Relation Discovery from the Semantic Web

Marta Sabou
Knowledge Media Institute
The Open University, UK
r.m.sabou@open.ac.uk

Mathieu d'Aquin
Knowledge Media Institute
The Open University, UK
m.daquin@open.ac.uk

Enrico Motta
Knowledge Media Institute
The Open University, UK
e.motta@open.ac.uk

ABSTRACT

Several Semantic Web specific tasks such as ontology learning/extension or ontology matching rely on identifying relations between two given concepts. Scarlet[1] is a technique for discovering relations between two given concepts by exploring ontologies available on the Semantic Web as a source of background knowledge. By relying on semantic web search engines such as Watson, Scarlet *automatically* identifies and combines relevant information from *multiple* and *heterogeneous* online ontologies. Scarlet has already been used successfully to support a variety of tasks, but is also available as a stand alone component that can be reused in various other applications. This poster will be accompanied by a demo of Scarlet's functionality available through its Web based user interface.

Keywords

relation discovery, semantic web, Scarlet

1. INTRODUCTION

The task of discovering semantic relations between concepts (e.g., subsumption, disjointness or named relations) is core to several Semantic Web tasks such as ontology matching, ontology learning or ontology enrichment. As such it has received considerable attention and lead to the development of a variety of methods which mainly differ by the type of information that they explore in order to identify relevant relations. On the one hand, the majority of approaches from the ontology learning community have primarily focused on exploring textual sources for relation learning (see an overview of the state of the art in [4]). On the other hand, methods identified in the ontology matching community mostly rely on ontologies to derive new relations. Most commonly ontology matchers rely on the ontologies to be matched for deriving relations. Besides these, there are also a few technique that rely on background knowledge provided by generic ontologies such as WordNet [6] or manually selected domain ontologies [2]. The DBPedia relation finder also belongs to the category of methods which discover relations from structured data, as it explores DBPedia (an RDF representation of WikiPedia) to derive relations [8].

The novelty of Scarlet is that it performs relation discovery by exploring multiple online ontologies available as part of the Semantic Web. These are automatically selected depending on the current relation discovery task. As such, Scarlet is similar to approaches that harvest the Semantic Web i.e., they *automatically* find and explore *multiple* and *heterogeneous* online knowledge sources, to accomplish tasks such as ontology learning [1] or word sense disambiguation [7].

[1]`http://scarlet.open.ac.uk/`

2. THE TECHNIQUE

Scarlet follows the paradigm of automatically selecting and exploring online ontologies *to discover relations between two given concepts.* For example, when relating two concepts labeled *Researcher* and *AcademicStaff*, Scarlet 1) identifies (at run-time) online ontologies that can provide information about how these two concepts inter-relate and then 2) combines this information to infer their relation. In [10] we have described two increasingly sophisticated strategies to identify and to exploit online ontologies for relation discovery. Hereby, we only provide their intuitive description.

The first strategy, S1, derives a relation between two concepts if this relation is defined within a single online ontology, e.g., stating that $Researcher \sqsubseteq AcademicStaff$. Figure 1 illustrates this strategy with an example where three ontologies are discovered (O_1, O_2, O_3) containing the concepts `A'` and `B'` corresponding to `A` and `B`. The first ontology contains no relation between the anchor concepts, while the other two ontologies declare a subsumption relation. For a given ontology (O_i) the following derivation rules are used:

- if $A'_i \equiv B'_i$ then derive $A \xrightarrow{\equiv} B$;
- if $A'_i \sqsubseteq B'_i$ then derive $A \xrightarrow{\sqsubseteq} B$;
- if $A'_i \sqsupseteq B'_i$ then derive $A \xrightarrow{\sqsupseteq} B$;
- if $A'_i \perp B'_i$ then derive $A \xrightarrow{\perp} B$;
- if $R(A'_i, B'_i)$ then derive $A \xrightarrow{R} B$.

The second strategy, S2, addresses those cases when no single online ontology states the relation between the two concepts by combining relevant information which is spread over two or more ontologies (e.g., that $Researcher \sqsubseteq ResearchStaff$ in one ontology and that $ResearchStaff \sqsubseteq AcademicStaff$ in another). In this strategy, relation discovery is a recursive task where two concepts can be matched because the concepts they relate to in some ontologies are themselves matched. Figure 2 illustrates this strategy where no ontology is available that contains anchor terms for both A and B, but where one of the parents (P_2) of the anchor term A'_2 can be matched to B in the context of a third ontology

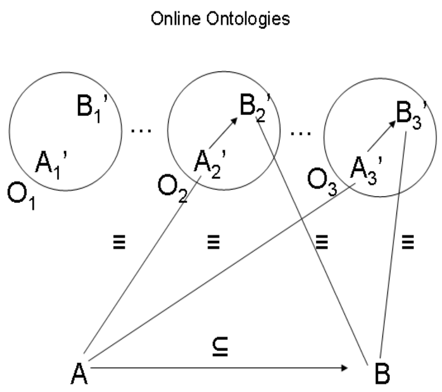

Figure 1: Derivation rules used by strategy S1.

