Investigating the Use of Background Knowledge for Assessing the Relevance of Statements to an Ontology in Ontology Evolution

Fouad Zablith, Mathieu d'Aquin, Marta Sabou, Enrico Motta

Knowledge Media Institute (KMi), The Open University
Walton Hall, Milton Keynes, MK7 6AA, United Kingdom
{f.zablith, m.daquin, r.m.sabou, e.motta}@open.ac.uk

Abstract. The tasks of learning and enriching ontologies with new concepts and relations have attracted a lot of attention in the research community, leading to a number of tools facilitating the process of building and updating ontologies. These tools often discover new elements of information to be included in the considered ontology from external data sources such as text documents or databases, transforming these elements into ontology compatible statements or axioms. While some techniques are used to make sure that statements to be added are compatible with the ontology (e.g. through conflict detection), such tools generally pay little attention to the *relevance* of the statement in question. It is either assumed that any statement extracted from a data source is relevant, or that the user will assess whether a statement adds value to the ontology. In this paper, we investigate the use of background knowledge about the context where statements appear to assess their relevance. We devise a methodology to extract such a context from ontologies available online, to map it to the considered ontology and to visualize this mapping in a way that allows to study the intersection and complementarity of the two sources of knowledge. By applying this methodology on several examples, we identified an initial set of patterns giving strong indications concerning the relevance of a statement, as well as interesting issues to be considered when applying such techniques.

1 Introduction

A substantial part of ontology dynamics include learning and evolving ontologies from external data sources. Such ontology updates and changes mainly involve finding and selecting statements to be added to the ontology. Statement selection is sometimes based on conflict detection to keep the ontology coherent after adding such statements [8]. However, a conflict-free statement is not necessarily relevant to an ontology. Even if extracted from a carefully selected data source, some statements might concern other domains than the one covered by the ontology, might not fit the conceptualization encoded in it, or might simply not add any value to the ontology. Little attention has been paid in ontology evolution tools to the (semi-)automatic assessment of the relevance of statements to an ontology.

In this paper, the question we are investigating is: How to automatically assess the relevance of a statement with respect to an ontology? This is not an easy problem to formulate because the notion of relevance itself is not easy to define. Significant research focuses on relevance in various areas such as artificial intelligence and cognitive science [12, 13], or information retrieval [2]. It is acknowledged that relevance is relative to the context of an action [12], which can be a specific event, a process or series of processes. Here, we present a methodology aiming at investigating the use of the context in which a statement appears in external ontologies to assess its relevance to the local ontology being evolved. We identify the following sub-problems to be tackled as part of this methodology:

– *How to define the context of a statement?*
– *How do we relate this context to the one of the ontology to evolve?*
– *How do we explore this relation to identify indications of relevance (or non-relevance)?*

To answer these questions, we use online ontologies to provide contexts where statements are used. We extract the context of a statement from an ontology as the sub-graph of the ontology surrounding the entities linked through the statement. We then use matching techniques to align such a context with the ontology to evolve, relating common entities and relations in both graphs. Finally, we devise a visualization tool that provides an overview of this mapping. Using such a visualization, we can study the intersection of the context of a statement with the ontology to evolve, in order to identify clues and indications to support the assessment of the relevance of the statement to the ontology.

Applying this methodology leads to interesting results and raises a number of issues related to the use of the proposed approach. Indeed, looking at various examples, general patterns emerge from specific contexts that seem to indicate either the relevance or the non-relevance of the statement. Also, while the large variety of ontologies available on the Semantic Web provides a great source of contexts for this approach, it also leads to new research questions, concerning for example the use of contexts having different levels of granularity from the considered ontologies, or that apply different modeling principles.

In Section 2 we discuss the motivations behind this work, including an overview of the Evolva ontology evolution framework. In Section 3 we present our methodology for exploring the ontological context of a statement for assessing its relevance. We discuss our observations from applying this methodology in Section 4, leading to future work and conclusions, presented in Section 5.

2 Motivation

Interest in developing techniques for automatically and semi-automatically building ontologies is continuously increasing [1, 3, 10]. However, in most cases, it is left to the user to judge the relevance and quality of added knowledge. Providing relevance check techniques would be important for helping the user, especially in information intensive environments. Purely relying on statistical methods (e.g. TFIDF [3]) is not applicable in all domains, firstly because such methods require a large corpus to increase the accuracy of statistical measures, and secondly, because in many cases the source of new knowledge is not extracted from text documents (e.g. it can be from another knowledge base).

