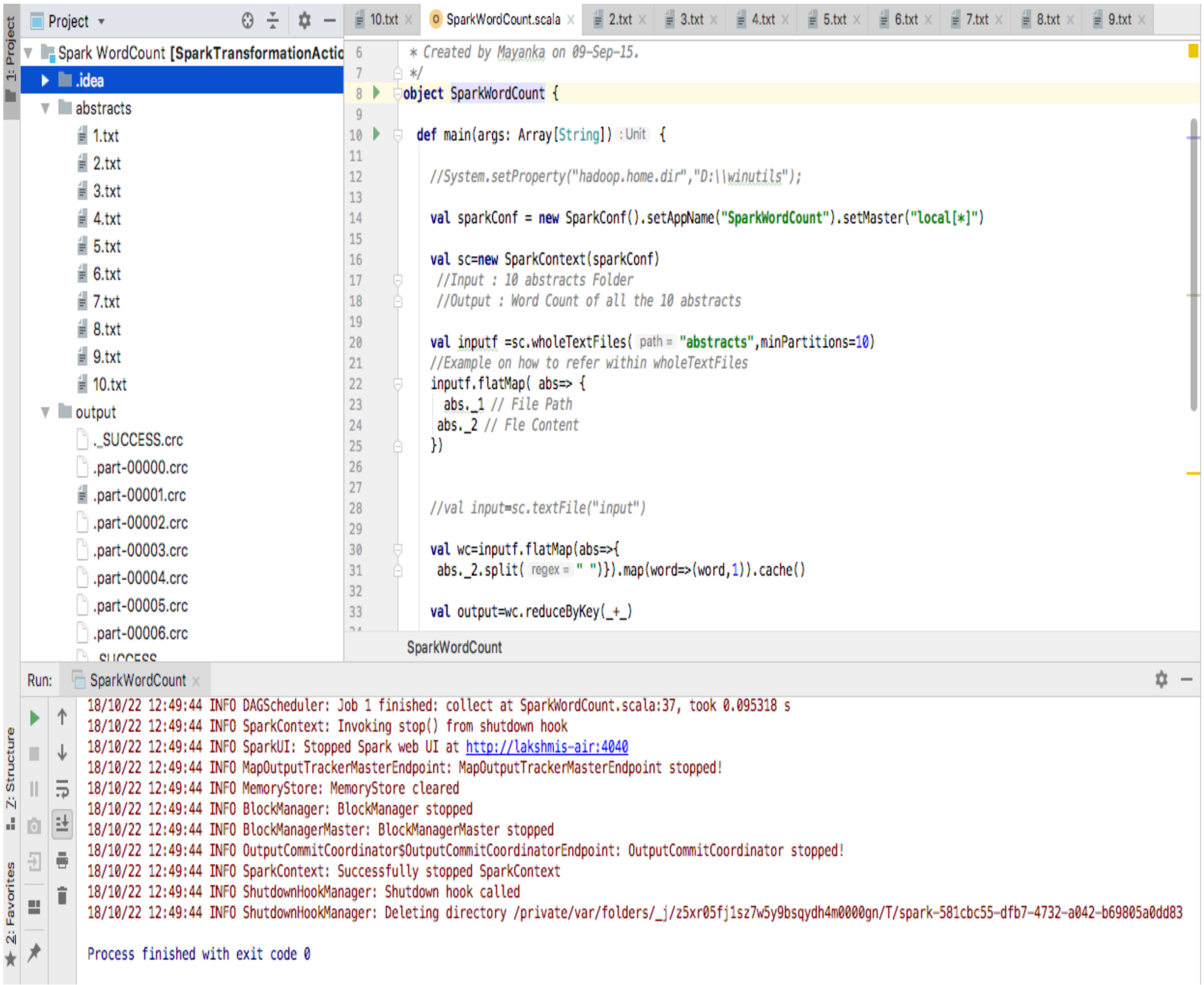
1. Create a Spark project, to perform the data statistics about 10 abstracts.

   Use at least two transformations and actions.

   Input: Choose 10 of abstracts from your dataset.

