Regex exercises SP 18

1. Write steps (pseudo code) to get a word count from a text file.
2. Fill in the pattern below for the findall function to ensure just valid emails are gathered.

str = 'purple alice@google.com, blah bob@abc.com blah dishwasher'  
  tuples = re.findall(r'\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_', str)  
  print (tuples)  ## [('alice', 'google.com'), ('bob', 'abc.com')]  
  for tuple in tuples:  
    print (tuple[0])  ## username  
    print (tuple[1])  ## host

1. What do the tuples print out from question #2 above given a pattern that would recognize only valid emails?
2. Greedy or non greedy? That is the question. Describe the outcome below as a greedy

Suppose you have text with tags in it: <b>foo</b> and <i>so on</i>

Further suppose you are trying to match each tag with the pattern '(<.\*>)' -- what does it match first?

1. Given the string str below, what will the pattern **a{4,}** match from the string?

str = 'aaaabbbabbaaaaaba'  
tuples = re.findall(r'a{4,}', str)  
print(tuples)