

## 正则表达式 2

戴一帆

2022 年 3 月 10 号

[toc] # 1、从文本提取所有英文单词的第一个字母 **结果：** ['H', 'P', 'a', 't', 'P', 's', 'S', 'i', 't', 'f', 'n', 'i', 't', 'H', 'P', 's', 'w', 'b', 'j', 'K', 'R', 'a', 'f', 'H', 'P']

```
import re
shuju = "Harry Potter and the Philosopher's Stone is the first novel in
the Harry Potter series written by J. K. Rowling and featuring Harry P
otter." # \w : [a-zA-Z0-9_]; \b: 空白字符
x = re.findall(r"\b\w", shuju) # re.findall: 找出所有匹配
print(x)
```

## 2、将三个快递信息全部抽取出来

- 格式按照[(定单编号, 寄件人, 收件人, 寄件地址, 到货地址), (...), (...)]的格式进行输出
- 另要求提取字符串中不应该含有空格。 **结果：** [['1049', '张三', '李四', '珠海市', '武汉市'], ['1050', '王五', '赵六', '天津市', '北京市'], ['1049', '钱七', '周八', '海口市', '广州市']]

```
import re
xinxi = '''
定单编号: 1049
寄件人: 张三
收件人: 李四
寄件地址: 珠海市
到货地址: 武汉市
定单编号: 1050
寄件人: 王五
收件人: 赵六
寄件地址: 天津市
到货地址: 北京市
定单编号: 1049
寄件人: 钱七
收件人: 周八
寄件地址: 海口市
到货地址: 广州市
'''
```



```
y = re.findall( r"定单编号: (\d{4})\n 寄件人: (.+)\n 收件人: (.+)\n
n 寄件地址: (.+)\n 到货地址: ?(.+)", xinxi) # ?: 代表对空格进行是否（该文
段空格不统一）;\n 代表换行; 只对（）内的文本进行匹配
print(y))
```