The approach we are proposing in this paper would help improving the accuracy and usability of many tools, e.g. the Watson [5] plugin and Evolva [14]. The Watson plugin for the NeOn toolkit[1] enables ontology builders to search and integrate statements from online ontologies. The user selects an ontology entity, then the Watson plugin returns a list of statements describing this entity. The user can browse these statements and choose which one to integrate in the base ontology. However, a lot of statements returned are not useful, and the user is responsible to select the ones that are appropriate (i.e., relevant) for the ontology. The solution discussed in this paper would make this process easier by automatically identifying the relevance of online statements with respect to the base ontology.

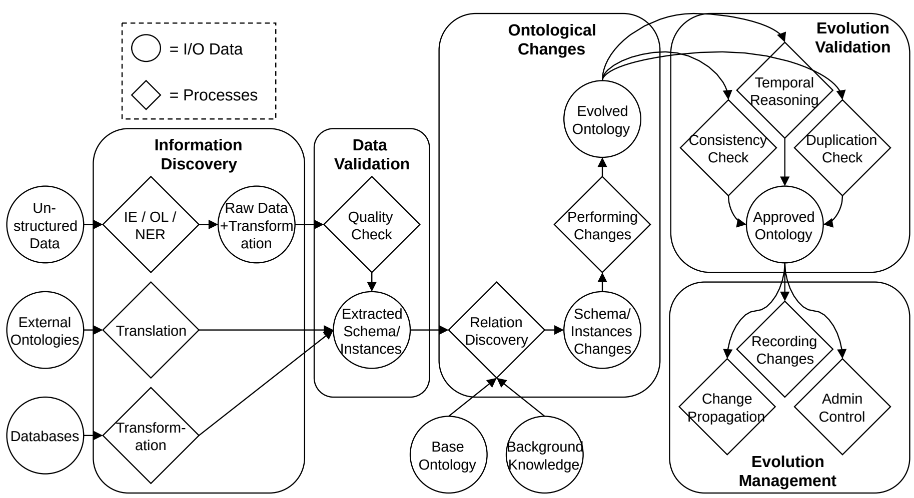

Fig. 1. Overview of the Evolva Framework.

Evolva is an ontology evolution tool that detects new information from external data sources, then links the new information to existing entities in the ontology using various sources of background knowledge. Figure 1 shows the framework components on which Evolva is based: in brief, the *Information Discovery* component analyzes external data sources (e.g. text) to identify potential ontology entities; the *Data Validation* component cleans the extracted data and

[1] http://www.neon-toolkit.org/

identifies the new entities that are not already in the ontology; the *Ontological Changes* component includes the "relation discovery" process that identifies links between the new entities and the ontology being evolved; the following components concern the validation and management of the changes applied.

In many cases, there are a lot of new statements, linking entities of the ontology to new entities, which are discovered and proposed to enrich the base ontology. However, not all of them are useful. For example, if a user is evolving an ontology in the academic domain, and the concept *Musician* occurs in the text document, Evolva would probably link *Musician* as a type of *Person* in the ontology. Of course the link by itself is correct and would not conflict with the ontology, but it's probably not relevant to add. With plenty of links to check, it's a burden on the user to manually select the relevant ones. Our work in this paper deals with assessing such relevance.

3 Exploring the Ontological Context of a Statement for Assessing its Relevance

One way to assess whether a statement is relevant to a particular ontology is to rely on additional information provided by background knowledge sources. More specifically, we consider that such background knowledge can be given by the contexts in which this statement has been used and applied. The main idea of our methodology is to explore the ontological contexts of statements–i.e., the contexts in which they are applied in other, external ontologies–to identify factors allowing to assess their relevance. In this work, it is assumed that the relevance of a statement relates to the added value of this statement with respect to a particular ontology. In other terms, we want to check whether the considered statement should be added to the considered ontology.

