**一. re中的sub函数**

使用Python 的re模块，re模块提供了re.sub用于替换字符串中的匹配项。

re.sub(pattern, repl, string, count=0)

参数说明：

* pattern：正则重的模式字符串
* repl：被拿来替换的字符串
* string：要被用于替换的原始字符串
* count：模式匹配后替换的最大次数，省略则默认为0，表示替换所有的匹配

**1.1 提取中文**

可以这样想：我们可以通过将不是中文的字符替换为空不就可以了

例如

import re

str = "重出江湖hello的地H方。。的,world"

str = re.sub("[A-Za-z0-9,。]", "", str)

print(str)

输出：神的孩子在唱歌

**1.2 提取英文**

import re

str = "重123出江湖hello的地H方。。的,world"

str = re.sub("[u4e00-u9fa5-9,。]", "", str)

print(str)

输出：helloHworld

**1.3 提取数字**

import re

str = "重123出江湖hello的地H方。。的,world"

str = re.sub("[A-Za-zu4e00-u9fa5,。]", "", str)

print(str)

输出：123

**二. re中的findall函数**

在字符串中找到正则表达式所匹配的所有子串，并返回一个列表，如果没有找到匹配的，则返回空列表。

语法格式为：

findall(string[, pos[, endpos]])

参数：

* string : 待匹配的字符串。
* pos : 可选参数，指定字符串的起始位置，默认为 0。
* endpos :可选参数，指定字符串的结束位置，默认为字符串的长度。 查找字符串中的所有数字：

扩展：正则中有match 和 search ，它们是是匹配一次，findall

匹配所有，具体了解可以到[菜鸟教程](https://www.runoob.com/python/python-reg-expressions.html)查看

2.1 提取中文



**2.2 提取英文**

通俗写法

import string#提供a-z的小写字母

dd = "神的孩子hello在H唱歌,world"

#准备英文字符

temp=""

letters=string.ascii\_lowercase#包含a-z的小写字母

for word in dd:#for循环取出单个词

if word.lower() in letters:#判断是否是英文

temp+=word#添加组成英文单词

print(temp)

输出：helloHworld

正则

#A-Za-z

import re

dd = "重出123江湖hello的地方的,world"

result = ''.join(re.findall(r'[A-Za-z]', dd))

print(result)

输出：helloHworld

**2.3 提取数字**

#0-9注意这个数字前面不能,要不然他连，都给算上

import re

dd = "神123的孩子hello在唱H歌。。,world"

result = ''.join(re.findall(r'[0-9]', dd))

print(result)

输出：123

**三. re中的compile函数**

compile函数用于编译正则表达式，生成一个正则表达式（ Pattern ）对象，供其他函数使用。

语法格式为：

re.compile(pattern[, flags])

参数：

pattern : 一个字符串形式的正则表达式

flags : 可选，表示匹配模式，比如忽略大小写，多行模式等，具体参数为：

* re.I 忽略大小写
* re.L 表示特殊字符集 w, W, , B, s, S 依赖于当前环境
* re.M 多行模式
* re.S即为 . 并且包括换行符在内的任意字符（. 不包括换行符）
* re.U 表示特殊字符集 w, W, , B, d, D, s,S 依赖于 Unicode 字符属性数据库
* re.X 为了增加可读性，忽略空格和 # 后面的注释

**3.1 同时匹配中英文数字去除其他字符**