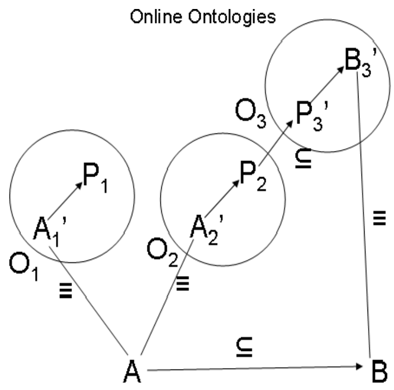

Figure 2: Derivation rules used by strategy S2.

(O_3). This strategy relies on a similar set of rules as S1. For a given ontology O_i the following rules are used:

1. for each P_i such that $A'_i \sqsubseteq P_i$, search for relations between P_i and B;
2. for each C_i such that $A'_i \sqsupseteq C_i$, search for relations between C_i and B;
3. derive relations using the following rules:
   - (r1) if $A'_i \sqsubseteq P_i$ and $P_i \xrightarrow{\sqsubseteq} B$ then $A \xrightarrow{\sqsubseteq} B$
   - (r2) if $A'_i \sqsubseteq P_i$ and $P_i \xrightarrow{\equiv} B$ then $A \xrightarrow{\sqsubseteq} B$
   - (r3) if $A'_i \sqsubseteq P_i$ and $P_i \xrightarrow{\perp} B$ then $A \xrightarrow{\perp} B$
   - (r4) if $A'_i \sqsubseteq P_i$ and $P_i \xrightarrow{R} B$ then $A \xrightarrow{R} B$
   - (r5) if $A'_i \sqsupseteq C_i$ and $C_i \xrightarrow{\sqsupseteq} B$ then $A \xrightarrow{\sqsupseteq} B$
   - (r6) if $A'_i \sqsupseteq C_i$ and $C_i \xrightarrow{\equiv} B$ then $A \xrightarrow{\sqsupseteq} B$

3. IMPLEMENTATION DETAILS

An initial prototype of both strategies has been implemented and evaluated by using the Swoogle'05 search engine (as reported in [11]). Recently, we have started re-implementing Scarlet to use the Watson semantic web gateway [5] for finding appropriate online ontologies. We have finalized the implementation of the first strategy and are currently working on the second, more complex, strategy. The implementation of the first strategy can be accessed through a Web based user interface. Additionally, a java distribution exists that can be integrated in other projects.

4. CURRENT APPLICATIONS

Scarlet originates from earlier work in the field of ontology matching, from the design of a matcher that exploits the entire Semantic Web as a source of background knowledge [9]. In essence, this matcher discovers semantic relations (mappings) between the elements of two ontologies by using the methods described above. A large-scale evaluation of this matcher lead to precision values of over 70% [10].

Scarlet's relation discovery functionality has also been used to semantically enrich folksonomy tagsets [3]. Given a set of implicitly related tags, we used Scarlet to identify relations between these tags and then merged them into a new knowledge structure (ontology).

Acknowledgements

This work was funded by the EC IST-FF6-027595 NeOn project. We thank Laurian Gridinoc for his help with the development of Scarlet's Web interface.

5. REFERENCES

[1] H. Alani. Position Paper: Ontology Construction from Online Ontologies. In *Proc. of WWW*, 2006.

[2] Z. Aleksovski, M. Klein, W. ten Katen, and F. van Harmelen. Matching Unstructured Vocabularies using a Background Ontology. In S. Staab and V. Svatek, editors, *Proc. of EKAW*, LNAI. Springer-Verlag, 2006.

[3] 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 Workshop on Bridging the Gap between Semantic Web and Web 2.0*, 2007.

[4] P. Cimiano. *Ontology Learning and Population from Text: Algorithms, Evaluation and Applications.* Springer, 2006.

[5] M. d'Aquin, C. Baldassarre, L. Gridinoc, M. Sabou, S. Angeletou, and E. Motta. Watson: Supporting Next Generation Semantic Web Applications. In *Proc. of WWW/Internet conference*, Vila Real, Spain, 2007.

[6] F. Giunchiglia, P. Shvaiko, and M. Yatskevich. Discovering Missing Background Knowledge in Ontology Matching. In *Proc. of ECAI*, 2006.

[7] J. Gracia, R. Trillo, M. Espinoza, and E. Mena. Querying the Web: A Multiontology Disambiguation Method. In *Proc. of ICWE*, 2006.

[8] J. Lehmann, J. Schuppel, and S. Auer. Discovering Unknown Connections - the DBpedia Relationship Finder. In *Proc. of CSSW*, 2007.

[9] 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.

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

[11] M. Sabou, J. Gracia, S. Angeletou, M. d'Aquin, and E. Motta. Evaluating the Semantic Web: A Task-based Approach. In *Proc. of ASWC/ISWC*, 2007.