# 实验 03 过滤器与正则表达式

班级：数据科学与大数据 2 班
学号：202026203039
姓名：赖丽婷
用户名：llt

一、实验目的
1．练习操作文件和目录的常用命令

二、实验要求
1．填写实验报告，请将关键命令及其结果进行截图(确保截图中的文字清晰可见)
2．导出为 pdf 文件，文件名为用户名-姓名-lab02.pdf，在规定截止时间之前上传作业）
3．以下步骤中**所有 s01 请换成你自己的用户名。**

三、实验步骤
1.对实验文件 sea.txt 中所有出现的单词进行计数，并按照出现次数从大到小的顺序打印出各单词及其出现次数。(要求统一大小写并且把标点符号和多余的空白符去除干净后进行统计）


cat 读取文件 把大写转为小写 删除连续的空格 把空格转为换行符 删除特殊字符 先排序为了能删除重复的行 然后在按字符出现的数字排序

2. 打印实验文件 rfile.txt 中包含合法数字（整数、小数、科学计数法，可正可负）的行
示例：
123 # 合法数字
12.3 # 合法数字
.123 # 合法数字
-3.14 # 合法数字
-3.14e9 # 合法数字
-.14E10 # 合法数字
-.14E-10 # 合法数字
1.2.3 # 非法数字
1.2.e3.5 # 非法数字
-1.-2E5 # 非法数字

```
llt@llt-virtual-machine:~/下载$ grep -E "^[-|+]?([0-9]*)?(.[0-9]+)?[E|e]?[-|+]?
([0-9]*)?$" rfile.txt

1
12
123
1234
12345
1234567
12345678
123456789
.1
0.1
0.12
12.3
12.34
123.45
123.456
1234.56
1234.567
-.1
-0.1
-12.3
-12.34
-123.45
-123.456
-1234.56
-1234.567
.1e10
-.1e10
-.1e-10
.1E8
-.1E8
-.1E-8
1.2e6
1.2E6
-1.2e6
-1.2E6
24.5e7
24.5E7
-24.5e7
-24.5E7
3e8
3E8
-3e8
-3E8



389092934
283987983893237
398478348957389459
49583749573593593845
209349891231838
897345151301381
084594991042948
094589200109193982

llt@llt-virtual-machine:~/下载$
```

3. 打印实验文件 rfile.txt 中包含合法单播 IP 地址的行

示例：

100.20.3.1 # 合法单播 IP 地址

293.38.20.2 # 非法单播 IP 地址

39.39.49.0 # 非法单播 IP 地址

39.49.493.39 # 非法单播 IP 地址

39.34.39.938 # 非法单播 IP 地址

34.34.34.255 # 非法单播 IP 地址

255.255.255.255 # 非法单播 IP 地址

254.0.0.254 # 合法单播 IP 地址

0.0.0.0 # 非法单播 IP 地址

0.1.2.3 # 非法单播 IP 地址

3.0.0.1 # 合法单播 IP 地址

```
llt@llt-virtual-machine:~/下载$ grep "^\(\(1[0-9][0-9]\.\)\|\(2[0-4][0-9]\.\)\|
\(25[0-4]\.\)\|\([1-9][0-9]\.\)\|\([1-9]\.\)\)\(\(1[0-9][0-9]\.\)\|\(2[0-4][0-9
]\.\)\|\(25[0-4]\.\)\|\([1-9][0-9]\.\)\|\([0-9]\.\)\)\{2\}\(\(1[0-9][0-9]\)\|\(
2[0-4][0-9]\)\|\(25[0-4]\)\|\([1-9][0-9]\)\|\([1-9]\)\)$" rfile.txt
100.20.3.1
254.0.0.254
130.34.34.130
3.0.0.1
llt@llt-virtual-machine:~/下载$
```

4. 打印实验文件 rfile.txt 中包含合法身份证号（15 位或 18 位）的行

身份证编码规则：

规则 1：15 位身份证号

- 6 位地址码：任意数字
- 6 位出生日期：2 位年+2 位月(01-12)+2 位日(01-31){{
- 3 位顺序码：任意数字

规则 2：18 位身份证号

- 6 位地址码：任意数字
- 8 位出生日期：
- 4 位年(18XX,19XX,2XXX)+2 位月(01-12)+2 位日(01-31)
- 3 位顺序码：任意数字
- 1 位校验码：任意数字或大写字母 X

示例：

283987983893237 # 非法

398478348957389459 # 非法

49583749573593593845 # 非法

349857394857935 78X # 非法

34895738953498934x # 非法

209349891231838 # 合法

897345151301381 # 非法

084594991042948 # 非法

094589200109193982 # 合法

956809200001395 68X # 非法

39485719991016487X # 合法