

pySpark code

```
import pyspark
from pyspark.sql import SparkSession
import re

spark = SparkSession.builder.appName("BibleCounter").getOrCreate()
sc = spark.sparkContext
sc.setLogLevel('WARN')

# Download the bible as txt
!curl -L "https://www.gutenberg.org/cache/epub/10/pg10.txt" > bible.txt

bible_lines = sc.textFile("bible.txt")
bible_words = bible_lines.flatMap(lambda line: [i for i in re.split('[^a-zA-Z]',
line.lower()) if i])
print(bible_words.count(), "words")
bible_word_counts = bible_words.map(lambda word: (word, 1))
bible_word_counts = bible_word_counts.reduceByKey(lambda count1, count2: count1 +
count2)
sorted_bible_counts = bible_word_counts.sortBy(lambda t: t[1], False)
sorted_bible_counts.take(10)
```

Output:

```
792193 words

[('the', 64116),
 ('and', 51696),
 ('of', 34722),
 ('to', 13598),
 ('that', 12912),
 ('in', 12667),
 ('he', 10419),
 ('shall', 9838),
 ('unto', 8997),
 ('for', 8971)]
```