**Homework 7**

To submit the homework, please upload an ipynb file to the Canvas Assignment.

The corpus contains 180 documents of movie reviews written by two reviewers. Each document is a review on one movie written by one reviewer. The first 80 reviews were written by Berardinelli and the rest 100 were by Schwartz.

**Problem 1**

Please calculate the entropy of the word “director” appearance in the 180 documents.

**Problem 2**

Calculate the mutual information between the “director” and the document author.

**Problem 3**

Find the top ten words with the highest mutual information with the document author and their respective mutual information. Explain (in Python comments) that what it means for a word by having a high mutual information with the document author.