# 数据总量

wc -l top10wan # 100000

# 微博发文user\_id总量

cat top10wan | awk -F"\t" '{print $5}' | sort | uniq | wc -l # 47661

# 微博文章weibo\_id总量

cat top10wan | awk -F"\t" '{print $9}' | sort | uniq | wc -l # 96688

# 计算发文(weibo\_id)最多的top 100个用户

cat top10wan | awk -F"\t" '{print $5,$9}' | sort | uniq | awk '{ct[$1]+=1};

  END {

   for(uid in ct) {print uid, ct[uid]}

  }

' | sort -k2 -nr | head -100

# 6529707947 38

# 2112339023 36

# 1750349294 30

# 5549358119 22

# 2275883675 22

# 6017112148 21

# 5666958393 21

...