

第 9 章 正则表达式

学习要点:

- 1.正则表达式语法 (Perl 风格)
- 2.正则表达式中的元素
- 3.Pperl 风格函数

主讲教师: 李炎恢

合作网站: <http://www.ibeifeng.com>

讲师博客: <http://hi.baidu.com/李炎恢>

处理字符串时,有很多较为复杂的字符串用普通的字符串处理函数无法干净地完成。比如说,可能需要验证一个 Email 地址是否合法,为此需要查看许多不容易检查的规则。这正是正则表达式的用武之地。正则表达式是功能强大而简明的字符组,其中可以包含大量的逻辑,特别值得一提的是正则表达式相当简短。

一. 正则表达式语法 (Perl 风格)

Perl 一直被认为是最伟大的解析语言之一,它提供了一种全面的正则表达式,即使是最复杂的字符串模式,也可以用这种正则表达式语言搜索和替换。PHP 开发人员认识到,与其重新发明正则表达式,不如让 PHP 用户直接使用声名赫赫的 Perl 正则表达式语言,即 Perl 风格的函数。

模式规则: /php/ 在字符串前后加上两条斜杠即可。

匹配函数: preg_match()函数在字符串中搜索模式,如果存在则返回 true,否则返回 false。

```
<?php  
    preg_match('/php/','php');  
?>
```

二. 正则表达式中的元素

正则表达式中包含三种元素分别为:量词、元字符、修饰符

量词

语法	描述
+	匹配任何至少包含一个前导字符串
*	匹配任何包含零个或多个前导字符串
?	匹配任何包含零个或一个前导字符串
.	匹配任意字符串
{x}	匹配任何包含 x 个前导字符串
{x,y}	匹配任何包含 x 到 y 个前导字符串

{x,}	匹配任何包含至少 x 个前导字符串
\$	匹配字符串的行尾
^	匹配字符串的行首
	匹配字符串的左边或者右边
()	包围一个字符分组或定义个反引用，可以使用\1\2 提取

元字符

语法	描述
[a-z]	匹配任何包含小写字母 a-z 的字符串
[A-Z]	匹配任何包含大写字母 A-Z 的字符串
[0-9]	匹配任何包含数字 0-9 的字符串
[abc]	匹配任何包含小写字母 a、b、c 的字符串
[^abc]	匹配任何不包含小写字母 a、b、c 的字符串
[a-zA-Z0-9_]	匹配任何包含 a-zA-Z0-9 和下划线的字符串
\w	匹配任何包含 a-zA-Z0-9 和下划线的字符串（同上）
\W	匹配任何没有下划线和字母数字的字符串
\d	匹配任何数字字符，和[0-9]相同
\D	匹配任何非数字字符，和[^0-9]相同
\s	匹配任何空白字符
\S	匹配任何非空白字符
\b	匹配是否到达了单词边界
\B	匹配是否没有达到单词边界
\	匹配正则中的特殊字符

修饰符

语法	描述
i	完成不区分大小写的搜索
m	在匹配首内容或者尾内容时候采用多行识别匹配
x	忽略正则中的空白
A	强制从头开始匹配
U	禁止贪婪匹配 只跟踪到最近的一个匹配符并结束

三. Perl 风格函数

PHP 为使用 Perl 兼容的正则表达式搜索字符串提供了 7 个函数，包括：preg_grep()、preg_match()、preg_match_all()、preg_quote()、preg_replace()、preg_replace_callback() 和 preg_split()。

搜索字符串：preg_grep() 函数搜索数组中的所有元素，返回由与某个模式匹配的所有元素组成的数组。

```
<?php
    $language = array('php','asp','jsp','python','ruby');
    print_r(preg_grep('/p$',$language));
?>
```

搜索模式：preg_match() 函数在字符串中搜索模式，如果存在则返回 true，否则返回 false。

```
<?php
    echo preg_match('/php[1-6]/','php5');
?>
```

电子邮件验证小案例（分组应用）

```
<?php
    $mode = '/([\w\.\_]{2,10})@(\w{1,}).([a-z]{2,4})/';
    $string = 'yc60.com@gmail.com';
    echo preg_match($mode,$string);
?>
```

匹配模式的所有出现：preg_match_all() 函数在字符串中匹配模式的所有出现，然后将所有匹配到的全部放入数组。

```
<?php
    preg_match_all('/php[1-6]/','php5sdfphp4sdflljkphp3sdfjphp2',$out);
    print_r($out);
?>
```

定界特殊的正则表达式：preg_quote() 在每个对于正则表达式语法而言有特殊含义的字符前插入一个反斜线。这些特殊字符包含：\$ ^ * () += { } [] | \ : < >。

```
<?php
    echo preg_quote('PHP的价格是：$150');
?>
```

替换模式的所有出现：preg_replace() 函数搜索到所有匹配，然后替换成想要的字符串返回出来。

```
<?php
    echo preg_replace('/php[1-6]/','python','This is a php5,This is a php4');
?>
```

ubb 小案例：贪婪问题+分组使用()

```
<?php
    $mode = '/[b\](.*)[/b\]/U';
    $replace = '<strong>\1</strong>';
    $string = 'This is a [b]php5[/b],This is a [b]php4[/b]';
    echo preg_replace($mode,$replace,$string);
?>
```

以不区分大小写的方式将字符串划分为不同的元素：preg_split()用来分割不同的元素。

```
<?php
    print_r(preg_split('/[. @]/','yc60.com@gmail.com' ));
?>
```

注：目前为 PHP 使用 POSIX 风格的正则表达式搜索字符串提供了 7 个函数，包括：ereg()、ereg_replace()、eregi()、eregi_replace()、split()、spliti() 和 sql_regcase()。

PS：和 Perl 风格基本相同，可以参考手册自行学习。

感谢收看本次教程！

本课程是由北风网(ibeifeng.com)

瓢城 Web 俱乐部(yc60.com)联合提供：

本次主讲老师：李炎恢

我的博客：hi.baidu.com/李炎恢/

我的邮件：yc60.com@gmail.com