#预处理数据

cut -f9,10,23 weibo.top10wan | sort -u | cut -f2,3 > weibo\_content.txt

#抓取Email

cat weibo\_content.txt | grep -E "[A-Za-z0-9.\_-]+@[A-Za-z0-9\_-]+[\.][A-Za-z.]\*" > email\_weibo

#可以用grep -o的方法查看自己grep到的email address是否正确。

#抓取小米品牌

cat weibo\_content.txt | grep -E "(小米|红米)[ A-Za-z0-9]+" > xiaomi\_product\_weibo

cat xiaomi\_product\_weibo | grep -E -o "(小米|红米)[A-Za-z0-9 ]+" | sed 's/[ \t]\*$//' | sort | uniq > xiaomi\_product

echo "小米一共有"$(cat xiaomi\_product | wc -l)"个独立品牌"

#这是我目前为止能抓到最多品牌的写法，如有更多的写法欢迎指教。