

苏州大学实验报告

院、系	计算机学院	年级专业	21 计科	姓名	方浩楠	学号	2127405048
课程名称	编译原理课程实践					成绩	
指导教师	王中卿	同组实验者	无	实验日期			

实验名称 初次实验（中缀转后缀，简单正则表达式）

一. 实验目的

了解 python 中的中缀表达式和正则表达式

二. 实验内容

1、实现中缀表达式转化为后缀表达式

如将 $9 - 5 + 2$ 转化为 $9\ 5 - 2 +$ ，要求能够中缀表达式中包含四则运算符和括号。

2、正则表达式练习

1) 匹配

验证 jpg,jpeg,gif,bmp 格式的文件名，大小写均可？

匹配日期格式：2/31/2006

电话号码(区号提取)：(0512) 68078800-6852

2) 提取（需要提取的内容在 re_example/ 文件夹中）

提取 list.html 中的所有超链接

提取 content 中的标题与正文

三. 实验步骤和结果

实现中缀表达式转化为后缀表达式

首先添加函数来判断运算符优先级

其次添加两个 list,一个当做栈,用来存放运算符,另一个用来存放最终输出结果

然后依次遍历中缀表达式,利用中缀表达式转后缀表达式的过程来运算

运算结果:

```
print("9 - 5 + 2的中缀表达式转后缀表达式的结果是:",end="")
print(infix_to_postfix('9 - 5 + 2'))
```

9 - 5 + 2的中缀表达式转后缀表达式的结果是:9 5 - 2 +

验证 jpg,jpeg,gif,bmp 格式的文件名，大小写均可

需要匹配有.jpg,.jpeg,.gif,.bmp 的字符串,并且必须是字符串的底部有这些字符

因此, 首先需要用正则表达式 `r' .+\. (jpg|jpeg|gif|bmp)'` 来匹配字符串,

使用 `is_match = re.match(pattern, lowercase_filename)` 来判断是否匹配.匹配后判断这些是否在字符串的尾部,函数最终返回

```
return is_match and lowercase_filename.split('.')[-1] in ['jpg', 'jpeg', 'gif', 'bmp']
```

运算结果:

```
print(validate_filename('0f0002JnyrTgRygqLGhM76.jpg'))
```

True

匹配日期格式:

正则表达式为 `r'\d{1,2}/\d{1,2}/\d{1,4}'`

用来匹配开头为一到两个数字,然后是/,然后是一到两个数字,然后/,结尾是四位数字的字符串
运行结果:

```
print(date_match('safsadsfas30/9/2001sfsadfa'))
```

30/9/2001

提取电话号码区号:

关键在于提取字符串中括号内的数字

使用正则表达式 `r'\((\d+)\)'`

输入(0512)\68078800-6852),运行结果:

0512

提取所有超链接:

正则表达式为 `r'<a\s+(?:[^\>]*?\s+)?href="([^\"]*)"'`

需要提取的内容为 `href="****"` 其中,双引号内的内容。

该正则表达式需要匹配形如 `clickhere` 的字符串,并且提取出 `https://www.example.com`

提取 list.html 的结果:

[output\link.html.txt](#)

提取正文和标题:

提取标题的正则表达式为: `r'<title>(.*?)</title>'`

该正则表达式用来提取两个 `title` 标签中的内容

提取正文的正则表达式为: `r'<p>(.*?)</p>'`

该正则表达式用于提取两个 `<p>` 标签内的正文内容

提取 content.html 的结果:

[output\content.html.txt](#)

四. 实验总结

本次实验让我了解到了 python 中的中缀表达式转后缀表达式,以及使用 python 的正则表达式