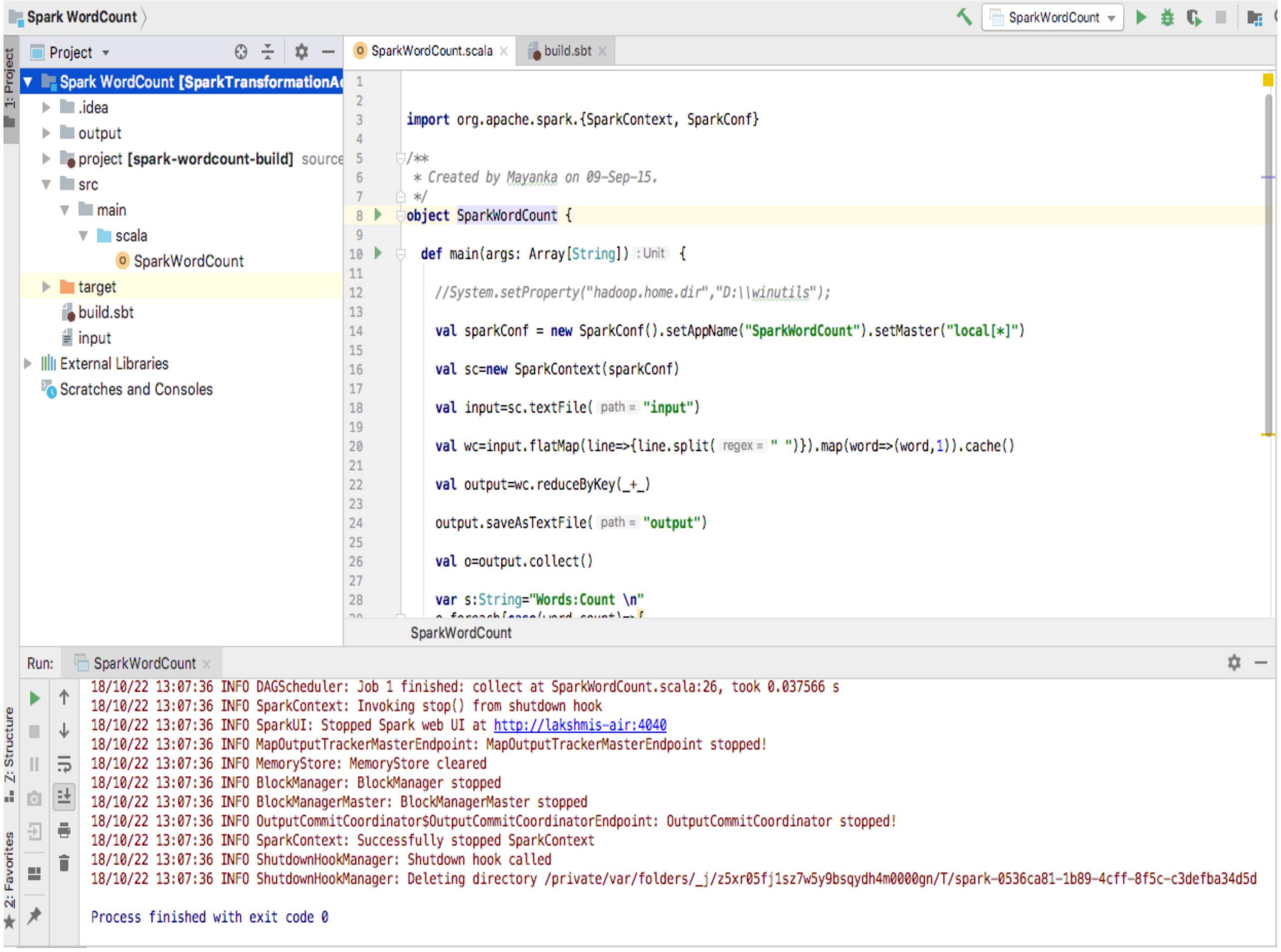
   Data Statistics Tasks:

   a. Word Count Statistics
   b. Parts of Speech Statistics
   c. WordNet Statistics
   d. Medical Word Statistics



```
 * Created by Mayanka on 09-Sep-15.