To analyze the use of ontological-contexts to assess the relevance of a statement s to an ontology O, three main tasks needs to be realized. First, such ontological contexts have to be discovered and extracted. We use a relation discovery engine on the Semantic Web, based on a Semantic Web gateway to identify online ontologies where s appears, and devise a technique to extract the surrounding of s in these ontologies (see Section 3.1). The result is a set of contexts $\{C_1, C_2, ..., C_n\}$ corresponding to sub-parts of ontologies surrounding the statement s. Second, these ontological contexts have to be aligned to the ontology O. We use simple matching techniques to identify common entities and relations in a context C_i and the ontology O (see Section 3.2). Then, the degree of overlap between the context of the statement and the ontology needs to be interpreted and translated into a relevance measure. However, this is not a trivial task. Therefore, in order to better understand what elements of the mapping might indicate relevance we perform a study of various cases of such mappings. For that purpose, we visualize the mapping between C_i and O and study their intersection and the complementarity in search for relevance factors. We implement a visualization tool that displays a merged graph based on this mapping, distinguishing clearly the common parts from the parts specific to the ontology and statement context (see Section 3.3).

3.1 Discovering and Extracting Statements' Contexts

In order to obtain ontological contexts in which a given statement appears, we use the Scarlet [11] relation discovery engine which is based on the Watson [5] ontology search engine. Watson collects, indexes and gives access to thousands of online ontologies, that can be searched and explored through its user interface, or through Web services. Scarlet exploits these Web services to find ontologies that (directly or indirectly) relate entities with each other. As a side effect, it can also identify in which ontologies these relations appear. In our case, considering a statement s of the form $s =< Subject, relation, Object >$, we use Scarlet to find ontologies that relate the entities *subject* and *object* through the relation $relation$.

Once the ontologies in which s appears have been identified, we extract the context surrounding s in each ontology. To achieve that, we traverse the existing relations between the entities *subject* and *object* linked by the statement s until a given recursion level is reached. This technique (which is very similar to the Prompt ontology view extraction feature [9] or some modularization techniques [6]) includes a number of parameters to customize the size and the content included in the context. In our case, we use separate parameters to limit the extraction depth of the entities' *parents*, *children* and other *named relations*.

As an example, Figure 2 shows the result of extracting the context of the statement $< Project, has - funding, Grant >$ from an online ontology[2], with a recursion level of 1 for all parameters.

3.2 Context Graph Matching

The goal of matching the statement context to the ontology, is to indicate how well the statement fits in the ontology. In other terms, we want to analyze the intersection, as well as the differences, between the context of the statement s and the ontology O, to derive potential factors indicating how much the considered statement would fit in O. The method we apply relates to the task of ontology matching [7], but considering only a sub-part of one ontology (the context). The purpose of our method is essentially different from ontology matching: while ontology matching aims to assess how similar two knowledge structures are, our aim is to identify how complementary the two structures are.

The approach we use for matching can be seen as a graph matching process. It starts by matching the names of the nodes (i.e., the entities) in the graphs of C_i and O using the Jaro-Winkler string similarity [4]. In a second step, it tries to align the edges of the graphs (i.e., the relations), first by comparing their names using the same similarity measure, and second by using the direction of the relation. We do not currently take into account the matching of relations in the analysis of our experiment.

[2] http://www.mindswap.org/2004/SSSW04/aktive-portal-ontology-latest.owl

Fig. 2. Example of ontological-context for the $< Project, has - funding, grant >$ Statement.

3.3 Visualizing Ontology-Statement Context Mappings

In order to analyze the mappings between statements' contexts and the ontology to evolve, we developed a tool to visualize such a mapping, clearly showing the intersection as well as the differences between the two graphs. The mappings resulting from the process described above allow us to divide the elements of the context and the ontology into three groups: 1- entities that are common to both C_i and O, 2- entities that are only present in C_i and 3- entities that are only present in O. Our visualization displays a unique graph based on these three groups of elements, using different shapes and colors to distinguish them. Entities from the first group are merged and displayed in green, and represented as star-shaped nodes. Entities from the second group are represented in red, with round-shaped nodes. Entities from the third group are represented in blue, with square-shaped nodes.

Figure 3 shows an example of the mapping visualization between the SWRC ontology[3], and the context of the statement $< Deliverable, subClassOf, Report >$ in an online ontology describing bibTeX entries[4].

