Introduction to Computers and Programming
Homework 8 (Week 10)
Due date: 2020/12/3(Thu.)

1. Write a program to prompt for a file name, and then read through the file and look for lines of the form:

   X-DSPAM-Confidence:0.8475

   When you encounter a line that starts with “X-DSPAM-Confidence:” pull apart the line to extract the floating-point number on the line. Count these lines and then compute the total of the spam confidence values from these lines. When you reach the end of the file, print out the average spam confidence.

   Test your file on the mbox.txt and mbox-short.txt files. You can download the file from www.py4e.com/code3/mbox-short.txt and www.py4e.com/code3/mbox.txt

   Example:

   ```
   Enter the file name: mbox.txt
   Average spam confidence: 0.894128046745
   Enter the file name: mbox-short.txt
   Average spam confidence: 0.750718518519
   ```

2. You will use the mbox.txt and mbox-short.txt files. This program will prompt the user for a file name and then open the file and read the entire file looking for lines that start with "Subject: [sakai] svn commit:" like the following:

   Subject: [sakai] svn commit: r39772 - content/branches/sakai_2-5-x/ ...
   Subject: [sakai] svn commit: r39771 - in bspace/site-manage/sakai_2-4-x ...
   Subject: [sakai] svn commit: r39770 - site-manage/branches/sakai_2-5-x/ ...
   Subject: [sakai] svn commit: r39769 - in gradebook/trunk/app/ui/src ...

   These indicate the revision number of a source modification (r39770) and the file(s) that were modified. The output we desire is to extract information from each of these lines that includes the revision number and top-level folder where the modification was done:

   ```
   r39772 content
   r39771 bspace
   r39770 site-manage
   r39769 gradebook
   ```

```
...
There were 27 subject lines
```

Also, print the number of properly formatted "Subject:" lines your program finds in the input data.

Note: There are two different line formats. Some have the word "in" in the subject line and others do not. So your code needs to check for the "in" in the subject line and extract the data so that you print the correct data regardless of the format of the input line. For the following lines, the bolded information is what we want to see:

*Subject: [sakai] svn commit: **r39772** - **content**/branches/sakai_2-5-x/ ...*
*Subject: [sakai] svn commit: **r39771** - in **bspace**/site-manage/sakai_2-4-x ...*
*Subject: [sakai] svn commit: **r39770** - **site-manage**/branches/sakai_2-5-x/ ...*
*Subject: [sakai] svn commit: **r39769** - in **gradebook**/trunk/app/ui/src ...*

Example:

```
Enter a file name: mbox-short.txt
r39772 content
r39771 bspace
r39770 site-manage
r39769 gradebook
r39766 site-manage
r39765 gradebook
   ...
r39749 bspace
r39746 bspace
r39745 providers
r39744 oncourse
r39743 gradebook
r39742 gradebook
There were 27 subject lines


Enter a file name: mbox.txt
r39772 content
r39771 bspace
r39770 site-manage
r39769 gradebook
   ...
r37110 ctools
r37109 osp
r37108 content
r37107 ctools
There were 1780 subject lines
```