

John Delshadi 75192867  
Ronnie Nguyen 22768608  
Jeffrey Fellows 34201703

Documents that have the term "conflict of interest"  
2,241

To find this answer, we used the file search feature on Macs to lookup all documents in the enron folder that contain the term conflict of interest. It doesn't care about cases already.

How many distinct terms are in this dataset  
2,622,925 terms for about 520,000 files

To find this term, we had actually done part of the developer section, but weren't able to finish it due to memory issues. We were able to run it for about 50,000 files, and ran it for 40,000, 30,000, 20,000, and 10,000 and used excel to find an equation to find it for larger numbers of files.

Estimated size of index\_plain.txt in MB  
83.30 MB

Like we mentioned above, we used the file we were able to create for the 50,000 files, and multiplied it by ten to get a file size that's representative of 520,000 files.