# 11 让正则飞起来：PCRE 库的相识相知

更新时间：2020-01-02 16:52:43

> 只有在那崎岖的小路上不畏艰险奋勇攀登的人,才有希望达到光辉的顶点。——马克思

**前言**

我们知道 Nginx 的 Server , Location ， Rewrite 语法非常的强大，我们可以灵活的配置从而达到各种各样的目的。这一切的一切都离不开强大的正则表达式。那么什么是正则表达式呢？

> 正则表达式（ Regular Expression ），是指一个用来描述或者匹配一系列符合某个句法规则的字符串的单个字符串 。通常用于检索或替换那些符合某个模式的文本内容。

其实，正则表达式在我们平时工作中应该经常的用到，几乎所有的编程语言都有自己的正则解析库。比如世界上最好的语言 PHP 就有支持 POSIX 和 PCRE 两套正则表达式。

> what? 正则表达式还有多种吗？
> 是的，现在有多种正则表达式标准，比如上面说到的 POSIX 和 PCRE 。其中， Nginx 使用的 PCRE 语法。大家不用担心，这两种标准基本上是一样的，只有一些细微处有些不同。

**PCRE详解**

PCRE 的全称是 Perl Compatible Regular Expression,也就是兼容 Perl 正则规范。 PCRE 规范的内容其实很多，但是经常使用的并不多，我在本文中给大家介绍一下我在工作中经常使用的正则语法，足以满足大家平时的工作要求。

> 二八原则：其实使用到的也就是整个规范中的 20% 而已。

## 元字符

和编程语言中的关键字一样，元字符就是那些在正则语法中具有特殊意义的字符。我把元字符分为了几种：

- 匹配特殊位置/字符的元字符
- 数量元字符
- 减少工作量的语法糖
- 模式字符

### 匹配特殊位置/字符

| 字符 | 特殊含义 |
| --- | --- |
| ^ | 匹配一行的开头 |
| $ | 匹配一行的结尾 |
| . | 匹配除换行符以外的所有字符 |
| [abc] | 可以匹配abc中的任意一个字符 |
| - | 和上面的方括号配合用于区间匹配 |

比如 ^abc.def$ 就可以匹配到以 abc 开头，以 def 结尾，并且 abc 和 def 中间有一个非换行的任意字符，比如 abcydef 字符串就满足匹配条件。

### 数量元字符

这些字符的作用就是对出现在前面的单元进行重复的匹配。

| 字符 | 特殊含义 |
| --- | --- |
| ? | 匹配前面的字符零次或一次 |
| + | 匹配前面的字符一次或多次 |
| * | 匹配前面的字符0次或多次 |
| {n} | 匹配前面的字符n次 |
| {n,} | 匹配前面的字符n次或更多次，至少匹配n次 |
| {m,n} | 匹配前面的字符的次数在m和n之间 |

## 语法糖

我们可以使用 [0-9] 来匹配数字，使用 [a-zA-Z] 来匹配，这样写完全没有问题。但是，这种类似的情况会经常出现在我们平时的工作中，作为程序员，我们要发扬"绝不多写一个字符"的优良传统，所以， PCRE 提供了一些 语法糖 来解放我们的双手。

| 字符 | 特殊含义 |
| --- | --- |
| \w | 匹配所有的字符，相当于 [0-9a-zA-Z_]，即匹配字母，数字和下划线 |
| \d | 相当于[0-9],匹配所有的数字 |
| \s | 匹配所有的空白字符，比如空格，制表符，换行符 |

PCRE 其实还提供了其他的类似 语法糖 ，但是在工作中经常用到的就是上面的三个。

## 分组捕获和反向引用

Q1: 什么是分组？

A1: 其实分组就是用括号把正则表达式括起来

Q2: 为什么要分组？

A2: 为了后面使用分组中的捕获的内容

Q3: 如何给分组编号？

A3: 从 1 开始，自左向右，遇到一个左括号编号就加 1 。

Q4: 如何使用捕获分组?

