

正则表达式快速参考手册

胡志飞 <WisdomFusion#gmail.com>

2012 年 6 月 2 日

目 录

1	简介	1
2	基本语法	1
3	高级语法	5
4	举些栗子	7
5	正则表达式“流派”	7
6	应用场景	8
6.1	正则表达式工具箱	8
6.2	应用案例	8

1 简介 INTRODUCTION

正则表达式，（**Regular Expression**，在代码中常简写为 **regex**、**regexp** 或 **RE**），计算机科学的一个概念。正则表达式使用字符来描述、匹配一系列符合某个句法规则的字符串。在很多文本编辑器里，正则表达式通常被用来检索、替换那些符合某个模式的文本。许多程序设计语言都支持利用正则表达式进行字符串操作。例如，在 **Perl**¹中就内建了一个功能强大的正则表达式引擎。正则表达式这个概念最初是由 **Unix** 中的工具软件（例如 **sed**²和 **grep**³）普及开的。

需要注意的是，用什么工具，用什么编辑语言，正则表达式的语法有些差别，特性的支持也参差不齐，称之为正则表达式“流派”（第5部分详述），所以要单独参考工具和编程语言本身的文档才行。

2 基本语法 BASIC SYNTAX

特性	语法	描述	举个栗子
字符	除 <code>[^\\$. ?*\+()</code> 以外的任意字符	除了 <code>[^\\$. ?*\+()</code> 以外的任意字符， <code>{</code> 和 <code>}</code> 也是文字文本，除了下面说到的成对出现的量词语法，如 <code>{n}</code> 和 <code>{m,n}</code> 等。	<code>a</code> 匹配 <code>about</code> 中的 <code>a</code>
转义	<code>\t</code> , <code>\?</code> , <code>*</code> , <code>\+</code> , <code>\.</code> , <code>\ </code> , <code>\{</code> , <code>\}</code> , <code>\\</code> , <code>\[</code> , <code>\]</code> , <code>\(</code> , <code>\)</code>		
	<code>\n</code> , <code>\r</code>	Windows 文件格式换行符是 <code>\r\n</code> ，UNIX 文件格式换行符是 <code>\n</code>	
	<code>\cA</code> 到 <code>\cZ</code> , <code>\ca</code> 到 <code>\cz</code>	Ctrl+A 到 Ctrl+Z, Ctrl+a 到 Ctrl+z	

To be continued...

¹**Perl**被称为“实用报表提取语言”（**Practical Extraction and Report Language**），正则表达式特性的推动者，文本处理非常方便。
²**sed**是一种 **UNIX/Linux** 平台下的轻量级流编辑器，日常一般用于处理文本文件。
³**grep**, **global search regular expression and print out the line**，是一种强大的文本搜索工具，它能使用正则表达式搜索文本，并把匹配的行打印出来。

(续表)

特性	语法	描述	举个例子
	<div>\a</div> <div>\f</div> <div>\v</div>		
基本特性	<div>.</div> （点）	匹配除换行符之外的任意字符，有些正则表达式“流派”还支持点是否匹配换行符的开关。	<div>.</div> 匹配 about 中的任意一个字符
	<div> </div>	管道，或的关系，匹配 的左侧或右侧的字符串	<div>abc def xyz</div> 匹配 <div>abc</div> 或 <div>def</div> 或 <div>xyz</div>
字符类	<div>[abc]</div> , <div>[\^\]]</div> <div>[^abc]</div> <div>\d</div> , <div>\D</div> <div>\w</div> , <div>\W</div> <div>\s</div> , <div>\S</div>		
POSIX	<div>[:alnum:]</div> <div>[:alpha:]</div> <div>[:ascii:]</div>		

To be continued...

(续表)

特性	语法	描述	举个栗子
	<code>[:blank:]</code>		
	<code>[:cntrl:]</code>		
	<code>[:digit:]</code>		
	<code>[:graph:]</code>		
	<code>[:lower:]</code>		
	<code>[:print:]</code>		
	<code>[:punct:]</code>		
	<code>[:space:]</code>		
	<code>[:upper:]</code>		
	<code>[:word:]</code>		
	<code>[:xdigit:]</code>		
锚点	<code>^</code>		
	<code>\$</code>		
	<code>\A</code>		

To be continued...

(续表)

特性	语法	描述	举个栗子
	\G		
	\Z		
	\b		
	\B		
	\<		
	\>		
	\`		
	\&		
	\'		
量词	?		
	??		
	?+		
	*		

To be continued...

(续表)

特性	语法	描述	举个栗子
	*?		
	*+		
	+		
	+?		
	++		
	{n}		
	{n,m}		
	{n,}		
	{,m}		
	{n,m}?		
	{n,m}+		

3 高级语法 **ADVANCED SYNTAX**

(续表)

特性	语法	描述	举个栗子
	\k<name>, \g{name}		
高级分组	(?#comment)		
	(?=Regex)		
	(?!Regex)		
	(?<=regex)		
	(?<!regex)		

4 举些栗子 REGEX EXAMPLES

一些栗子

5 正则表达式“流派” REGEX FLAVORS

流派

6 应用场景 APPLICATION SCENARIOS

6.1 正则表达式工具箱 REGEX TOOLBOX

总有一款适合你，Windows 下的记事本太鸡肋，Word 处理方式主要是“通配符”而不是正则表达式。

RegexBuddy

JGsoft 开发的一个强大的正则表达式测试工具，

grep

PowerGREP

RegexBuddy 的兄弟软件，同是 JGsoft 开发，是 grep 在 Windows

平台的实现和增强。

UltraEdit, Notepad++

Vim

GNU Emacs

sed & awk

6.2 应用案例 APPLICATION CASES

Dreamweaver 表格处理

VBA 中使用正则表达式

```
1 Sub IndentParaWithRegEx()  
2 ' PowerPoint VBA 批量给指定字符开头段落加动画  
3 Dim oSld As Slide  
4 Dim oShp As Shape  
5 Dim i As Integer  
6 ' 正则相关变量  
7 Dim regx As Object, oMatch As Object  
8  
9 ' 这里写查找的正则，参考 http://msdn.microsoft.com/en-us/library/ms974570.aspx  
10 strPattern = "^开头字符串"  
11
```

```
12 Set regx = CreateObject("vbscript.regexp")
13 With regx
14     .Global = True
15     .IgnoreCase = True
16     .Pattern = strPattern
17 End With
18
19 For Each oSld In ActivePresentation.Slides
20     For Each oShp In oSld.Shapes
21         If oShp.HasTextFrame Then
22             If oShp.TextFrame2.HasText Then
23                 With oShp.TextFrame2.TextRange
24                     For i = 1 To .Paragraphs.Count
25                         With .Paragraphs(i)
26                             ' 可能会出现多个匹配项的
27                             If (regx.Test(.Text) = True) Then
28                                 .ParagraphFormat.FirstLineIndent = 0
29                             End If
30                         End With
31                     Next i 'para
32                 End With
33             End If 'has text
34         End If 'has textframe
35     Next oShp
36 Next oSld
37 End Sub
```

InDesign GREP

使用 **Perl** 正则表达式处理文件
神的编辑器之正则