

```

1
2
3 from mrjob.job import MRJob
4
5 class BigramCount(MRJob):
6
7     def configure_args(self):
8         super(BigramCount, self).configure_args()
9         self.add_passthru_arg('--stopwords', type=str
, help='Comma-separated list of stopwords to exclude'
)
10
11     def mapper(self, _, line):
12         # Tokenize the line into words
13         words = line.split()
14
15         # Get the list of stopwords from the command
line argument
16         stopwords = set(self.options.stopwords.split(
', '))
17
18         # Generate bigrams and emit them as key-value
pairs
19         for i in range(len(words) - 1):
20             word1 = words[i].lower()
21             word2 = words[i + 1].lower()
22
23             # Exclude bigrams with stopwords
24             if word1 not in stopwords and word2 not
in stopwords:
25                 yield f"{word1},{word2}", 1
26
27     def reducer(self, bigram, counts):
28         # Sum the counts for each bigram
29         yield bigram, sum(counts)
30
31 if __name__ == '__main__':
32     BigramCount.run()
33

```