```
 1
 2
 3 from mrjob.job import MRJob
 4 import re
 5
 6 # List of stopwords
 7 STOPWORDS = set(["the", "and", "of", "a", "to", "in"
   , "is", "it"])
 8
 9 # Regular expression to split words by non-alphabetic
    characters
10 WORD_RE = re.compile(r"\b\w+\b")
11
12 class NonStopWordCount(MRJob):
13
14     def mapper(self, _, line):
15         # Split the line into words using the regular
    expression
16         words = re.findall(WORD_RE, line.lower())
17         # Emit each non-stop word as a key with a
   value of 1
18         for word in words:
19             if word not in STOPWORDS:
20                 yield (word, 1)
21
22     def reducer(self, word, counts):
23         # Sum the counts for each word
24         yield (word, sum(counts))
25
26 if __name__ == '__main__':
27     NonStopWordCount.run()
```