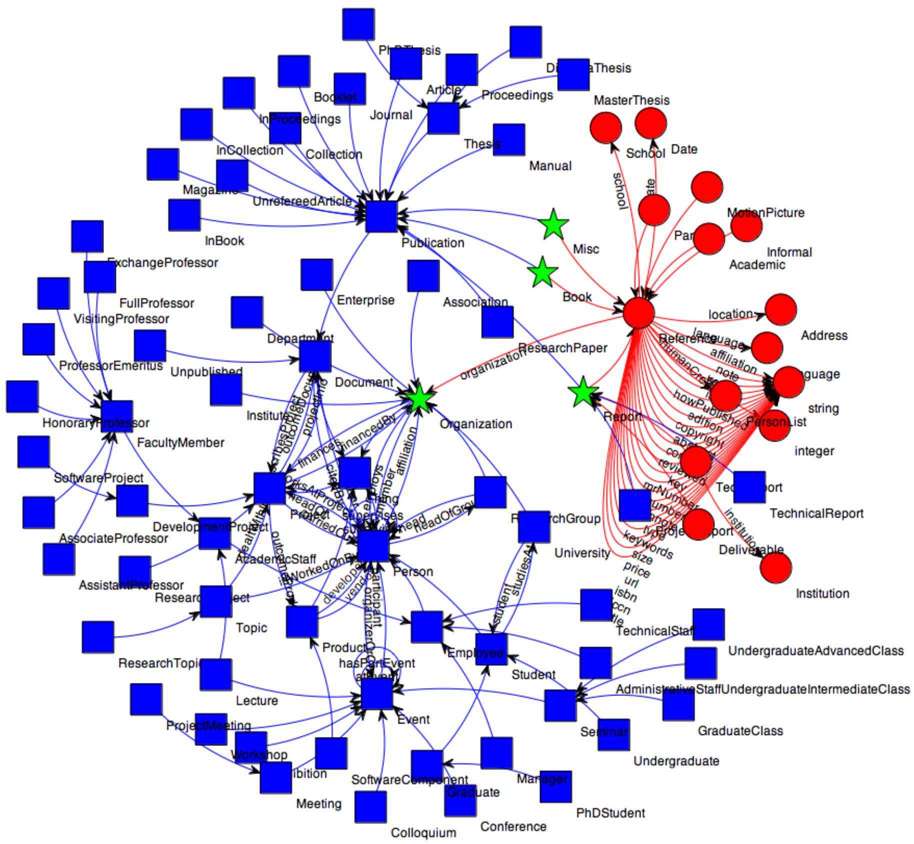

Fig. 3. Example of visualizing the mapping between the SWRC ontology and the $< Deliverable, subClassOf, Report >$ statement. Stars depict joint concepts, round nodes depict concepts that are only found in the statement's context and square nodes correspond to concepts only found in the ontology.

As can be seen from this example, depending on the size of the ontology and of the context, many elements might be displayed that are not useful to assess the relevance of a particular statement. Optional filters can then be applied to obtain a simpler visualization, keeping only the surrounding concepts of the

[3] http://ontoware.org/frs/download.php/354/swrc_updated_v0.7.1.owl
[4] http://oaei.ontologymatching.org/2004/Contest/103/onto.rdf

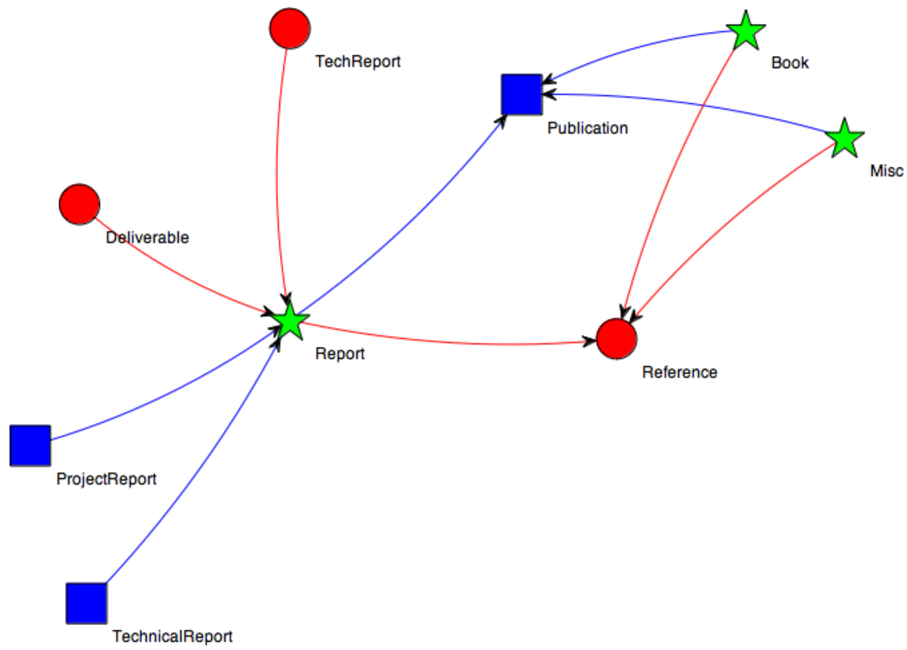

Fig. 4. Pattern 1 example and simplified visualization of the mapping between the SWRC ontology and the $< Deliverable, subClassOf, Report >$ statement.

statement's nodes that are common in the two graphs (see Figure 4 for the same example with this filter applied).

