

## Python 程序设计 实验 5：字符串的应用

### 注意事项：

- (1) 实验报告提交**截止日期：2021.04.15， 23:59pm**，迟交扣 20%，缺交 0 分。
- (2) 实验报告内容包括：解决问题的思路与方法（如代码的解释）、遇到的问题以及收获（简单描述即可）、代码运行结果的展示。
- (3) 实验报告提交方法：**blackboard**。
- (4) 提交要求：实验报告+源代码，打包上传，命名：学号\_姓名\_实验报告\_5。
- (5) **禁止抄袭**，一经发现 **0 分处理**（包括抄袭者和提供代码或实验报告者）！

**1. 字符串推导式：**我们之前学过列表推导式。例如，生成前 4 个奇数，我们可以写

```
[2 * num - 1 for num in range(1,5)] # 生成[1, 3, 5, 7]
```

仿照上面写法，使用推导式完成以下字符串操作：

- 1) ['apple', 'orange', 'pear'] -> ['A', 'O', 'P'] # 第一个字母大写
- 2) ['apple', 'orange', 'pear'] -> ['apple', 'pear'] # 是否包含字母'p'
- 3) ["TA\_parth", "student\_poohbear", "TA\_michael", "TA\_guido", "student\_htiek"] -> ["parth", "michael", "guido"] #是否以 TA\_开头
- 4) ['apple', 'orange', 'pear'] -> [('apple', 5), ('orange', 6), ('pear', 4)] #生成列表
- 5) ['apple', 'orange', 'pear'] -> {'apple': 5, 'orange': 6, 'pear': 4} #生成字典

**2. 特殊单词：**有一种特殊的英文单词，它的相邻字母对之间的“距离”不断增加。如单词 subway，它的相邻字母对分别为 (s, u), (u, b), (b, w), (w, a), (a, y)，字母之间的距离依次为 2, 19, 21, 22, 24（如 a 是第 1 个字母，y 是第 25 个字母，a 和 y 的距离为 24）。编写函数 is\_special\_word(word)，word 为输入单词字符串，如为特殊单词，返回 True；否则返回 False。

**3. Caesar 加密：**Caesar 加密将文本中每个字符向前移动三个字母，即

A->D, B->E, ..., X->A, Y->B, Z->C

如单词 PYTHON，将加密为 SBWKRQ。

编写加密函数 `encrypt_caesar(plaintext)`和 `decrypt_caesar(ciphertext)`，每个函数都接受一个参数，表示将要加密或解密的字符串，然后返回加密或解密之后的字符串。

注：（1）只考虑大写字母；（2）非字母字符不需要通过加密或解密来修改。如 `encrypt_caesar("FIRST POST")`应该返回 `"IIUVW S0VW"`；（3）`encrypt_caesar("")`应返回 `" "`，即空字符串。

**4. 杨辉三角：编写函数 `print_yanghui_triangle(N)`，输出杨辉三角的前 N 行，要求使用 `format` 函数进行格式化输出。**

例子：`print_yanghui_triangle(10)`，输出

```

      1
     1 1
    1 2 1
   1 3 3 1
  1 4 6 4 1
 1 5 10 10 5 1
1 6 15 20 15 6 1
1 7 21 35 35 21 7 1
1 8 28 56 70 56 28 8 1
1 9 36 84 126 126 84 36 9 1
```

## 5. 可读性等级

不同书的阅读群体不一样。例如，一本书中可能有许多很长的复杂单词；而较长的单词可能与较高的阅读水平有关。同样，较长的句子也可能与较高的阅读水平相关。研究者开发了许多可读性测试，给出了计算文本阅读水平的公式化过程。其中一个可读性测试是 Coleman Liau 指标：文本的 Coleman Liau 指标旨在划分理解文本所需的阅读水平等级。Coleman Liau 指标公式如下

$$\text{index} = 0.0588 * L - 0.296 * S - 15.8$$

其中，L 是文本中每 100 个单词的平均字母数，S 是文本中每 100 个单词的平均句子数。

考虑以下文本：

Congratulations! Today is your day. You're off to Great Places! You're off and away!

该文本有 65 个字母，4 个句子，14 个单词。文本中每 100 个单词的平均字母数是  $L=65/14*100=464.29$ ；文本中每 100 个单词的平均句子数是  $S=4/14*100=28.57$ 。

代入 Coleman Liau 指标公式，并向最近的整数取整，我们得到可读性指数为 3 级。

我们将编写一个函数 `readability`，输入参数为字符串，返回可读性等级。

**实现要求：**

- (1) 若计算结果小于 1，输出“Before Grade 1”；
- (2) 若计算结果大于或等于 16，输出“Grade 16+”；
- (3) 除 (1) 和 (2) 外，输出“Grade X”，X 为相应等级。
- (4) 字母包括大写字母和小写字母（不考虑数字和标点符号）；
- (5) 以空格分隔作为标准区分单词，如；

It was a bright cold day in April, and the clocks were striking thirteen. Winston Smith, his chin nuzzled into his breast in an effort to escape the vile wind, slipped quickly through the glass doors of Victory Mansions, though not quickly enough to prevent a swirl of gritty dust from entering along with him.

55 words

- (6) 句号 (.)、感叹号 (!) 或问号 (?) 表示句子的结尾。如

Mr. and Mrs. Dursley, of number four Privet Drive, were proud to say that they were perfectly normal, thank you very much.

3 sentences

- (7) 可读性等级的参考例子见 `readability.txt`。