open.michigan

UNIVERSITY OF MICHIGAN SCHOOL OF INFORMATION

SI502

Assignment 4 – Being Careful With Your Input

Due Date: Friday February 6, 2009 at 5:00PM

In this assignment you will search through the file and look for lines that begin with "From". You will parse the From line and print out the second word for each From line and then you will also count the number of From lines and print out a count at the end.

This is a sample good output with a few lines removed:

```
python fromcount.py
Enter a file name: mbox-short.txt
stephen.marquard@uct.ac.za
louis@media.berkeley.edu
zqian@umich.edu
rjlowe@iupui.edu
  [ … snip … ]
louis@media.berkeley.edu
ray@media.berkeley.edu
cwen@iupui.edu
cwen@iupui.edu
cwen@iupui.edu
27
```

The following is a partially completed program that blows up - you need to adapt this program and fix the errors in the program so the output looks as shown above.

```
import string
fname = raw_input("Enter a file name: ")
# Make sure to deal with "file not found" gracefully
infile = open(fname, "r")
# Set our counter to zero
x = 0
for line in infile:
    # Split the line into words based on whitespace
    words = string.split(line)
    # debug print the number of words we have and the words
    print len(words), words
    # The following line fails when there are no words on a line
    if ( words[0] == 'From' ) :
        x = x + 1
        print x, words[1]

print "There were",x,"lines in the file with From as the first word"
```

When the program runs and is given a bad fie name

```
python frompart.py
Enter a file name: fred
Traceback (most recent call last):
  File "frompart.py", line 6, in <module>
    infile = open(fname, "r")
IOError: [Errno 2] No such file or directory: 'fred'
```

The fixed program should recover from the bad file name and print the following:

```
python fromcount.py
Enter a file name: fred
File not found: fred
```

When this program runs with mbox-short.txt, it produces the following output:

```
python frompart.py
Enter a file name: mbox-short.txt
7 ['From', 'stephen.marquard@uct.ac.za', 'Sat', 'Jan', '5', '09:14:16', '2008']
stephen.marquard@uct.ac.za ===============
2 ['Return-Path:', '<postmaster@collab.sakaiproject.org>']
5 ['Received:', 'from', 'murder', '(mail.umich.edu', '[141.211.14.90])']
6 ['by', 'frankenstein.mail.umich.edu', '(Cyrus', 'v2.3.8)', 'with', 'LMTPA;']
6 ['Sat,', '05', 'Jan', '2008', '09:14:16', '-0500']

  [ … snip … ]

'[194.35.219.184])']
5 ['BY', 'holes.mr.itd.umich.edu', 'ID', '477F90B0.2DB2F.12494', ';']
5 ['5', 'Jan', '2008', '09:14:10', '-0500']
5 ['Received:', 'from', 'paploo.uhi.ac.uk', '(localhost', '[127.0.0.1])']
7 ['by', 'paploo.uhi.ac.uk', '(Postfix)', 'with', 'ESMTP', 'id', '5F919BC2F2;']
7 ['Sat,', '5', 'Jan', '2008', '14:10:05', '+0000', '(GMT)']
7 ['Date:', 'Sat,', '5', 'Jan', '2008', '09:12:18', '-0500']
9 ['X-Authentication-Warning:', 'nakamura.uits.iupui.edu:', 'apache', 'set', 'sender',
'to', 'stephen.marquard@uct.ac.za', 'using', '-f']
2 ['To:', 'source@collab.sakaiproject.org']
2 ['From:', 'stephen.marquard@uct.ac.za']
7 ['Subject:', '[sakai]', 'svn', 'commit:', 'r39772', '-', 'content/branches/sakai_2-
5-x/content-impl/impl/src/java/org/sakaiproject/content/impl']
3 ['X-Content-Type-Outer-Envelope:', 'text/plain;', 'charset=UTF-8']
2 ['X-DSPAM-Result:', 'Innocent']
6 ['X-DSPAM-Processed:', 'Sat', 'Jan', '5', '09:14:16', '2008']
2 ['X-DSPAM-Confidence:', '0.8475']
2 ['X-DSPAM-Probability:', '0.0000']
0 []
Traceback (most recent call last):
  File "frompart.py", in <module>
    if ( words[0] == 'From' ) :
IndexError: list index out of range
```

Turn in your completed Python program to CTools as an attachment. You are welcome and encouraged to include debug print statements in your program that are commented out. Your program should produce the output shown at the beginning of this assignment.

Your Tasks

You are to fix the program to recover from the bad file name.

You are to fix the words[0] error at least **two** ways using two different techniques chosen from the following list:

- Use two if statements nested within one another
- Use the Boolean operator (and)
- Use an if statement and a continue

You are to hand in two separate versions of the program. Add a comment to each version describing how you fixed the error. If you want you can hand in more than two versions.