

To launch the mapreduce:

Create the jar file, and with the following name class:

```
hadoop jar TFIDF.jar cs.bigdata.TFIDF.TFIDF [input1] [input2] [output]
```

Words with the highest tf-idf scores in these documents are: (Since punctuations are not removed, I put a bit more than 20 words)

Tf-idf	(word, docid)
0.005496667175439178	("Buck", callwild.txt)
0.001330542451197579	("dogs", callwild.txt)
0.0012869181085353632	("Thornton", callwild.txt)
0.0012545959976238653	("myself", defoe-robinson-103.txt)
9.597355385687453E-4	("Buck's", callwild.txt)
9.379233672376377E-4	("Spitz", callwild.txt)
9.161111959065297E-4	("Francois", callwild.txt)
8.288625105820983E-4	("John", callwild.txt)
8.288625105820983E-4	("sled", callwild.txt)
7.19801653926559E-4	("Buck,", callwild.txt)
6.761773112643434E-4	("dogs,", callwild.txt)
6.501088351323666E-4	("Friday", defoe-robinson-103.txt)
5.816763261710648E-4	("shore,", defoe-robinson-103.txt)
5.47460071690414E-4	("-", defoe-robinson-103.txt)
5.453042832776963E-4	("Perrault", callwild.txt)
5.36054653530197E-4	("However,", defoe-robinson-103.txt)
5.234921119465884E-4	("Hal", callwild.txt)
5.075411081296547E-4	("God", defoe-robinson-103.txt)
4.961356899694377E-4	("me.", defoe-robinson-103.txt)
4.961356899694377E-4	("me;", defoe-robinson-103.txt)
4.7986776928437267E-4	("trail", callwild.txt)
4.5805559795326483E-4	("ice", callwild.txt)
4.448113082484614E-4	("thoughts", defoe-robinson-103.txt)
4.36243426622157E-4	("team", callwild.txt)
4.36243426622157E-4	("Thornton's", callwild.txt)
4.277031810081359E-4	("indeed", defoe-robinson-103.txt)
4.1443125529104914E-4	("around", callwild.txt)
3.926190839599413E-4	("Dave", -callwild.txt)
3.820815083672681E-4	("viz.", defoe-robinson-103.txt)
3.7637879928715965E-4	("I,", defoe-robinson-103.txt)
3.7080691262883345E-4	("camp,", callwild.txt)
3.7080691262883345E-4	("Buck.", callwild.txt)
3.7067609020705115E-4	("towards", defoe-robinson-103.txt)