

After the vision of the Semantic Web was broadcasted at the turn of the millennium, ontology became a synonym for the solution to many problems concerning the fact that computers do not understand human language: if there were an ontology and every document were marked up with it and we had agents that would understand the mark-up, then computers would finally be able to process our queries in a really sophisticated way. Some years later, the success of Google shows us that the vision has not come true, being hampered by the incredible amount of extra work required for the intellectual encoding of semantic mark-up – as compared to simply uploading an HTML page. To alleviate this acquisition bottleneck, the field of ontology learning has since emerged as an important sub-field of ontology engineering.

It is widely accepted that ontologies can facilitate text understanding and automatic processing of textual resources. Moving from words to concepts not only mitigates data sparseness issues, but also promises appealing solutions to polysemy and homonymy by finding non-ambiguous concepts that may map to various realizations in – possibly ambiguous – words.

Numerous applications using lexical-semantic databases like WordNet (Miller, 1990) and its non-English counterparts, e.g. EuroWordNet (Vossen, 1997) or CoreNet (Choi and Bae, 2004) demonstrate the utility of semantic resources for natural language processing.

Learning semantic resources from text instead of manually creating them might be dangerous in terms of correctness, but has undeniable advantages: Creating resources for text processing from the texts to be processed will fit the semantic component neatly and directly to them, which will never be possible with general-purpose resources. Further, the cost per entry is greatly reduced, giving rise to much larger resources than an advocate of a manual approach could ever afford. On the other hand, none of the methods used today are good enough for creating semantic resources of any kind in a completely unsupervised fashion, albeit automatic methods can facilitate manual construction to a large extent.

The term ontology is understood in a variety of ways and has been used in philosophy for many centuries. In contrast, the notion of ontology in the field of computer science is younger – but almost used as inconsistently, when it comes to the details of the definition.

The intention of this essay is to give an overview of different methods that learn ontologies or ontology-like structures from unstructured text. Ontology learning from other sources, issues in description languages, ontology editors, ontology merging and ontology evolving transcend the scope of this article. Surveys on ontology learning from text and other sources can be found in Ding and Foo (2002) and Gómez-Pérez and Manzano-Macho (2003), for a survey of ontology learning from the Semantic Web perspective the reader is referred to Omelayenko (2001).

Another goal of this essay is to clarify the notion of the term ontology not by defining it once and for all, but to illustrate the correspondences and differences of its usage. In the remainder of this section, the usage of ontology is illustrated very briefly in the field of philosophy as contrasted to computer science, where different types of ontologies can be identified.

In section 2, a variety of methods for learning ontologies from unstructured text sources are classified and explained on a conceptual level. Section 3 deals with the evaluation of automatically generated ontologies and section 4 concludes.

934845-593 453 4535 ---- 3 @#%#\$\$^&\$ @#4@#%^~#^

934845-593 453 4535 ---- 3 @#%#\$\$%

934845-593 453 4535 ---- 3 @#%#\$\$^

() (&^ %\$#(:'''))

934845-593 453 4535 ---- 3 @#%#\$\$%, , , , ^&\$ @#4@#%^~#^

934845-593 453 4535 ---- 3 @#%#\$\$^&\$ @#4 , , , . . . @#%^~#^

934845-593 453 4535 ---- 3 @#%#\$\$^&\$ @#4@#%^~#^