Explore how term-document matrices and weightings can be used for document classification. You will be attempting to distinguish between documents from different categories in the Brown corpus.

Use the provided script as a starting point. Before beginning, read and understand what it's doing. Then implement three sorts of document vectors:

1. Raw counts of terms in each document.
2. TF-IDF weighting, using the specific scheme described by Jurafsky and Martin (ch. 6).
3. Another weighting of your own invention/discovery. This may be another TF-IDF variant, or something else entirely!

You may use only built-in Python modules and numpy. You may work in a group of 1 or 2. Submissions will be graded without regard for the group size. **You should turn in a document describing the third method that you used and discussing all of the results.** There is no need to rehash the first two methods. The results/discussion should include a) the percent correct for each method, and b) a brief explanation of the relative performance, i.e. why does method A lead to better classification performance than method B? **You should also turn in 3 Python scripts, one for each of the above approaches.** These will be mostly the same and mostly consisting of the provided boilerplate.