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

寄件人:

收件人: 周八 寄件地址: 海口市 到货地址: 广州市

xinxi = '''

钱七



y = re.findall(r"定单编号: (\d{4})\n 寄件人: (.+)\n 收件人: (.+)\ 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))

3、将所有的服务器全部替换成 qq 服务器可以使用的邮箱

-形如 XXXXX@qq.com,.... 结果: 11120015613@qq.com, 11abcakj@qq.com, 11xaksjdnas23@qq.com, abc!askdnla@qq.com, 56378925@qq.com, 1568#@qq.com, 18923@qq.com, 11683@qq.com。

import re

youxiang = '11120015613@qq.com, 11abcakj@google.com, 11xaksjdnas23@1230
6.com, abc!askdnla@gmail.com, 56378925@yoole.com, 1568#@ert.com, 18923@
zhbit.com, 11683@hhhh.com。'

M =re.sub(r"@\w+.com", "@qq.com", youxiang) # re.sub:替换,@\w+.com: 匹配该文段并用@qq.com 替换print(M)

4、按照所有的中英文的标点符号(除%和/n 外)对上文文本进行切分

结果: ['2022 年 CPI 同比涨幅会如何变动?如何实现 3%的目标','政府出台的对中小微企业的帮扶释放了哪些信号','','就业目标新增 1100 万人以上',',','如何看待灵活就业',",'实现共同富裕应当走什么样的路径','','今天','3 月 9 日','晚上','','两会夜话','直播特别节目第五期,全国政协委员杨伟民','王一鸣,全国人大代表张兆安邀您云上夜话,一起来聊聊和 GDP 有关的那些事',",'总台记者','薛晨','刘苏','王新琦','李竟成','颜蔚然',';']

import re

wenzi = '2022 年 CPI 同比涨幅会如何变动?如何实现 3%的目标?政府出台的对中小微企业的帮扶释放了哪些信号?'就业目标新增 1100 万人以上',"如何看待灵活就业?"实现共同富裕应当走什么样的路径?\今天(3月9日)晚上.《两会夜话》直播特别节目第五期,全国政协委员杨伟民/王一鸣,全国人大代表张兆安邀您云上夜话,一起来聊聊和GDP 有关的那些事。(总台记者 薛晨 刘苏 王新琦 李竟成 颜蔚然);'

print(re.split(r"[?/.。''"'《》()()\]", wenzi))# \\:z 在 python 代表\, 因题要求故不能使用\W, 只能对文本中除%和/n 外的符号进行匹配并用 re.split 切分

6、解释在代码行里

