coursera

Lab Help

Parts-of-Speech Tagging - Working with tags and Numpy

In this lecture notebook you will create a matrix using some tag information and then modify it using different approaches. This will serve as hands-on experience working with Numpy and as an introduction to some elements used for POS tagging.

In [1]:
```
import numpy as np
import pandas as pd
```

Some information on tags

For this notebook you will be using a toy example including only three tags (or states). In a real world application there are many more tags which can be found here (https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html).

In [2]:
```
# Define tags for Adverb, Noun and To (the preposition) , respectively
tags = ['RB', 'NN', 'TO']
```

In this week's assignment you will construct some dictionaries that provide useful information of the tags and words you will be working with.

One of these dictionaries is the `transition_counts` which counts the number of times a particular tag happened next to another. The keys of this dictionary have the form `(previous_tag, tag)` and the values are the frequency of occurrences.

Another one is the `emission_counts` dictionary which will count the number of times a particular pair of `(tag, word)` appeared in the training dataset.

In general think of `transition` when working with tags only and of `emission` when working with tags and words.

In this notebook you will be looking at the first one:

In [3]:
```
# Define 'transition_counts' dictionary
# Note: values are the same as the ones in the assignment
transition_counts = {
    ('NN', 'NN'): 16241,
    ('RB', 'RB'): 2263,
    ('TO', 'TO'): 2,
    ('NN', 'TO'): 5256,
    ('RB', 'TO'): 855
```