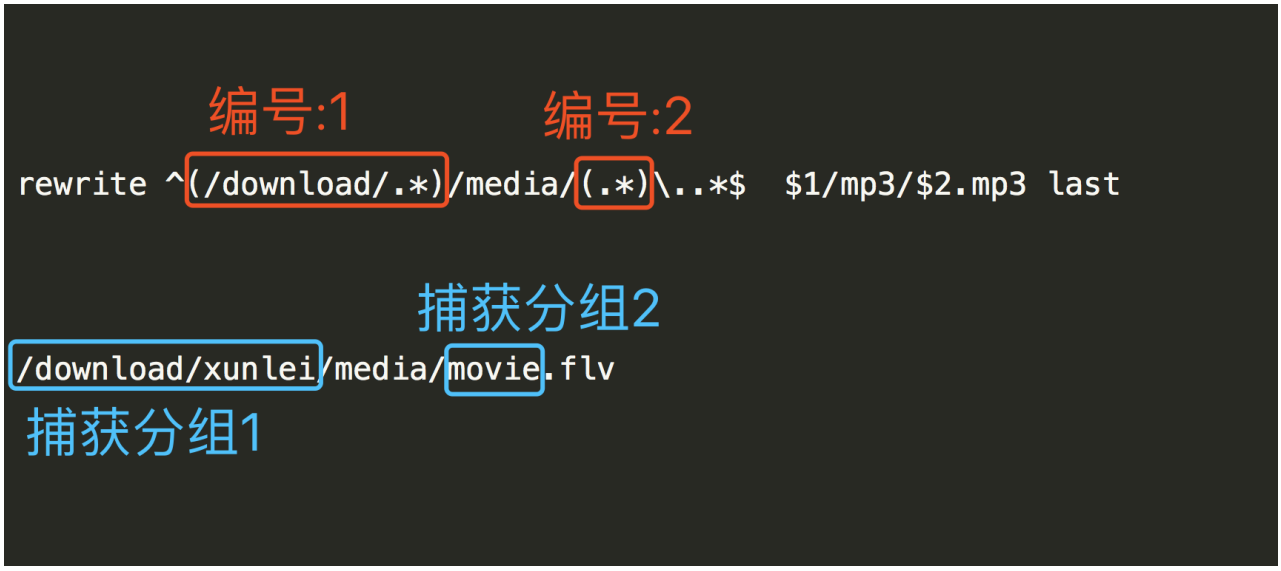
A4: 使用 $n 即可，其中 n 是分组的编号。

如果大家明白了上面的几个问题，那么这部分内容就应该理解了。

我们用一个例子来说明这部分内容非常合适。

```
rewrite ^(/download/.*)/media/(.*)\..*$  $1/mp3/$2.mp3 last
```

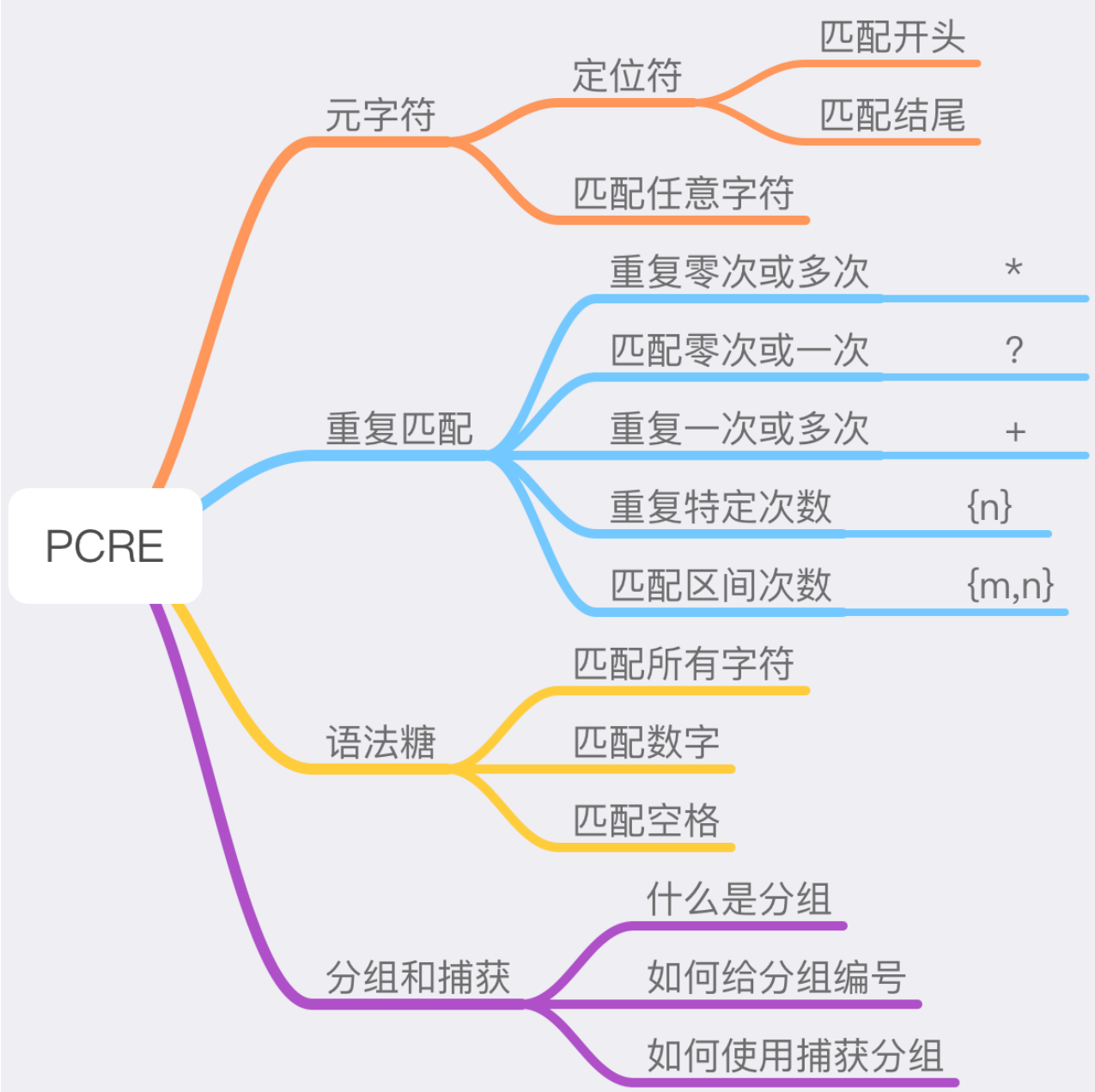
如果我们请求的 path 是 /download/xunlei/media/movie.flv ，那么实际的请求资源就是 /download/xunlei/mp3/movie.mp3 。




> 犹记得我刚学习正则表达式的时候，学习分组和捕获这部分内容时超级痛苦，完全搞不懂什么是捕获。

写在最后

在本章中，我们介绍了Nginx使用的PCRE常用语法，这部分比较理论化，但是对理解后面的内容非常重要，希望大家多看多练。




}

10 条条大路通罗马，形形色色的虚拟主机

12 深入浅出 Rewrite 语法