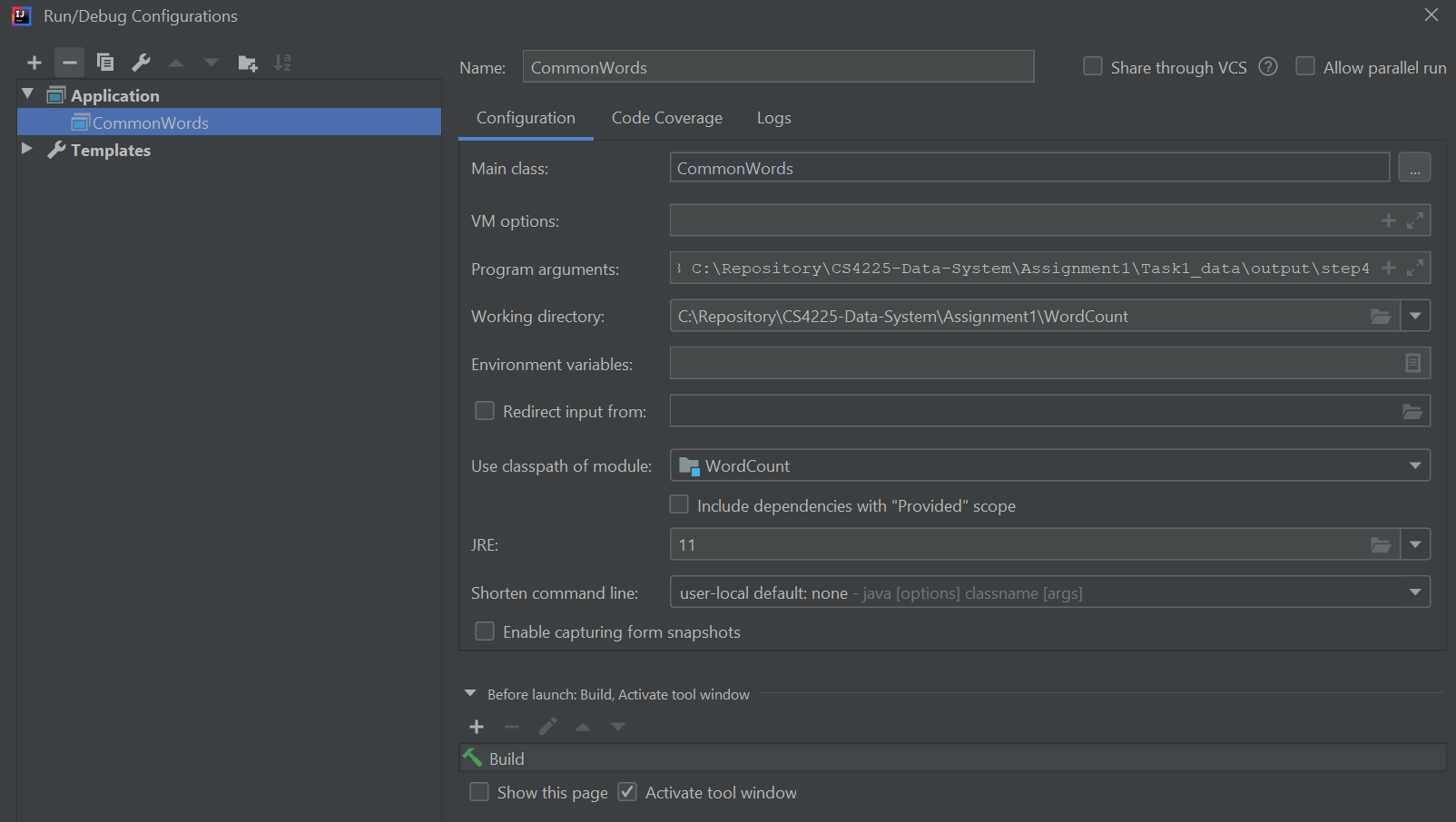
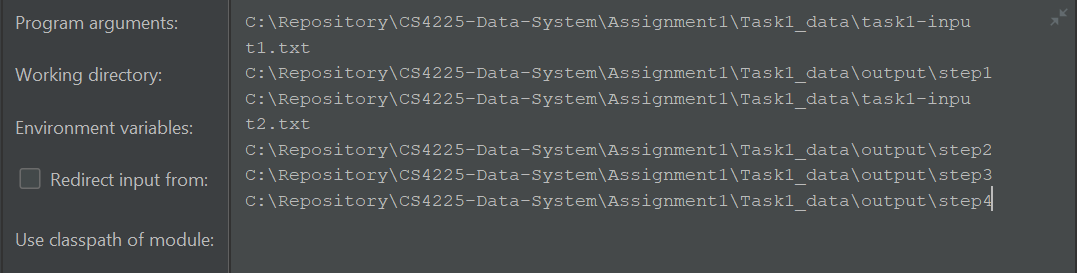
**Assignment 1 Task 1**

**How to run Task 1 code in IntelliJ**

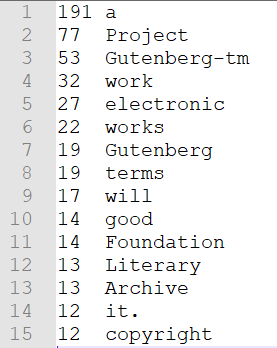
* Set up Maven project following “setup\_IDE\_win\_macos\_linux.docx”
* Edit running configuration



* Add file paths to program arguments



* Expected output



**How to run Task 1 code in HDFS**

* File structure

|  |  |
| --- | --- |
| Folder | Files |
| assignment1\_task1 | stopwords.txt task1-input1.txt task1-input2.txt |
|  | WordCount.jar |

docker cp WordCount.jar d46320c58d59:WordCount.jar

docker cp assignment1\_task1/stopwords.txt d46320c58d59: assignment1\_task1/stopwords.txt

docker cp assignment1\_task1/task1-input1.txt d46320c58d59:assignment1\_task1/task1-input1.txt

docker cp assignment1\_task1/task1-input2.txt d46320c58d59:assignment1\_task1/task1-input2.txt

* Set up directory and files in HDFS

docker exec -it master bash

hadoop fs -mkdir -p assignment1\_task1

hdfs dfs -put ./assignment1\_task1/\* assignment1\_task1

* Run jar file in Hadoop

hadoop jar WordCount.jar CommonWords assignment1\_task1/task1-input1.txt assignment1\_task1/output/step1 assignment1\_task1/task1-input2.txt assignment1\_task1/output/step2 assignment1\_task1/output/step3 assignment1\_task1/output/step4

* Check output

Output directory is assignment1\_task1/output/step4

