



STUDYDADDY

**Get Homework Help
From Expert Tutor**

Get Help

Topics

This assignment will give you practice with line-based and token-based file processing, writing to a file, and String methods.

Instructions

You are going to write a program that will read the contents of a series of emails and determine which emails should be considered spam. The analysis will be printed in a summary report that is written to a file.

Email Messages

Each email message will contain:

- Sender
- Recipients
- Subject
- Message body
- `---eom---` (this exact String will be on a line of its own and designate the end of message)

For example:

```
From: Russell Wilson
To: Tyler Locket
cc:
bcc:
Subject: SP for PC
```

Hey,

The surprise party for Pete is coming up. Do we need to get anything else? We're down to the wire.

Let me know if I need to collect more funds from the team.

```
---eom---
```

Your program will use line-based file processing to access each line in the email message, but will need to use token-based processing to analyze the message body for each email.

emails.txt

The contents of all emails (one after the other) **will be stored in a other file called Sample**

cthat you can use for testing your program.

Analyzing each email

Each email will be analyzed to determine how likely it is to be spam. Our program is not very smart, so it simply counts the number of times that a spam-like word appears in the email. Words to look for include:

- offer
- wire
- bank
- fund
- transfer
- lottery

Your program should count the number of occurrences of these keywords in a single email. Note that keyword searching should be case-insensitive and the words may be partial words of a larger word ("fund" in "Fundraising" counts as an occurrence).

Consider the email above from Russell Wilson to Tyler Lockett, there are 2 keywords present "wire" and "fund" (in "funds").

Threshold for spam keywords

You should create a class constant at the top of your program. If the number of spam keywords for an email is greater than or equal to the threshold, then that message should be considered spam.

In the case of the email from Russell Wilson above, if the threshold is 2, the message would be considered spam. If the threshold is 3, the message would not be considered spam (since there are only 2 keywords in the email).

Writing the summary to a file

As you analyze each email, you should print to the summary to a new file called `summary.txt` using a `PrintStream`. The summary should include the subject of

each email; however, if an email is deemed spam, the marker ****SPAM**** should appear in front of the subject.

Check the other attached file called email....

```
Ignore the robots reading your emails...
```

```
    I ran out of cookies
```

```
    From the bottom of my heart...
```

```
**SPAM** Immediate Attention Requested
```

```
**SPAM** You're a winner!
```

```
**SPAM** Your trees are so happy!
```

```
    (no subject)
```

```
    Don't forget!
```

```
**SPAM** SP for PC
```

```
8 emails processed.
```

In order to print the subject of each email, you will need to "remember" this information from the beginning of the message until after the entire message is processed (the `---` `eom---` is reached).

Finally, you should print a count of the number of email analyzed.

Program Development

- You must break your program into a minimum of 3 methods, including the main. Each method should accomplish a specific task and be appropriately named.
- I recommend writing the program without the file output to start. Once you get that part working, then print it to the file instead of the console.
- Likewise, don't get caught up in remembering the subject right off the bat. Start by analyzing each email and print ****SPAM**** (or not) accordingly before working on the subject.



STUDYDADDY

**Get Homework Help
From Expert Tutor**

Get Help