

华侨大学工学院

实验报告

课程名称: Pyhton 编程 (实验)

实验项目名称: 字符串与正则表达式

学 院: 工学院

专业班级: 物联网工程 1 班

姓 名: 李昊唐

学 号: 1995131017

指导教师: 王佳斌

2021 年 6 月 2 日

预 习 报 告

一、 实验目的

通过 Python 字符串与正则表达式的学习，掌握 Python 字符串格式化方法，字符串常用方法，掌握正则表达式中 re 模块的主要方法，掌握正则表达式是在处理字符串中的应用，为进一步使用 Python 语言编程，以及编写复杂程序打下基础。

二、 实验仪器

三、 实验原理

1. python 中文件的读写 ab, rb 等
2. 写入时如何加入\n,因为版本不同，所以要加\r\n
3. 如何查找整个字符串并替换数字。
4. re 模块的 filter 函数，对序列做过滤处理；digit 判断序列是否全为数字；两者组合可筛出数组里的数字：filter(str.isdigit, char)
5. string 模块:<http://www.cnblogs.com/rollenholt/archive/2011/11/25/2263722.html>
6. a = a.lower()变小写；
7. 新建字典并赋值：for i in string.lowercase: #lowercase:小写字母
count: 统计频率 dict[i]=a.count(i)
8. set 函数：给出一个有若干字符串组成的列表，set()函数返回列表

预 习 报 告

中不同的字符串组成的集合。把它想象成一个 `for` 循环。从列表里拿出第一个元素，放到集合。第二，三，四，五个，等等，它已经在集合里面了，因为 `Python` 集合不允许重复，所以它只被列出了一次。第六个。第七个又是一个重复的，所以它只被列出了一次。原来的列表甚至不需要事先排好序。

给出一个字符串列表, `".join(a_list)` 将所有的字符串拼接成一个。

9. `b.sort()` 排序

四、实验内容及步骤

1. 教材基本字符串与正则表达式操作练习

- (1) 第 7，8 章的教材内容练习
- (2) 第 7，8 章的精彩案例练习

2. `python` 字符串与正则表达式拓展实验（一）

1) 实验内容

已知一个文件 `test.txt`，内容如下：

2012 来了。

2012 不是世界末日。

2012 欢乐多。

- 1. 请输出其内容。
- 2. 请计算该文本的原始长度。

实验报告

五、 实验原始数据

- 3.请去除该文本的换行。
- 4.请替换其中的字符"2012"为"2015"。
- 5.请提取内容为数字的子串（2012， 2013）。

实验要求

1. 命令行：python 2012.py filename
 - a) 2012.py 为编写的 Python 代码文件
 - b) Filename 为文件路径
2. 按序号完成上述 5 个任务
3. 任务 3，去掉换行符后再保存到新的文件
4. 任务 4，替换后再保存到新的文件
5. 任务 1、2、5，提取后直接输出到 stdout

```
import sys
import re

if __name__ == "__main__":
    with open(sys.argv[1], mode='r') as f:
        text = f.read()
    print(text)
    print(len(text))
    with open('test1.txt', mode='w') as f:
        f.write(text.replace('\n', ''))
    with open('test2.txt', mode='w') as f:
        f.write(text.replace('2012', '2015'))
    print(re.findall(r'(\w*[0-9]+\w*', text))
```

指导老师签名：_____

时 间：_____

实验报告

六、数据处理

python 字符串与正则表达式拓展实验（二）

一、实验要求

字符串 `a = "aAsmr3idd4bgs7Dlsf9eAF"`

1. 请将 `a` 字符串的数字取出，并输出成一个新的字符串。
2. 请统计 `a` 字符串出现的每个字母的出现次数（忽略大小写，`a` 与 `A` 是同一个字母），并输出成一个字典。 例 `{'a':3,'b':1}`
3. 请去除 `a` 字符串多次出现的字母，仅留最先出现的一个,大小写不敏感。例：'`aAsmr3idd4bgs7Dlsf9eAF`'，经过去除后，输出 '`asmr3id4bg7lf9e`'
4. 按 `a` 字符串中字符出现频率从高到低输出到列表，如果次数相同则按字母顺序排列。

```
import string
a = "aAsmr3idd4bgs7Dlsf9eAF"
num = list(filter(str.isdigit, a))
num = ''.join(num)
print(num)
a = a.lower()
cnt = {}
for i in string.ascii_lowercase:
    cnt[i] = a.count(i)
    if a.count(i) == 0:
        del cnt[i]
print(cnt)
temp = []
for i in a:
    if i not in temp:
        temp.append(i)
        print(i, end='')
print()
t = sorted(cnt.items(), key=lambda v: (v[1], v[0]), reverse=True)
print(t)
```

实 验 报 告

七、 实验结论及分析讨论

预习报告成绩	实验报告成绩	实验操作成绩	总成绩