Homework 5: Co-occurrences

Dr. Benjamin Roth
Computerlinguistische Anwendungen

Due: Wednesday, June 05, 2019, 14:00

In this homework you will create a co-occurrence matrix and weight it with point-wise mutual information. You can check your progress using `unittest`s:
`python3 -m unittest -v hw05_cooccurrence/test_cooccurrence.py`

Exercise 1: Collecting the Cooccurrences [2 points]

Complete the function `cooccurrences(text, n)` that traverses through each position in a text, and counts all co-occurrences of the word at that position with neighboring n words (in each direction). This function returns a dictionary mapping word pairs to their counts.

Exercise 2: Creating a Cooccurrence Matrix [2 points]

Complete the function `cooc_dict_to_matrix(cooc_dict)` that turns the dictionary from the previous exercise into a Scipy sparse matrix.

Exercise 3: Computing the Point-Wise Mutual Information [4 points]

Complete the function `ppmi_weight(cooc_matrix)` that returns a PPMI weighted version of its input. Estimate the probabilities by using the counts of the matrix, as described in the docstring.