4 Experiment, Observations and Discussions

The ultimate goal of our work is to automatically assess the relevance of a statement to an ontology based on the level of matching between its context and the ontology (therefore exploiting the result of the techniques presented above). In preparation for devising a method that can translate a certain mapping situation into a relevance measure, we conducted a study of what factors could be taken into account for computing relevance. This study also provides a better understanding of the additional improvements required for such a structure matching based approach, and the issues that need to be tackled. Concretely, in the study presented in this section, we applied the techniques above to obtain contexts for 10 statements and then assessed relevance with respect to the SWRC ontology. We chose this set of statements so that it contained both relevant and irrelevant ones. Table 1 shows some examples of the selected statements.

In this section, we report on the observations we derived from analyzing the visualization of the corresponding mappings. We first discuss two "relevance patterns" we detected in these examples. These patterns correspond to general

Statement	Context Ontology	Relevant	Figure
$< Project, has-funding, Grant >$	http://www.mindswap.org/2004/SSSW04/ aktive-portal-ontology-latest.owl	Yes	2
$< Deliverable, subClassOf, Report >$	http://ontoware.org/frs/download.php/ 354/swrc_updated_v0.7.1.owl	Yes	3
$< Organization, subClassOf, >$ ActingEntity	http://www.atl.external.lmco.com/projects/ ontology/ontologies/basketball_soccer/ basketball.daml	Yes	5
$< Player, subClassOf, Person >$	http://www.atl.external.lmco.com/projects/ ontology/ontologies/basketball_soccer/ basketball.daml	No	5
$< Media, disjoint, Event >$	http://watson.kmi.open.ac.uk/ontologies/ LT4eL/CSnCSv0.01Lex.owl	No	N/A

Table 1. Example of Statements with Relevance with respect to the SWRC Ontology.

situations appearing in the mapping graphs and indicating that the statement is relevant. We then discuss particular phenomena that emerged in our examples and that should be taken into account when designing a statement relevance assessment method based on the ontological contexts of statements. Finally, we discuss general issues that point out further cases to investigate in our approach.

4.1 Identifying Relevance Patterns

We identified two patterns occurring in mapping graphs when the relation of the considered statement is *subClassOf*.

Pattern 1 - adding a sub-class to a joint concept. Pattern 1 occurs when the statement to assess includes a concept being a *subClassOf* a concept occurring in both the context and the ontology (see the example in Figure 4, where the considered statement is $< Deliverable, subClassOf, Report >$ and the shared concept is *Report*). The factor indicating relevance in this case is the fact that other sub-classes of the common concept are also shared between the context and the ontology, are close to each other (like $TechReport$ and $TechnicalReport$ in our case) or are somehow related to each other. Such a pattern clearly indicates the relevance of the statement, as it shows that, while the context and the ontology have many sub-classes of the same concept in common, the one to be added through the statement is missing in the ontology.

Conversely, if the conditions of the pattern do not hold for a particular context mapping, it also provides an indication of the irrelevance of the statement. Such a counter example is presented in Figure 5, where the statement considered is $< Player, subClassOf, Person >$. In this case, the concept to be added (*Player*) and its siblings are not related to the sub-classes of *Person* in the SWRC ontology.

