

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
20         total_count = sum(values)
21         yield word, total_count
22
23     # Run class
24 if __name__ == '__main__':
25     UniqueWordCount.run()
26
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL SQL CONSOLE COMMENTS

\$ cd 11aug2000

user@Dee MINGW64 ~/Desktop/Assignments/11aug2000

\$ py count.py file1.txt

No configs found; falling back on auto-configuration

No configs specified for inline runner

Creating temp directory C:\Users\user\AppData\Local\Temp\count.user.20230303.033458.383926

Running step 1 of 1...

job output is in C:\Users\user\AppData\Local\Temp\count.user.20230303.033458.383926\output

Streaming final output from C:\Users\user\AppData\Local\Temp\count.user.20230303.033458.383926\output...

```
"11" 1
"12" 1
"13" 1
"14" 1
"15" 1
"16" 1
"17" 1
"18" 1
"19" 1
"20" 1
"21" 1
"a" 74
"able" 1
"about" 9
"abruptly" 1
"absences" 1
"absurd" 1
"accidents" 1
"according" 1
"action" 1
"added" 1
```

Ln 2, Col 1 (30 selected) Spaces: 4 UTF-8 CRLF Python 3.11.2 64-bit

count.py X

11aug2000 > count.py > ...
1 # Import necessary libraries
2 from mrjob.job import MRJob

PROBLEMS

OUTPUT

DEBUG CONSOLE

TERMINAL

SQL CONSOLE

COMMENTS

"asked" 1

"asking" 1

"asleep" 3

"assorted" 1

"at" 16

"attended" 1

"attracting" 1

"audible" 1

"aunt" 2

"awake" 1

"aware" 2

"away" 3

"awkward" 1

"awoke" 1

"awoken" 3

"back" 10

"balding" 1

"ball" 2

"barely" 2

"be" 14

"became" 1

"because" 3

"beckoned" 1

"become" 1

"becoming" 1

"bed" 5

"bedroom" 3

"bedside" 2

"been" 23

"before" 7

"began" 1

Ln 2, Col 1 (30 selected) Spaces: 4 UTF-8 CRLF Python 3.11.2 64

count.py X

11aug2000 > count.py > UniqueWordCount

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

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL SQL CONSOLE COMMENTS

bash

```
"or" 4
"orange" 2
"other" 2
"our" 3
"out" 9
"outside" 3
"over" 8
"owl" 1
"p" 11
"pain" 3
"painful" 1
"painfully" 1
"painless" 1
"panic" 1
"paralyzed" 1
"parchment" 1
"parents" 1
"part" 2
"passageway" 1
"passing" 1
"path" 1
"patterned" 1
"peered" 1
"people" 1
"perched" 1
"perfectly" 1
"perform" 2
"performed" 1
"pet" 1
"peter" 1
"petunia" 2
```

count.py X

11aug2000 > count.py > UniqueWordCount

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

PROBLEMS	OUTPUT	DEBUG CONSOLE	TERMINAL	SQL CONSOLE	COMMENTS
----------	--------	---------------	----------	-------------	----------

"wizard"	7				
"wizarding"	2				
"wizardry"	1				
"wizards"	4				
"woke"	1				
"woman"	1				
"wooden"	1				
"word"	1				
"words"	1				
"worked"	1				
"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...

non_eng.py X

11aug2000 > non_eng.py > ...

```
1  from mrjob.job import MRJob
2  import enchant
3
4  # create a dictionary for US English words
5  d = enchant.Dict("en_US")
6
7  class NonEnglishWordsCount(MRJob):
8
9      # the mapper function reads each line and splits it into word
10     def mapper(self, _, line):
11         for word in line.split():
12             if not d.check(word):
13                 yield word.lower(), 1
14
15     # the reducer function receives a word and a list of counts for that word
16     def reducer(self, word, counts):
17         yield word, sum(counts)
18
19 if __name__ == '__main__':
20     NonEnglishWordsCount.run()
21
```

non_eng.py X

11aug2000 > non_eng.py > ...

```
1 from mrjob.job import MRJob
2 import enchant
3
4 # create a dictionary for US English words
5 d = enchant.Dict("en_US")
6
7 class NonEnglishWordsCount(MRJob):
8
```

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...

"(gryffindor," 1

"\u2013" 11

"\u2014" 10

"\u2014\u201d" 2

"\u2018fortuna" 1

"\u2018sibyll" 1

"\u201c" 1

"\u201ca" 1

"\u201caha!\u201d" 1

"\u201call" 2

"\u201cand" 1

"\u201canyway," 1

"\u201care" 1

Ln 1, Col 1 Spaces: 4 UTF-8 CR LF Python 3.11.2 64-bit

non_eng.py X

11aug2000 > non_eng.py > ...

```
1 from mrjob.job import MRJob
2 import enchant
3
4 # create a dictionary for US English words
5 d = enchant.Dict("en_US")
6
7 class NonEnglishWordsCount(MRJob):
8
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL SQL CONSOLE COMMENTS

```
"another,"      2
"applause,"     1
"appointment,\u201d" 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
```

Ln 1, Col 1 - Spaces: 4 - UTF-8 - CRLF - (/) Python - 3.11.2 64-bit

non_eng.py X

11aug2000 > non_eng.py > ...

```
1 from mrjob.job import MRJob
2 import enchant
3
4 # create a dictionary for US English words
5 d = enchant.Dict("en_US")
6
7 class NonEnglishWordsCount(MRJob):
8
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL SQL CONSOLE COMMENTS

```
"harry,\u201d" 1
"harry?\u201d" 1
"hastily," 1
"hat," 1
"he," 1
"heart," 1
"hermione" 13
"hermione," 4
"hermione." 2
"him," 1
"him,\u201d" 1
"him;" 1
"himself,\u201d" 1
"hogwarts!" 1
"hones\u2019ly." 1
"however," 2
"hufflepuff," 1
"importance,\u201d" 1
"insides," 1
"it," 1
"j.k." 11
"job," 1
"kettleburn" 1
"kettleburn," 1
"knaves," 1
```