

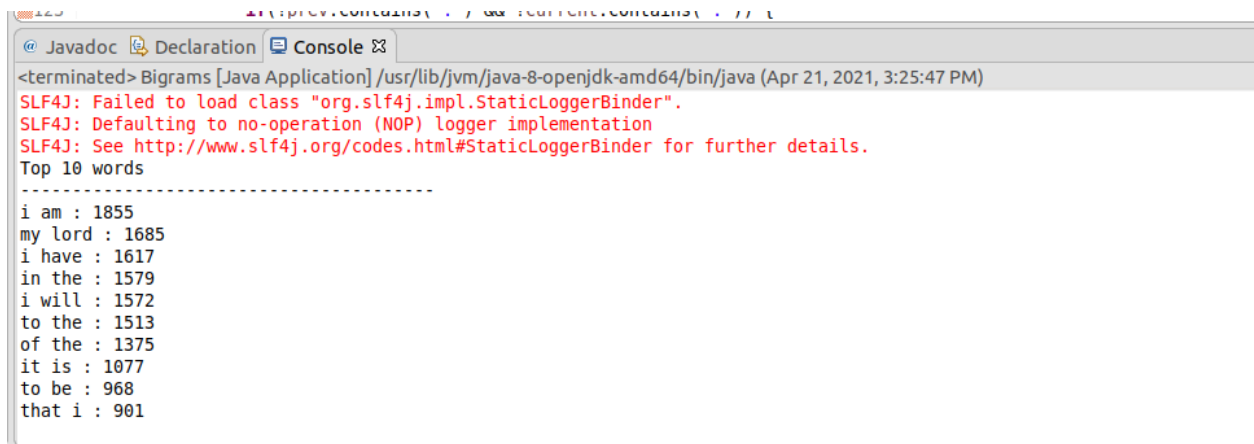
Lab 7

- 1) The below image is a screenshot obtained from the Exp1 on the Shakespeare dataset. The words were first grouped together using a custom FlatMapFunction and then the counting and sorting were implemented in a sink function.

```
<terminated> WordCount (1) [Java Application] /usr/lib/jvm/java-8-openjdk-amd64/bin/java (Apr 21, 2021, 1:58:02 PM)
SLF4J: Failed to load class "org.slf4j.impl.StaticLoggerBinder".
SLF4J: Defaulting to no-operation (NOP) logger implementation
SLF4J: See http://www.slf4j.org/codes.html#StaticLoggerBinder for further details.
Top 10 words
-----
the : 26856
and : 24116
i : 22412
to : 19225
of : 16018
you : 14097
a : 13986
my : 12283
that : 11171
in : 10640
|
```

Results of exp1 bonus are included as a text file called wordcountwindow.txt

- 2) For this experiment, for bigrams, I replaced words like "i'll" with "ill" so "i" "ll" will not be included as a bigram. All the question marks and exclamation marks were replaced with "." for simplicity. It was not clear to me whether the question required to list all the



```
<terminated> Bigrams [Java Application] /usr/lib/jvm/java-8-openjdk-amd64/bin/java (Apr 21, 2021, 3:25:47 PM)
SLF4J: Failed to load class "org.slf4j.impl.StaticLoggerBinder".
SLF4J: Defaulting to no-operation (NOP) logger implementation
SLF4J: See http://www.slf4j.org/codes.html#StaticLoggerBinder for further details.
Top 10 words
-----
i am : 1855
my lord : 1685
i have : 1617
in the : 1579
i will : 1572
to the : 1513
of the : 1375
it is : 1077
to be : 968
that i : 901
```

bigrams or just the top 10. So I have included code for printing the top 10. And also a text file of results if all bigrams are to be printed. Screenshot of exp 2 is pasted below:

The results of exp2 bonus are included under bigramwindow.txt.