C#正则表达式（二）

正则表达式是一种匹配文本的模式。.NET框架提供了允许这种匹配的正则表达式引擎，模式由一个或多个字符、运算符和结构组成。

C#中正则表达式有以下几种：

1. 字符转义

正则表达式中的反斜杠“\”指示以下值之一：

①后接字符为特殊字符，例如，\b 是指示正则表达式匹配应从单词边界开始的定位点，\t表示制表符，而 \x020 表示空间。

②未转义语言构造的字符应按字面意思进行解释，例如，大括号 ({) 开始定义限定符，而反斜杠后接大括号 (\{) 表示正则表达式引擎应匹配大括号。 同样，单个反斜杠标记转义的语言构造的开始，而两个反斜杠 (\\) 表示正则表达式引擎应匹配反斜杠。

正则表达式中的反斜杠字符（\）指示其后跟的字符是特殊字符，或应按原义解释该字符。

下表列出了转义字符：



正则表达式 \G(.+)[\t|\u007c](.+)\r?\n 可以解释为下表中所示内容。

| **模式** | **描述** |
| --- | --- |
| \G | 从上次匹配结束处开始匹配。 |
| (.+) | 一次或多次匹配任何字符。 这是第一个捕获组。 |
| [\t\u007c] | 匹配制表符 (\t) 或垂直条 (|)。 |
| (.+) | 一次或多次匹配任何字符。 这是第二个捕获组。 |
| \r? \n | 匹配零或一个出现回车符后接新行的次数。 |

1. 字符类

一个字符类定义一组字符，其中的任一字符均可出现在输入字符串中以便成功匹配。

下表列出了字符类：



正字符组：[]

负字符组：[^]

任意字符：.

Unicode类别或Unicode块：\p{}

负Unicode类别或Unicode块：\P{}

https://msdn.microsoft.com/zh-cn/library/20bw873z(v=vs.110).aspx

<http://www.runoob.com/csharp/csharp-regular-expressions.html>

1. 定位点

定位点或原子零宽度断言会使匹配成功或失败，具体取决于字符串中的当前位置，但它们不会使引擎在字符串中前进或使用字符。

下表列出了定位点：



1. 分组构造

分组构造描述了正则表达式的子表达式，通常用于捕获输入字符串的子字符串。

下表列出了分组构造：



1. 限定符
2. 反向引用构造
3. 备用构造
4. 替换
5. 杂项构造