

Exp1)

```
cp419@cp419-VirtualBox:~/Desktop$ hadoop fs -cat /lab2/output/wordcount/* | head -10
2022-02-14 13:25:50,838 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
!      10526
!'By    1
!'twas  1
!,      1
!As     1
!Ay     1
!Come   1
!Give't 1
!Handkerchief  1
!Hear    1
cat: Unable to write to output stream.
2022-02-14 13:25:51,123 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
```

Exp2)

Q: Just get rid of the useless symbols such as , . ! ?, and store the previous word in a variable before go to next token.

Shakespeare:

```
cp419@cp419-VirtualBox:~/hadoop/sbin$ hadoop fs -cat /user/hamamoto/lab2/output/exp2/*
2022-02-16 09:01:25,567 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
to the 1512
it is 1083
i am 1855
i ll 1783
to be 971
in the 1580
my lord 1699
i will 1571
i have 1620
of the 1374
```

Gutenberg:

```
cp419@cp419-VirtualBox:~/Downloads$ hadoop fs -cat /user/hamamoto/lab2/output/exp2/*
2022-02-16 09:40:56,075 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
to the 636851
in a 272811
on the 410838
and the 453613
of a 303399
it was 370330
at the 304649
to be 388179
in the 927220
of the 1575063
```

