```
count.py X

11aug2000 > count.py > UniqueWordCount

1  # Import necessary libraries
2  from mrjob.job import MRJob
3  import re

4

5  # Regular expression for finding words in a string
6  WORD_PATTERN = re.compile(r"\w+")

7

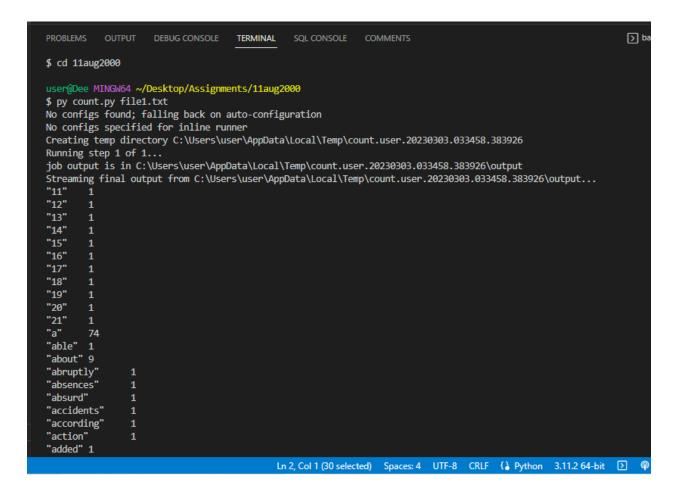
8  # Define a class for counting words
9  class UniqueWordCount(MRJob):
10

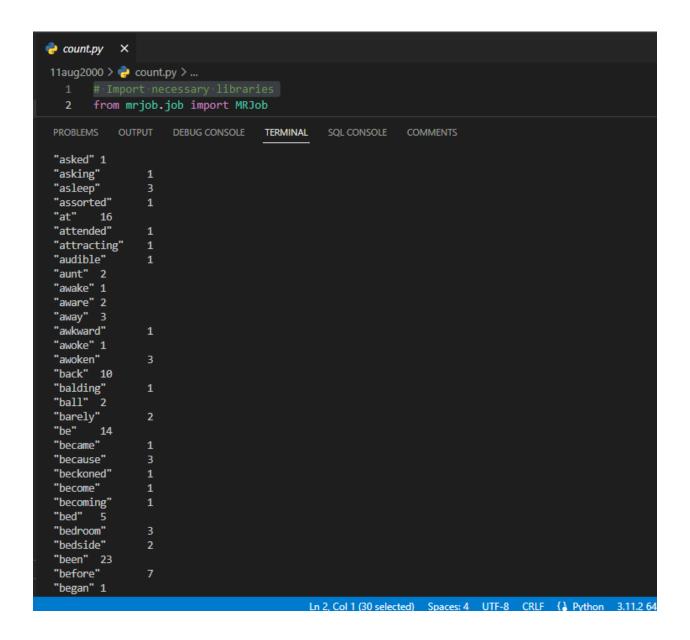
11  # define the mapper function that reads each line.
12  def mapper(self, _, line):
13  for word in WORD_PATTERN.findall(line):
14  # Yield each word
15  yield word.lower(), 1

16

17  # define reducer function that receives a word and a list of counts for that word.
