Homework Number: 11

Name: Andrew Wu

ECN Login: wu1795

Due Date: 4/11/24

Recipe 1:

For recipe 1, I needed to find a way to see if there were multiple entries in the "From" header. After looking a few of the junk mail files, I realized that the entries were separated by a comma. I then created a regex that looked for a comma in the "From" header. To make sure that there would be no error with a random comma being placed. I more specifically looked for a closed bracket, a comma, and an open bracket, signifying that there is another entry in the "From" header.

Recipe 2:

For recipe 2, I looked at common words in the subject line of the specified junk emails. I chose words that either were very specific to the email or appeared in multiple emails, cutting down the number of words needed in the regex. I then created the regex, utilizing logic or for similar or misspelt words in the subject line.

Recipe 3:

For recipe 3, I once more looked at common words that appeared in all the junk emails, now in the body of the email. I then looked through the junk emails for the recipe and found two specific words that would capture those emails, the words being GEMSTONE and FREE. Since regex is case sensitive, the two words being in capital letters works well with trying to capture those emails without capturing any other emails.

Recipe 4:

For recipe 4, I looked more specifically at the characteristics of the junk emails. I realized that the junk emails all had content types multipart/alternative, text/plain, and text/html. I also realized that all the transfer encoding for the content were either 8bit, 7bit, or quoted-printable. I also noticed that personal emails had a different charset and utilized a recipe from the notes to allow personal emails to pass through. Through those observations, I created regex expressions to look for those characteristics in the emails to try and capture them into recipe 4.