Ontology Learning: Where are we?
And where are we going?

Paul Buitelaar

DFKI GmbH
Language Technology Lab & Competence Center Semantic Web
Saarbrücken, Germany

Ontology learning concerns the development of automatic methods for the extraction of a domain model from a relevant, i.e. domain-specific data set. In the context of ontology evolution, a specific domain model is already given and the task of ontology learning reduces to the extension or adaptation of this domain model on the basis of a changing underlying data set.

Ontology learning largely builds on methods previously developed in knowledge acquisi- tion, natural language processing and machine learning although with the specific purpose of automatically deriving an ontology, i.e. an explicit, shared and formally defined logical model. Unfortunately, the current state-of-the-art in ontology learning cannot be said to have reached this goal yet, although progress is made on various levels over the last couple of years.

Ontology learning is in fact not really one task but rather a collection of tightly connected subtasks that can be organized in a layered representation of increasing complexity, i.e. term extraction, synonym and translation detection, concept formation, instantiation, relation extraction, paraphrase and rule derivation, axiomatization. On each of these levels, methods and tools have been developed that address one or more subtasks. Methodologies are still needed however that address all subtasks in a coherent way and provide benchmarks for evaluation of methods on all levels, separately and in combination.

Ontology learning tools need to perform well on all levels of analysis, but even this is no ultimate guarantee for being actually useful. In addition to performance considerations, ontology learning tools need to be fully integrated into the knowledge engineering life-cycle, working in the background and providing the human domain expert with relevant input for ontology construction or evolution. Usability of ontology learning tools will thus be measured in terms of productivity of the human domain expert.

Ontology learning until recently has been based mostly on knowledge extraction from textual data, although some work has been done on extraction from tables and other structured data. Currently however, more and more semi-structured data becomes available in the form of Wikis and User Tags that shows a number of advantages for ontology learning as these data sets carry a lot of implicit knowledge (i.e. relations by linking or by social grouping) that can be more easily extracted than similarly implicit knowledge available in textual data. Additionally, more and more ontologies become publicly available that may be used as input by ontology learning tools, possibly in combination with knowledge derived from Wikis and User Tags and from more traditional textual data sets.

Ontology learning is a relatively new field of research, although building on long-standing methods in AI. In the developing context of the Semantic Web it is and will remain a central field of attention as ontologies form the semantic backbone of the Semantic Web, whereas their construction is complex and therefore knowledge- and cost-intensive. Automating this process through ontology learning thus remains an attractive proposition.