

```
1 from mrjob.job import MRJob
2 import re
3
4 # Regular expression to split words by non-alphabetic
  characters
5 WORD_RE = re.compile(r"\b\w+\b")
6
7 class WordCount(MRJob):
8
9     def mapper(self, _, line):
10         # Split the line into words using the regular
      expression
11         words = re.findall(WORD_RE, line.lower())
12         # Emit each word as a key with a value of 1
13         for word in words:
14             yield (word, 1)
15
16     def reducer(self, word, counts):
17         # Sum the counts for each word
18         yield (word, sum(counts))
19
20 if __name__ == '__main__':
21     WordCount.run()
```