Determining Named Entity Class Senses in WordNet and GermaNet

Goal
Add lexical information to Named-Entity classes in WordNet and GermaNet:
- 'name-bearing' or not: the fact that a word can refer to an object, person, etc. that can be named with an individual name
- 'role' or 'type': being a 'type' is to have a property persistently over time and being a 'role' is to have a property over a finite period of time

Data and Pre-processing
- 450 noun lemmas from WordNet
  - 300 classified as 'name-bearing'
  - 150 classified as 'not name-bearing'
- Data annotation by CrowdFlower
- Libraries used for editing WordNet:
  - JAWS (Java API for WordNet Search)
  - extJWNL (extended Java WordNet Library)

Classifier 1
Idea: Derive information from the gloss of the given noun sense and the WordNet hierarchy of hyponyms and hypernyms.
Approach: Automatic classification and evaluation with WEKA.
Features:
- Gloss content, e.g. 'someone' or 'somebody' are strong indicators for a name-bearing noun
- Ontological features from WordNet hierarchy, for example:
  - Number of hyponyms
  - Number and percentage of hyponyms with instances etc.

Classifier 2
Idea: Fetch the sentences out of UKWaC which contain our name-bearing nouns and use corpus features for classification.
Approach: Run the classifier in two classification steps and evaluate manually.
Features:
- Several regular expressions from Rudify like "is/was no longer a/the" which are indicators for 'role'
- A few ontological features from WordNet hierarchy:
  - Some lemmas are typically role, like 'worker'. If the current noun contains 'worker' it will be classified as 'role'
  - After the first step, all words are classified as 'type' or 'role'. However, if in the hypernym structure of a target word y another target word x is found that is classified as 'role', x will be classified as 'role' afterwards, too.

Evaluation
Classifier 1:
- development test set: average precision 75%
- test set: average precision 79%
  - F1-Measure 79%
  - Kappa: 0.57
  - Best Feature: Lexical file name

Classifier 2 (test set):
- 'role' precision 84%, recall 58%
- 'type' precision 64%, recall 42%
- overall F1-Measure 60%
- Best feature: Regular expressions from Rudify

Conclusion
There are several things to do in future. One of them was that we weren’t able to apply our data on WordNet and GermaNet until we could provide some good results. Nevertheless, we prepared all noun synsets from WordNet as a WEKA-classifiable .arff file and we also wrote the code for editing the dictionary and tried it out locally.

References
- Espresso: Leveraging Generic Patterns for Automatically Harvesting Semantic Relations
  - Patrick Pantel, Marco Pennacchiotti 2006
- Cross-Lingual Evaluation of Ontologies with Rudify
  - Amanda Hicks, Axel Herold 2011
- Automatically Harvesting and Ontologizing Semantic Relations
  - Patrick Pantel, Marco Pennacchiotti 2008

Note that we collected all WordNet synonyms of each sense from UKWaC, that's why there are more instances in Classifier 2.