```
//Section1: Question 1: uploading text file from Gutenberg Project
val text=sc.textFile("/FileStore/tables/1260_0-2.txt")
```

```
text: org.apache.spark.rdd.RDD[String] = /FileStore/tables/1260_0-2.txt MapPartitionsRDD[276]
at textFile at command-1912341606623214:2
```

```
//Question 2: Extract name entities in the text file

//Install John Snows Labs NLP
//Check the Spark and NLP's version
spark.version
```

```
res81: String = 3.1.0
```

```
import com.johnsnowlabs.nlp.SparkNLP

SparkNLP.version
```

```
import com.johnsnowlabs.nlp.SparkNLP
res82: String = 3.0.1
```

```
//Import NLP packages
import com.johnsnowlabs.nlp.base._

import com.johnsnowlabs.nlp.annotator._
```

```
import com.johnsnowlabs.nlp.base._
import com.johnsnowlabs.nlp.annotator._
```

```
//Filter the vocabularies with less than 1 character
val words=text.flatMap(line=>line.split("""\W+"""))
val longWords=words.filter(x=>x.length>=1)
val longWordsframe=longWords.toDF("text")
```

```
words: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[277] at flatMap at command-36966928
22597035:2
longWords: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[278] at filter at command-36966
92822597035:3
longWordsframe: org.apache.spark.sql.DataFrame = [text: string]
```