

Yucheng Jiang

Report

The code for this assignment was written by python. To run the code, enter 'python SpamLord.py data dev/dev/ data dev/devGOLD' to the command line. I wrote 4 regular expressions, 3 for emails and 1 for phone numbers. They can catch all the e-mails and phone numbers in the devGOLD file without false positive or false negative errors. To deal with fax numbers, I utilized negative lookbehind assertion to exclude numbers that after word "fax". However, the contained pattern of negative lookbehind assertion must only match strings of some fixed length, so I handled it case by case by counting the number of characters between "Fax" and number, which means if there are additional test cases with fax number, my code may not work. To deal with phone number for lab, I used an if statement to exclude them.