

Report on the Sheffield Visit - 13.09-15.10

Marta Sabou

Vrije Universiteit Amsterdam, The Netherlands

1 Goal of the Visit

The goal of this research visit was implementing an ontology learning algorithm using the GATE framework. This experiment was a case study for assessing the support of GATE for implementing domain dependent ontology learning tools. A secondary goal was one of knowledge transfer, i.e. consulting and learning from the NLP experts of the Sheffield University since NLP is not a core expertise of the Vrije Univeristeit.

2 Topic of the visit

The broad topic of this visit is ontology learning. Ontology learning is an important research topic within the Semantic Web community as it promises alleviating the current knowledge acquisition bottleneck. Our focus is on ontology learning in the domain of web services. Our goal is to automatically learn domain ontologies that can be used in the semantic description of web services therefore facilitating the development of the Semantic Web Services field.

3 Results of the visit

3.1 Feedback on the extraction method

The members of the NLP group at Sheffield University provided insightful and useful advices for further developing our ontology learning methods.

3.2 Prototype implementation in GATE

As planned, the extraction prototype was entirely implemented in GATE.

Figure 1 provides a schematic representations of the major components in our extraction methods. Currently there is a wrapper that transforms Minipar’s output in annotation sets assigned to each document. While one still needs to run Minipar externally, the GATE group plan to fully integrate Minipar in GATE. Most of the extraction patterns were re-implemented in terms of Jape rules. They form the lexical extraction part of my method. Currently six patterns are implemented, where some pattern rely on the output of the previous patterns. It is still possible to implement 3-4 patterns which are less important (i.e. cover very rarely appearing lexical information). Besides these patterns, two other modules were implemented and integrated as processing resources in GATE. First, an ontology builder module centralizes the results of the lexical extractors and translates them into an ontology. It also maintains the link between the extracted concepts and the documents where these concepts appear. The ontology pruner module implements algorithms for filtering potentially un-interesting concepts from the ontology and removing them from the ontology. Both modules are still quite simple, however the evaluation of the extracted ontology suggested many improvements to these modules.

We have reused many of the components already existing in GATE, most notably classes that implemented Documents, AnnotationSets, Annotations, Corpora, ProcessingResources and Ontologies. This saved us a lot of implementation time.

3.3 Knowledge Transfer

During the visit several (15) research interviews were conducted. A research interview consisted in 1-2 hours discussion with one researcher, in which he gave a detailed presentation of his own research. Weekly department and GATE meetings were also valuable events that facilitated knowledge transfer.

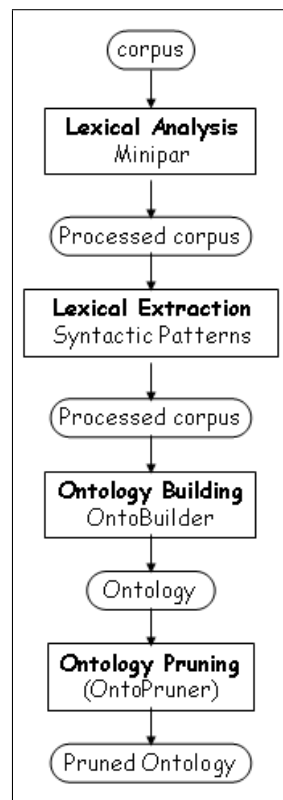


Figure 1: Extraction components.