 */
object SparkWordCount {

  def main(args: Array[String]) : Unit  {

    //System.setProperty("hadoop.home.dir","D:\\winutils");

    val sparkConf = new SparkConf().setAppName("SparkWordCount").setMaster("local[*]")

    val sc=new SparkContext(sparkConf)
     //Input : 10 abstracts Folder
     //Output : Word Count of all the 10 abstracts

    val inputf =sc.wholeTextFiles( path = "abstracts",minPartitions=10)
    //Example on how to refer within wholeTextFiles
    inputf.flatMap( abs=> {
      abs._1 // File Path
     abs._2 // Fle Content
    })


    //val input=sc.textFile("input")

    val wc=inputf.flatMap(abs=>{
     abs._2.split( regex = " ")}).map(word=>(word,1)).cache()

    val output=wc.reduceByKey(_+_)
```

```
18/10/22 12:49:44 INFO DAGScheduler: Job 1 finished: collect at SparkWordCount.scala:37, took 0.095318 s
18/10/22 12:49:44 INFO SparkContext: Invoking stop() from shutdown hook
18/10/22 12:49:44 INFO SparkUI: Stopped Spark web UI at http://lakshmis-air:4040
18/10/22 12:49:44 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
18/10/22 12:49:44 INFO MemoryStore: MemoryStore cleared
18/10/22 12:49:44 INFO BlockManager: BlockManager stopped
18/10/22 12:49:44 INFO BlockManagerMaster: BlockManagerMaster stopped
18/10/22 12:49:44 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
18/10/22 12:49:44 INFO SparkContext: Successfully stopped SparkContext
18/10/22 12:49:44 INFO ShutdownHookManager: Shutdown hook called
18/10/22 12:49:44 INFO ShutdownHookManager: Deleting directory /private/var/folders/_j/z5xr05fj1sz7w5y9bsqydh4m0000gn/T/spark-581cbc55-dfb7-4732-a042-b69805a0dd83

Process finished with exit code 0
```



```scala
import org.apache.spark.{SparkContext, SparkConf}

/**
 * Created by Mayanka on 09-Sep-15.
 */
object SparkWordCount {

  def main(args: Array[String]) :Unit  {

    //System.setProperty("hadoop.home.dir","D:\\winutils");

    val sparkConf = new SparkConf().setAppName("SparkWordCount").setMaster("local[*]")

    val sc=new SparkContext(sparkConf)

    val input=sc.textFile( path = "input")

    val wc=input.flatMap(line=>{line.split( regex = " ")}).map(word=>(word,1)).cache()

    val output=wc.reduceByKey(_+_)

    output.saveAsTextFile( path = "output")

    val o=output.collect()

    var s:String="Words:Count \n"
```

```
18/10/22 13:07:36 INFO DAGScheduler: Job 1 finished: collect at SparkWordCount.scala:26, took 0.037566 s
18/10/22 13:07:36 INFO SparkContext: Invoking stop() from shutdown hook
18/10/22 13:07:36 INFO SparkUI: Stopped Spark web UI at http://lakshmis-air:4040
18/10/22 13:07:36 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
18/10/22 13:07:36 INFO MemoryStore: MemoryStore cleared
18/10/22 13:07:36 INFO BlockManager: BlockManager stopped
18/10/22 13:07:36 INFO BlockManagerMaster: BlockManagerMaster stopped
18/10/22 13:07:36 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
18/10/22 13:07:36 INFO SparkContext: Successfully stopped SparkContext
18/10/22 13:07:36 INFO ShutdownHookManager: Shutdown hook called
18/10/22 13:07:36 INFO ShutdownHookManager: Deleting directory /private/var/folders/_j/z5xr05fj1sz7w5y9bsqydh4m0000gn/T/spark-0536ca81-1b89-4cff-8f5c-c3defba34d5d

Process finished with exit code 0
```