**#求weibo\_cnt**

cut -f5,6,9 weibo.top10wan | sort -k3n | awk -F"\t" 'BEGIN{s=0;}{if($3!=s) print $0; s=$3;}END{if($3!=s) print $0;}' | cut -f1 | sort | uniq -c | awk -F" " '{print $2" "$1;}' > userid+weibocount

**#求changweibo\_cnt**

cut -f4,5 weibo\_changweibo.2018-05-17 | sort -k2n | awk -F"\t" 'BEGIN{s=0;}{if($2!=s) print $0; s=$2;}END{if($2!=s) print $0;}' | cut -f1 | sort | uniq -c | awk -F" " '{print $2" "$1;}'> userid+changweibocount

**#合并共有**

join userid+weibocount userid+changweibocount > result.txt

**#合并只发了短微博的用户**

join -v 1 userid+weibocount userid+changweibocount > tmp1.txt

awk '$2=$2" 0"' tmp1.txt >> result.txt

**#合并只发了长微博的用户**

join -v 2 userid+weibocount userid+changweibocount > tmp2.txt

awk '$1=$1" 0 "' tmp2.txt >> result.txt

**#按照userid的发文数量做第一次降序排序，长微博文章量做第二次降序排序**

sort result.txt -k2,2rn -k3,3rn > hw2\_result.txt

# Tips:

# 1. 本次作业出现较多问题在于

1. 未将只发了长微博或者只发了短微博的userid进行合并排序
2. 排序时未完全按照短微博发文量作为第一次降序排序，例如：  
   answer1.txt  
   3555110134 92

2009574761 78

6015764972 65

1325344423 61

5090053433 61

5315625446 61

5051077448 49

1742566624 47

1906537253 46

3253198047 42

1663072851 40

5103645868 40

5915215161 38

6529707947 38

2112339023 36 4

1741249710 36

6522710140 36

2626472787 34

2628297355 33

1819638285 32  
  
answer2.txt

6529707947 38 0

2112339023 36 4

1750349294 30 1

2275883675 22 10

5549358119 22 1

1644114654 21 37

1789250890 21 12

2000961721 21 12

3275115274 21 11

1402977920 21 8

1664176597 21 6

2028810631 21 6

5470635923 21 1

5666958393 21 0

6017112148 21 0

1845864154 20 64

2292896411 20 30

1653603955 20 29

1267454277 20 27

3057540037 20 27

这两种答案中，answer2.txt是正确的。

# 2. 本次作业的最优写法请参考 脚本作业3参考1 中的正确参考写法。