**TnT POS Tagger**

**Background**

The TnT POS Tagger is based off the paper “TnT – A Statistical Part-of-Speech Tagger” by Thorsten Brants.

**Definitions**

One of the key parts of this tagger is combination of a Part of Speech tag and the capitalization of the word the tag occurs on. This will be referred to as a *PosCap* and is represented in the model class edu.umn.biomedicus.syntax.model.PosCap, and with the word as the model class edu.umn.biomedicus.syntax.model.WordPosCap.

**Training**

The TnT POS Tagger takes three models:

1. The Known Words Model that contains probabilities that words will occur conditional on a PosCap is implemented in the class edu.umn.biomedicus.syntax.tnt.KnownWordModel. It is trained by the class edu.umn.biomedicus.syntax.tnt.KnownWordModelTrainer.
2. The Suffix Model that contains the probability that a suffix will occur conditional on a PosCap is implemented in the class edu.umn.biomedicus.syntax.tnt.SuffixModel. It is trained by the class edu.umn.biomedicus.syntax.tnt.SuffixModelTrainer. It takes 3 variables
3. The PosCap Trigram Model that contains probabilities that a PosCap will occur conditional on the two previous PosCaps is found in the class edu.umn.biomedicus.syntax.tnt.PosCapTrigramModel. This is a smoothed probability, so it also factors the probability conditional on one previous PosCap, and the probability of a PosCap itself. The method of smoothing is described in the Brants paper and in the trainer class edu.umn.biomedicus.syntax.tnt.PosCapTrigramModelTrainer.

The known words and suffix models both implement the WordProbabilityModel interface and are interchangeable. Which one is used depends on whether a word is seen in the known words model. If it does have a known words probability, that is used, otherwise the suffix probability is used.

**Tagging**

**Included Trained models**

The included trained model is trained using

**UIMA resources and annotators**