IT412: Natural Language Processing

Assignment 4: Text Chunking

Instructor: Prasenjit Majumder

Lab Outcome: After performing this assignment you will be able to do chunking and parsing.

1 Chunking

Chunking is a process of extracting phrases from unstructured text. Instead of just simple tokens which may not represent the actual meaning of the text, its advisable to use phrases such as "Albert Einstein" as a single word instead of 'Albert' and 'Einstein' separate words. Text chunking, also referred to as shallow parsing, is a task that follows Part-Of-Speech Tagging and that adds more structure to the sentence. Chunking works on top of POS tagging, it uses pos-tags as input and provides chunks as output. Chunking is very important when you want to extract information from text such as Locations, Person Names etc.

Figure 1 shows an sentence with its Part of speech tagging. Figure 2 shows chunking based on Part of speech tagging.

the	little	yellow	dog	barked	at	the	cat
DT	JJ	JJ	NN	VBD	IN	DT	NN

Figure 1: Sentence with POS tagging

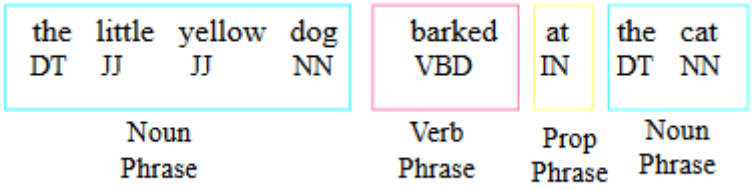

Figure 2: Chunked Sentence

2 Implementation

2.1 Dataset

For this assignment you are provided with train and test set of ConLL 2000 Chunking Competition. Create a parser to chunk given sentences into chunks as mentioned by ConLL competition.

2.2 Exercise

1. Create Parser for given train data and test your parser on test set. Compute Precision, recall and F1 score for test set.

3 References

- ConLL 2000 Shared task: Chunking
- http://deepdive.stanford.edu/example-chunking
- https://nlp.stanford.edu/software/lex-parser.shtml

3.1 Codes

- https://nlpforhackers.io/text-chunking/
- https://medium.com/greyatom/learning-pos-tagging-chunking-in-nlp-85f7f811a8cb