

正则

why

补充知识

形象记忆

定义正则表达式

一般匹配

高级匹配

重复匹配

特殊含义的字符组成一套规则

匹配出符合规则的子字符串

1、用户注册

2、爬虫程序

re.split

可以加入多个字符切分匹配

加入group()

得到返回值

re.rearch

搜索到字符的索引

不能分组

判断是否有

re.match

只取第一个

re.sub

相当于replace

只是加了正则

理解

贪婪匹配

匹配换行符以外的任意一个字符

能多拿就多拿

\*

长度任意可以为0

非贪婪匹配

.\*?

拿到第一个就不拿了

分组

加()

匹配完后进行删选

小练习

联系非贪婪可以找到所以网站信息

(?:字符)

点.

匹配换行符以外的任意一个字符

[]

加了^是取反

匹配一个字符,该字符属于中括号内指定的字符

\*

必须与其他字符连用,代表左侧的字符出现0次或者无穷次

?:

?: 必须与其他字符连用,代表左侧的字符出现0次或者1次

+

必须与其他字符连用,代表左侧的字符出现1次或者无穷次

自己设置{n,m}: 必须与其他字符连用

pattern=re.compile('alex')

语法

re.findall('表达式','字符串')

\w

匹配字母数字或下划线

\W

匹配非字母数字下滑线

\s

匹配任意空白字符

\S

匹配任意非空白字符

\d

匹配任意数字

\D

匹配非任意数字

\n

匹配换行符

\t

匹配制表符

^

匹配字符串开头

\$

匹配字符串结尾

火车头

匹配字符串结尾

火车后面有人丢美元

也可以多个字符匹配

a|b

a或b有就取一个

注意

某些符号要加转义\