Pattern 2 - adding a super-class to a joint concept. Pattern 2 is the reverse of Pattern 1. It occurs when the statement considered would add a super-

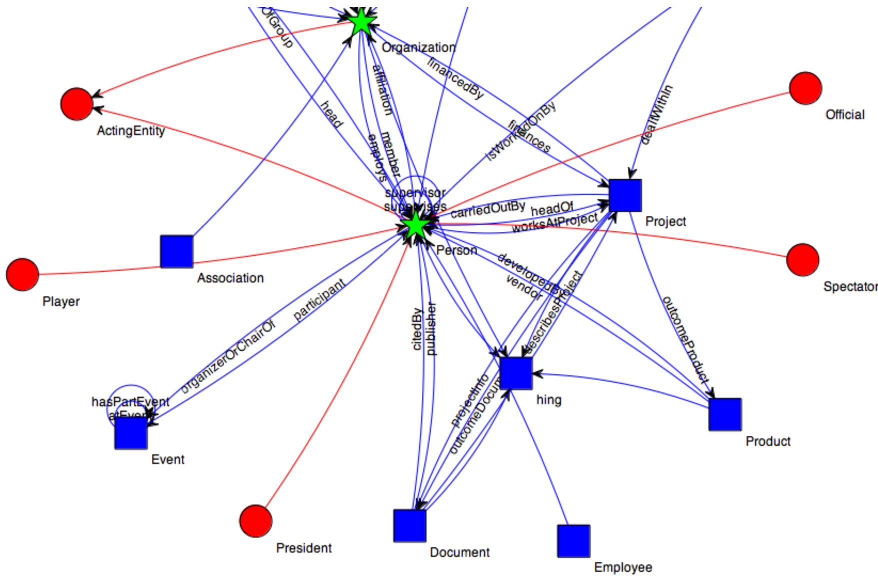

Fig. 5. Pattern 2 example for $< Organization, subClassOf, ActingEntity >$. Pattern 1 counter example for $< Player, subClassOf, Person >$.

class to a concept common to the context and the ontology. In this case, similarly, if the other sub-classes of the added concept are also occuring in both ontologies or are somehow related, this represents an indication of the relevance of the statement. For example, this pattern appears with the statement $< Organization, subClassOf, ActingEntity >$ in Figure 5 where the joint concept $Person$ is also a sub-class of $ActingEntity$, therefore re-inforcing its addition to the ontology.

Figure 4 provides another interesting case of Pattern 2 for $< Book, subClassOf,$ $Reference >$. The addition of this statement to the ontology is enforced by the fact that $Reference$ could be the super-class of two joint concepts. However, since $Reference$ and $Publication$ denote similar concepts, adding $Reference$ would be redundant and would add little value to the ontology. In contrast, $ActingEntity$ is sufficiently different from other concepts in the ontology to be worthwhile added.

4.2 Design Considerations

Studying examples of context mappings gave us the possibility to extract relevance patterns that can be used in an automatic statement relevance assessment tool. Above, we focused on the easy case where the statement concerns a hierarchical relation, but applying the same methodology on a larger number of examples, we expect to be able to identify such patterns for other types of relations as well.

However, in addition to identifying patterns, the examples we considered also made emerge a number of issues that would have to be taken into consideration in the design of a statement relevance assessment technique. The first one of these design consideration concerns the quality of the element level matching. Indeed, it appears clearly that, in the example Figure 4, using a more sophisticated matching would make the two nodes $TechnicalReport$ and $TechReport$ to be merged, thus making Pattern 1 to appear more clearly. In addition, while some of the nodes that are not in the intersection between the context and the ontology should not be matched, the fact that they are strongly related can be used as an additional information for relevance (e.g., in the same example, the nodes $Reference$ and $Publication$). We can use techniques based on the co-occurence of the terms on the Web to assess the relatedness of these terms.

Another consideration concerns the assessment of a statement as a whole, or as a part only. Indeed, so far, when talking about statements, we were referring to two nodes connected through a certain relation. However, an actual statement can be derived from the ontology, while not appearing explicitly in it. For example, the statement $< Interview, subClassOf, Event >$ is actually derived from the following path[5]: $< Interview, subClassOf, Human-Contact, subClassOf, Communicative-Event, subClassOf, Mental-Event, subClassOf, Event >$. An interesting phenomenon in such case is when part of the path that led to the considered statement in the context is already defined in the ontology. For example the statement $<$ *Ontology, developedBy, Organization*$>$, has the following path[6]: $<$ *Ontology, subClassOf, KAMethodology, subClassOf, Product, developedBy, Organization* $>$. The part $<$ *Product, developedBy, Organization* $>$ is already defined in the ontology. While this gives a hint on the relevance of the statement, it also indicates that it should not be assessed as a whole: only the part which is not common between the ontology and the context should be considered, as only this part will potentially be added to the ontology.

