Automatic Entity Disambiguation in Unstructured Text. In *Proc. of ISWC*, 2006.
18. A. Hotho, R. Jaschke, C. Schmitz, and G. Stumme. Information Retrieval in Folksonomies: Search and Ranking. In *Proc. of ESWC*, 2006.
19. J. Lee and R. Goodwin. The Semantic Webscape: a View of the Semantic Web. IBM Research Report, 2004.
20. E. Motta and M. Sabou. Language Technologies and the Evolution of the Semantic Web. In *Proc. of the Int. Conf. on Language Resources and Evaluation*, 2006.
21. E. Motta and M. Sabou. Next Generation Semantic Web Applications. In *Proc. of ASWC*, 2006.
22. R. Navigli and P. Velardi. Structural Semantic Interconnections: A Knowledge-Based Approach to Word Sense Disambiguation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(7):1075–1086, 2005.
23. C. Patel, K. Supekar, Y. Lee, and E. K. Park. OntoKhoj: A Semantic Web Portal for Ontology Searching, Ranking and Classification. In *Proc. of WIDM*, 2003.
24. T. Pedersen, S. Banerjee, and S. Patwardhan. Maximizing Semantic Relatedness to Perform Word Sense Disambiguation, 2005. Research Report UMSI 2005/25.
25. R. Porzel and R. Malaka. A Task-based Approach for Ontology Evaluation. In *Proc. of the ECAI Workshop on Ontology Learning and Population*, 2004.
26. M. Sabou, M. d'Aquin, and E. Motta. Using the Semantic Web as Background Knowledge for Ontology Mapping. In *Proc. of the Ontology Matching WS*, 2006.
27. M. Sabou, M. d'Aquin, and E. Motta. Using the Semantic Web as Background Knowledge for Ontology Matching: Investigating a New Paradigm. Submitted for peer review, 2007.
28. M. Sabou, V. Lopez, E. Motta, and V. Uren. Ontology Selection: Ontology Evaluation on the Real Semantic Web. In *Proc. of the EON Workshop*, 2006.
29. P. Schmitz. Inducing Ontology from Flickr Tags. In *Proc. of the Collaborative Web Tagging Workshop at WWW'06*, 2006.
30. P. Shvaiko and J. Euzenat. A Survey of Schema-based Matching Approaches. *Journal on Data Semantics*, IV, 2005.
31. L. Specia and E. Motta. Integrating Folksonomies with the Semantic Web. In *Proc. of ESWC*, 2007.
32. W. van Hage, S. Katrenko, and G. Schreiber. A Method to Combine Linguistic Ontology-Mapping Techniques. In *Proc. of ISWC*, 2005.
33. J. Völker, D. Vrandecic, and Y. Sure. Automatic Evaluation of Ontologies (AEON). In *Proc. of the ISWC*, 2005.
34. T.D. Wang, B. Parsia, and J.A. Hendler. A Survey of the Web Ontology Landscape. In *Proc. of ISWC*, 2006.
35. X. Wu, L. Zhang, and Y. Yu. Exploring Social Annotations for the Semantic Web. In *Proc. of WWW*, 2006.