1. TrumpEmailPdfParser.py parses through all trump campaign emails in pdf form and reads them into a single text document (A).
   1. Uses ‘Tika’ module’s parser to deal w/ pdf
   2. For errors (unrecognized Unicode, etc.) it just skips.
   3. No inspection to see what the words are, and all hyperlinks are super long and unwieldy.
2. \_\_\_\_\_.py opens the Output file (A) and tokenizes words, replacing hyperlinks with (hyperlink)
3. \_\_\_\_\_.py tokenizes sentences; sentences can be grouped by purpose
   1. Call to action
   2. Attacks
   3. Ask for money

# Analysis

Tokenize words

Do a freq distribution of words, look at most common ones and see what stop words there are

Maybe remove stop words using NLTK’s stopwords corpus

Stem remaining words

Tag remaining words’ parts of speech

'From': '"Donald J. Trump" <contact@victory.donaldtrump.com>', 'Body': "\n++HYPERLINK++ \n\n++RECIPIENT\_FIRNAME++ for the past two weeks we've seen just how low the Democrats will sink in order to obstruct our agenda and everything we've accomplished since 2016.\nNancy Pelosi and Chuck Schumer shut down YOUR government by choosing illegal immigrants over hard-working American patriots like YOU. NO MORE!\nThe only reason Democrats shut down the government is because they know they can't win in 2020. They don't care about your safety, they only care about Presidential Harassment!\nWe showed them in 2016, now let's show them again that the silent majority is back and will NEVER be forgotten again.\nWe're going to remind each and every Democrat that the American people voted for border security, the rule of law and to BUILD THE WALL!\n\n++HYPERLINK++ \nI need you to add your name to publicly to stand with me in this fight. >> \n++HYPERLINK++ \nTELL DEMOCRATS TO BUILD THE WALL \n\nPlease confirm your information below:\n++HYPERLINK++ \n++RECIPIENT\_FULLNAME++ \n\n++HYPERLINK++ \n++RECIPIENT\_EMAIL++ \n++HYPERLINK++ \nSIGN MY NAME \n\nSign the official Build the Wall petition by 11:59 PM TONIGHT to get on the list we send to EVERY Democrats' office demanding they finally put Americans FIRST.\nRemember, when we stand united as Americans, we're unstoppable. N, you are a part of the greatest political movement in the history of mankind and it's called MAKE AMERICA GREAT AGAIN!\nI will not rest until the Democrats finally understand that Americans WILL ALWAYS COME FIRST!\nLet's show the Democrats the TRUE force of nature that is the silent majority.\nI need 1,OOO,OOO patriots to step up right now and add their name to publicly to stand with me: \n++HYPERLINK++ \n++HYPERLINK++ \nMAKE AMERICA GREAT AGAIN! \n\n++HYPERLINK++ \n\nDonald J. Trump\n\nPresident of the United States of America \n\n++HYPERLINK++ \nSIGN MY NAME \nContributions to the Trump Make America Great Again Committee are not deductible for federal income tax purposes.\n\nPaid for by Trump Make America Great Again Committee, a joint fundraising committee authorized by and composed of Donald J. Trump for President, Inc. and the Republican National Committee.\n\n You are receiving this email at ++RECIPIENT\_EMAIL++\nTrump Make America Great Again Committee, 138 Conant Street, 2nd Floor, Beverly, MA 01915\n\nWe believe this is an important way to reach our grassroots supporters with the most up-to-date information regarding the efforts of the Trump Make America Great Again Committee and President Trump, and we're glad you're on our team. It's because of grassroots supporters like you that we will Make America Great Again, and we appreciate your support. Thank you for all that you do!\n\n++HYPERLINK++ \nPrivacy Policy | \n++HYPERLINK++ \nUnsubscribe | \n++HYPERLINK++ \nUpdate Profile \n"}