Finally, until now, we have assumed that the assessment of relevance would be realized one statement at a time. However, in scenarios like the one of Evolva, a set of statements would have to be considered. While the statement relevance assessment technique could be applied sequentially, it appears that the presence of other statements in the same context as the one considered could have an influence on its relevance. For example, we consider the statements $< Deliverable, subClassOf, Report >$ and $< Book, subClassOf, Reference >$. They both appear in the same context of Figure 4, giving a higher indication that the context strongly complements the ontology. These two statements would be assessed with a certain level of relevance separately, but if taken together, their relevance should re-enforce each other's: a statement should be considered more relevant if assessed in the same context as other relevant statements.

[5] Context ontology: http://morpheus.cs.umbc.edu/aks1/ontosem.owl
[6] Context ontology: http://139.91.183.30:9090/RDF/VRP/Examples/ka2.rdf

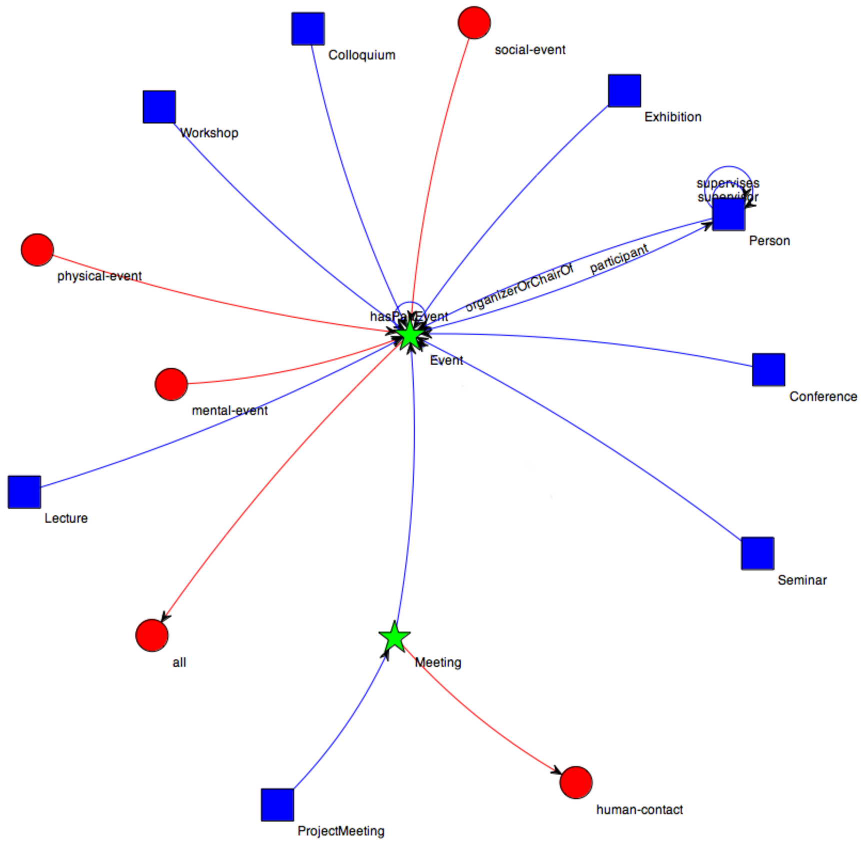

Fig. 6. Granularity discrepancy between the ontology and the statement context.

4.3 Further Cases to Investigate

In several examples, one general issue has been identified that needs special attention in the approach of using ontological-contexts from the Semantic Web to assess statement relevance: the heterogeneity of the contexts we can find and extract from online ontologies. First, different levels of granularity between the context and the ontology can affect the application of this approach. Indeed, considering the statement $< Human-Contact, subClassOf, Meeting >$ in the context mapping in Figure 6, we can see that, while the ontology has a direct and focused representation of events (workshop, conference, etc.), the context considers a more fine-grained representation of the same concept (starting, at the general level, from physical-event, mental-event, etc.) This makes it difficult to reconcile the context with the ontology and to understand the relevance of the given statement.

Another case of heterogeneity concerns modeling discrepancies between various ontologies. For example, in the SWRC ontology, the concept *Student* is modeled as a sub-class of *Person*. However, other online ontologies use a different representation pattern where student is not a sub-category of person, but is a role that a person can take (concretely represented by a *hasRole* relation between *Person* and *Student*). Assessing the relevance to the SWRC ontology of a statement involving a role in such a context becomes difficult as, even if from a general point of view the statement could be relevant, it does not fit the modeling pattern employed by the ontology, and so, does not really fit the ontology. However, additional processes could be considered to realize that the ontology and the context use different patterns, notify the user, and possibly transform the statement from one modeling approach to the other.