18  def reducer(self, word, values):
19  # Sum the counts for each word and emit the result
19  total_count = sum(values)
20  yield word, total_count
21

22  # Run class
23  # Run class
24  if __name__ == '__main__':
25  UniqueWordCount.run()
```





```
🥏 count.py 🛛 🗙
11aug2000 > 🎨 count.py > ધ UniqueWordCount
   2 from mrjob.job import MRJob
 PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL
                                                                                                                       ) bash
 "or" 4
 "orange"
 "other" 2
"our" 3
"out" 9
 "outside"
 "over" 8
"owl" 1
"p" 11
"pain" 3
"painful"
 "painfully"
 "painless"
 "panic" 1
 "paralyzed"
 "parchment"
 "parents"
 "part" 2
 "passageway"
 "passing"
 "path" 1
  "patterned"
 "peered"
 "people"
 "perched"
 "perfectly"
 "perform"
 "performed"
 "pet" 1
 "peter" 1
 "petunia"
```

```
🥏 count.py 🛛 🗙
11aug2000 > 🧓 count.py > 😭 UniqueWordCount
       from mrjob.job import MRJob
 PROBLEMS OUTPUT
                     DEBUG CONSOLE
                                                SQL CONSOLE
                                                             COMMENTS
                                     TERMINAL
 "wizard"
 "wizarding"
 "wizardry"
 "wizards"
 "woke" 1
 "woman" 1
 "wooden"
                1
 "word" 1
 "words" 1
 "worked"
 "world" 3
 "wormtail"
               24
 "worries"
                1
 "would" 15
 "wrath" 1
 "wrong" 1
 "www" 18
"year" 3
 "years" 2
"yes" 1
"yet" 4
 "you" 39
 "your" 4
 "zooming"
                 1
 "ztcprep"
                 18
 Removing temp directory C:\Users\user\AppData\Local\Temp\count.user.20230303.033458.383926...
```

```
inaug2000 > i
```

```
🥏 non_eng.py 🗙
11aug2000 > 👶 non_eng.py > ...
  1 from mrjob.job import MRJob
       import enchant
     d = enchant.Dict("en_US")
     class NonEnglishWordsCount(MRJob):
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL SQL CONSOLE COMMENTS
user@Dee MINGW64 ~/Desktop/Assignments
$ cd 11aug2000
user@Dee MINGW64 ~/Desktop/Assignments/11aug2000
$ py non_eng.py file2.txt
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory C:\Users\user\AppData\Local\Temp\non_eng.user.20230303.033814.937496
Running step 1 of 1...
job output is in C:\Users\user\AppData\Local\Temp\non_eng.user.20230303.033814.937496\output
Streaming final output from C:\Users\user\AppData\Local\Temp\non_eng.user.20230303.033814.937496\output...
"\u2013"
"\u2014"
                 10
"\u2014\u201d" 2
"\u2018fortuna" 1
"\u2018sibyll" 1
"\u201c"
"\u201ca"
"\u201caha!\u201d"
 "\u201call"
 "\u201cand"
"\u201canyway," 1
"\u201care"
```

```
🥏 non_eng.py 🗙
11aug2000 > 👘 non_eng.py > ...
   1 from mrjob.job import MRJob
   2 import enchant
   5 d = enchant.Dict("en_US")
   PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL
                                                            SQL CONSOLE COMMENTS
"another," 2
"applause," 1
"appointment,\u201d" 1
"arithmancy," 1
"arithmancy," 1
"around," 1
"azkaban" 12
"azkaban," 1
"back," 2
"be,\u201d" 1
"beard," 1
"bed," 1
 "beds," 1
 "been," 1
 "begin!\u201d" 1
 "blandly," 1
"book?\u201d" 1
"braggart!" 1
 "business.\u201d" 1
"cadogan!\u201d" 1
 "castle," 1
"ceiling," 1
 "charge!\u201d" 1
 "cheerful," 1
"clapping," 1
"cloaks,\u201d" 1
```

```
🥏 non_eng.py 🗙
11aug2000 > 👶 non_eng.py > ...
       from mrjob.job import MRJob
       import enchant
   4 # create a dictionary for US English words
       d = enchant.Dict("en_US")
       class NonEnglishWordsCount(MRJob):
 PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL
                                                SQL CONSOLE
                                                              COMMENTS
 "harry,\u201d" 1
 "harry?\u201d" 1
 "hastily,"
 "hat," 1
"he," 1
"heart,"
"hermione" 13
"hermione," 4
"hermione." 2
"him."
 "him," 1
 "him, \u201d" 1
 "him;" 1
 "himself,\u201d"
 "hogwarts!" 1
 "hones\u2019ly."
 "however,"
 however," 2
"hufflepuff," 1
 "importance,\u201d"
                         1
 "insides,"
 "it," 1
"j.k." 11
 "job," 1
 "kettleburn" 1
 "kettleburn," 1
 "knaves,"
```