# 实验03过滤器与正则表达式

班级：数据科学与大数据技术1班

学号：202026203005

姓名：张华

用户名：s13

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）

文本

中度可信度描述已自动生成

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

文本

描述已自动生成

形状, 矩形

描述已自动生成

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

文本

描述已自动生成

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

34985739485793578X # 非法

34895738953498934x # 非法

209349891231838 # 合法

897345151301381 # 非法

084594991042948 # 非法

094589200109193982 # 合法

95680920000139568X # 非法

39485719991016487X # 合法

文本

描述已自动生成