5 Conclusion and Future Work

Being able to assess the relevance of a statement to be added to an ontology is crucial in ontology evolution as well as for other tasks such as ontology learning or enrichment. However, such a task is generally left to the ontology developer, as it requires background knowledge to understand the added value of the statement to the considered ontology. In this paper, we envisaged an approach based on the use of the context in which a statement is used in external ontologies, to assess its relevance with respect to the ontology to evolve. We presented a methodology to study the relation between this ontology and ontological contexts extracted from online ontologies, with the goal to make relevance factors emerge from this study. Applying this methodology on several concrete examples taken from the Evolva scenario, we were indeed able to extract relevance patterns from the graph-based visualization employed to study the mapping between the ontology and the contexts in which the considered statements appear. We also observed several interesting phenomena, helping us to better understand how to design an automatic relevance assessment technique, as well as potential limitations of our approach that we need to overcome.

At short term, we plan to extend the present study to a higher number of examples to extract more relevance patterns and to implement the patterns we have already identified to test and validate our hypothesis that they can be used to automatically assess the relevance of statements with respect to an ontology. Ultimately, our goal is to develop a complete relevance assessment mechanism, to be included in the Evolva system. It is worth mentioning here that the techniques developed in this paper for ontology-context matching and visualization are intended to be reused in this tool, to provide the user not only with the final result of relevance assessment, but with all the information to justify such a result.

Acknowledgements

This work was funded by the NeOn project sponsored by the European Commission as part of the Information Society Technologies (IST) programme under grant number IST-FF6-027595.

References

1. E. Blomqvist. Pattern ranking for semi-automatic ontology construction. In *Proceedings of the 2008 ACM symposium on Applied computing*, pages 2248–2255. ACM New York, NY, USA, 2008.
2. P. D. Bruza and T. W. C. Huibers. A study of aboutness in information retrieval. *Artificial Intelligence Review*, 10(5):381–407, October 1996.
3. P. Cimiano and J. Volker. Text2Onto - a framework for ontology learning and data-driven change discovery. In *Proceedings of the 10th International Conference on Applications of Natural Language to Information Systems (NLDB 2005), Alicante*, pages 15–17, 2005.
4. W. W. Cohen, P. Ravikumar, and S. E. Fienberg. A comparison of string distance metrics for name-matching tasks. In *Proceedings of the IJCAI-2003 Workshop on Information Integration on the Web (IIWeb-03)*, 2003.
5. M. d'Aquin, C. Baldassarre, L. Gridinoc, M. Sabou, S. Angeletou, and E. Motta. Watson: Supporting next generation semantic web applications. In *Proceedings of WWW/Internet conference*, 2007.
6. M. d'Aquin, A. Schlicht, H. Stuckenschmidt, and M. Sabou. Criteria and evaluation for ontology modularization techniques. In *Modular Ontologies: Concepts, Theories and Techniques for Knowledge Modularization*, pages 67–89. Springer-Verlag, 2009.
7. J. Euzenat and P. Shvaiko. *Ontology Matching*. 2007.
8. P. Haase and L. Stojanovic. Consistent evolution of OWL ontologies. In *Proceedings of the Second European Semantic Web Conference, Heraklion, Greece, 2005*, pages 182–197, 2005.
9. N. F. Noy and M. A. Musen. Specifying ontology views by traversal. In *The Semantic Web ISWC 2004*, pages 713–725. 2004.
10. K. Ottens, M. Gleizes, and P. Glize. A multi-agent system for building dynamic ontologies. In *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*, pages 1–7, Honolulu, Hawaii, 2007. ACM.
11. M. Sabou, M. d'Aquin, and E. Motta. Exploring the semantic web as background knowledge for ontology matching. *Journal on Data Semantics*, (XI), 2008.
12. D. Sperber and D. Wilson. *Relevance*. 1986.
13. R. J. Sternberg. *Metaphors of mind*. 1990.
14. F. Zablith. Evolva: A comprehensive approach to ontology evolution. In *Proceedings of ESWC 2009: The Semantic Web: Research and Applications - PhD Symposium*, pages 944–948. 2009.