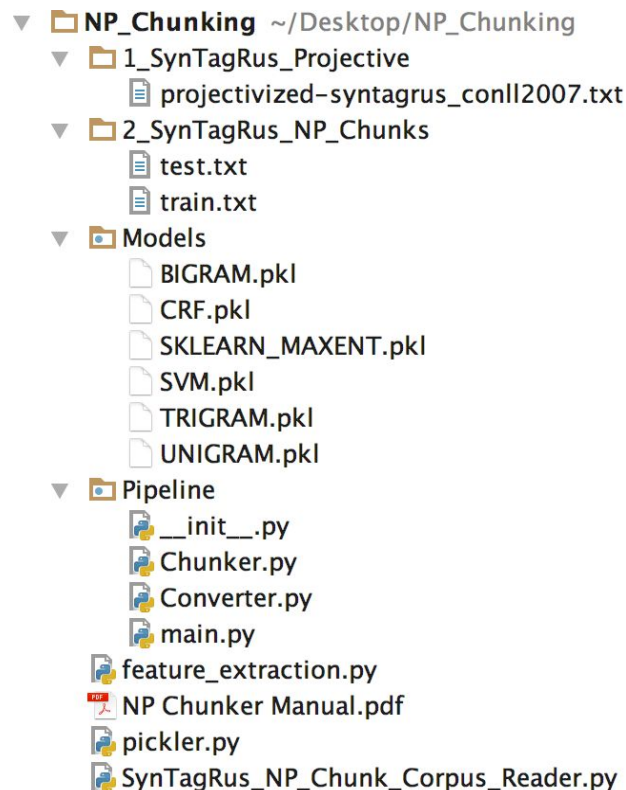


Code Structure and Usage



- *1_SynTagRus_Projective* contains the text file with the dependency trees of SynTagRus
- *2_SynTagRus_NP_Chunks* contains the test and train data for the Chunker
- *Models* contains the models that will be pickled during Chunker training
- The *Pipeline* package contains:
 - *Converter.py*: the code for converting the dependency trees from *1_SynTagRus_Projective* into the Conll-style train and test data, storing it in *2_SynTagRus_NP_Chunks*
 - *Chunker.py*: the Chunker class with the methods to extract the train and test data and apply the different models to it. A Chunker object must be instantiated inside *main.py* with one of the following parameters:
{'UNIGRAM', 'BIGRAM', 'TRIGRAM', 'SKLEARN_MAXENT', 'CRF', 'SVM'}
(See commented code in *main.py*)
 - *main.py*: the code used to train and evaluate models
- *feature_extraction.py* contains the code to extract the different features from data. Edit features inside the *features* method of this script
- *SynTagRus_NP_Chunk_Corpus_Reader.py*: adapted from nltk conll corpus reader to read converted train and test data
- *pickler.py* is basic code to pickle and load models

Required Installations

- *pycrfsuite* (<https://python-crfsuite.readthedocs.io/en/latest/>)
- *sklearn* (<http://scikit-learn.org/stable/>)