Named Entity Recognition (NER) Lab

Objective :

To use an already trained NER model to extract named entities from a text snippet.

NER Tool:

Python nltk ne_chunk library.

Task:

1. Download the python file “NER Exercise.py” from AutOnline.
2. Run it in PyCharm or an environment of your choice and ensure that it runs error free.
3. Examine the code to understand how it identifies the named entities or “chunks” the sample sentence.
4. Examine the output from the sample code.
5. **Your task:**
    a. Read in a newspaper article contained in an ASCII text file from a directory called “data” in your working directory and identify all the **persons** and **locations** mentioned.
    b. In addition you should also output the sentence number in which they were mentioned.
    c. Manually calculate the recall and precision for your results.
    d. Upload your work as the lab work for this session.
6. **Bonus task:**
    a. Extend your solution to read in a directory of txt files and output the filename and sentence number for all the person and location entities mentioned. For data, use about 5 articles from a newspaper site of your choice in a particular genre, such as current affairs, politics or sports.
7. Once you finish this lab you should have a good idea on how you might tackle the second assignment task.
8. Investigate coreference resolution using *Spacy/en_coref_lg* and *Stanford corenlp*.