A High-Performance Syntactic and Semantic Dependency Parser

Anders Björkelund, Bernd Bohnet, Love Hafdell, Pierre Nugues

This demonstration presents a high-performance syntactic and semantic dependency parser. The system is based on top performing systems of the CoNLL 2009 Shared Task. The full semantic pipeline achieves an average labeled semantic F1 of 80.90. The system consists of a pipeline of modules that carry out the following steps:

Input: Speculators are calling for a degree of liquidity that is not there in the market.

Output:
- A tokenized, tagged, lemmatized, and parsed sentence
- The interface shows the semantic annotations with one line per predicate

Architecture:
- Tokenization
- Part-of-speech tagging
- Lemmatization
- Dependency parsing
- Predicate Identification
- Predicate Disambiguation
- Argument Identification
- Argument Classification

Data:
- We use the CoNLL 2009 Shared Task data for training

Preprocessing:
- OpenNLP is used for tokenization
- We use a language-independent trainable lemmatizer and POS classifier
- The English lemmatizer has an accuracy of 99.46 and the German of 98.28
- The English tagger has an accuracy of 97.64 and the German one of 97.23

Dependency Parsing:
- The parser is an improved version of our contribution to the CoNLL 2009 Shared Task
- The dependency parser ranked first for English and German
- The improved version uses a hash kernel that improves the accuracy by 0.45 (English) and 0.62 (German) and the speed by a factor of 3.5
- The parser has a parallel feature extraction and parsing algorithm. Additional speed up by 3.4 on a 4 core CPU

Semantic Role Labeling:
- Each step uses different classifiers for nouns and verbs
- The classifiers are trained using L2-regularized linear logistic regression
- With an automatic predicate identification stage we obtain 80.90
- With an oracle identifying predicates (similar to CoNLL 2009 Shared Task), we get 85.58

Download page: http://code.google.com/p/mate-tools/

Demonstration pages:
http://barbar.cs.lth.se:8081/
http://en.sempar.ims.uni-stuttgart.de/
http://de.sempar.ims.uni-stuttgart.de/

Contact:
anders.bjorkelund@cs.lth.se
bohnet@ims.uni-stuttgart.de