

常用正则表达式

正则表达式用于字符串处理、表单验证等场合，实用高效。现将一些常用的表达式收集于此，以备不时之需。

匹配中文字符的正则表达式：[\u4e00-\u9fa5]

评注：匹配中文还真是个头疼的事，有了这个表达式就好办了

匹配双字节字符(包括汉字在内)：[\x00-\xff]

评注：可以用来计算字符串的长度（一个双字节字符长度计 2，ASCII 字符计 1）

匹配空白行的正则表达式：\n\s*\r

评注：可以用来删除空白行

匹配 HTML 标记的正则表达式：<(\S*?)[^>]*.*?</\1><.*?/>

评注：网上流传的版本太糟糕，上面这个也仅仅能匹配部分，对于复杂的嵌套标记依旧无能为力

匹配首尾空白字符的正则表达式：^\s*\s*\$

评注：可以用来删除行首行尾的空白字符(包括空格、制表符、换页符等等)，非常有用的表达式

匹配 Email 地址的正则表达式：\w+([-+.]\w+)*@\w+([-.\w+)*\.\w+([-.\w+)*

评注：表单验证时很实用

匹配网址 URL 的正则表达式：[a-zA-z]+://[^\s]*

评注：网上流传的版本功能很有限，上面这个基本可以满足需求

匹配帐号是否合法(字母开头，允许 5-16 字节，允许字母数字下划线)：^[a-zA-Z][a-zA-Z0-9_]{4,15}\$

评注：表单验证时很实用

匹配国内电话号码：\d{3}-\d{8}|\d{4}-\d{7}

评注：匹配形式如 0511-4405222 或 021-87888822

匹配腾讯 QQ 号：[1-9][0-9]{4,}

评注：腾讯 QQ 号从 10000 开始

匹配中国邮政编码: [1-9]\d{5}(?!d)

评注: 中国邮政编码为 6 位数字

匹配身份证: \d{15}|\d{18}

评注: 中国的身份证为 15 位或 18 位

匹配 ip 地址: \d+\.\d+\.\d+\.\d+

评注: 提取 ip 地址时有用

匹配特定数字:

^[1-9]\d*\$ //匹配正整数

^-[1-9]\d*\$ //匹配负整数

^-?[1-9]\d*\$ //匹配整数

^[1-9]\d*0\$ //匹配非负整数 (正整数 + 0)

^-[1-9]\d*0\$ //匹配非正整数 (负整数 + 0)

^[1-9]\d*\.\d*0\.\d*[1-9]\d*\$ //匹配正浮点数

^-([1-9]\d*\.\d*0\.\d*[1-9]\d*)\$ //匹配负浮点数

^-?([1-9]\d*\.\d*0\.\d*[1-9]\d*0?\.\d*0+0)\$ //匹配浮点数

^[1-9]\d*\.\d*0\.\d*[1-9]\d*0?\.\d*0+0\$ //匹配非负浮点数 (正浮点数 + 0)

^-([1-9]\d*\.\d*0\.\d*[1-9]\d*)0?\.\d*0+0\$ //匹配非正浮点数 (负浮点数 + 0)

评注: 处理大量数据时有用, 具体应用时注意修正

匹配特定字符串:

^[A-Za-z]+\$ //匹配由 26 个英文字母组成的字符串

^[A-Z]+\$ //匹配由 26 个英文字母的大写组成的字符串

^[a-z]+\$ //匹配由 26 个英文字母的小写组成的字符串

^[A-Za-z0-9]+\$ //匹配由数字和 26 个英文字母组成的字符串

^[w]+\$ //匹配由数字、26 个英文字母或者下划线组成的字符串

评注: 最基本也是最常用的一些表达式

原载地址: <http://lifesinger.3322.org/myblog/?p=185>

原创文章如转载, 请注明: 转载自月光博客 [<http://www.williamlong.info/>]

本文链接地址: <http://www.williamlong.info/archives/433.html>