

Java中正则表达式使用方法详解（一）PDF转换可能丢失图片或格式，建议阅读原文

https://www.100test.com/kao_ti2020/145/2021_2022_Java_E4_B8_AD_E6_AD_A3_c104_145090.htm 如果你曾经用过Perl或任何其他内建正则表达式支持的语言，你一定知道用正则表达式处理文本和匹配模式是多么简单。如果你不熟悉这个术语，那么“正则表达式”（Regular Expression）就是一个字符构成的串，它定义了一个用来搜索匹配字符串的模式。许多语言，包括Perl、PHP、Python、JavaScript和JScript，都支持用正则表达式处理文本，一些文本编辑器用正则表达式实现高级“搜索-替换”功能。那么Java又怎样呢？本文写作时，一个包含了用正则表达式进行文本处理的Java规范需求

（Specification Request）已经得到认可，你可以期待在JDK的下一版本中看到它。然而，如果现在就需要使用正则表达式，又该怎么办呢？你可以从Apache.org下载源代码开放的Jakarta-ORO库。本文接下来的内容先简要地介绍正则表达式的入门知识，然后以Jakarta-ORO API为例介绍如何使用正则表达式。

一、正则表达式基础知识 我们先从简单的开始。假设你要搜索一个包含字符“cat”的字符串，搜索用的正则表达式就是“cat”。如果搜索对大小写不敏感，单词“catalog”、“Catherine”、“sophisticated”都可以匹配。也就是说：

1.1 句点符号 假设你在玩英文拼字游戏，想要找出三个字母的单词，而且这些单词必须以“t”字母开头，以“n”字母结束。另外，假设有一本英文字典，你可以用正则表达式搜索它的全部内容。要构造出这个正则表达式，你可以使用一个通配符句点符号“.”。这样，完整的表达式就是“t.n

”，它匹配“tan”、“ten”、“tin”和“ton”，还匹配“t#n”、“tpn”甚至“tn”，还有其他许多无意义的组合。这是因为句点符号匹配所有字符，包括空格、Tab字符甚至换行符：

1.2 方括号符号 为了解决句点符号匹配范围过于广泛这一问题，你可以在方括号（“[]”）里面指定看来有意义的字符。此时，只有方括号里面指定的字符才参与匹配。也就是说，正则表达式“t[aeio]n”只匹配“tan”、“Ten”、“tin”和“ton”。但“Toon”不匹配，因为在方括号之内你只能匹配单个字符：

1.3 “或”符号 如果除了上面匹配的所有单词之外，你还想要匹配“toon”，那么，你可以使用“|”操作符。“|”操作符的基本意义就是“或”运算。要匹配“toon”，使用“t(a|e|i|o|oo)n”正则表达式。这里不能使用方括号，因为方括号只允许匹配单个字符；这里必须使用圆括号“（）”。圆括号还可以用来分组，具体请参见后面介绍。

100Test 下载频道开通，各类考试题目直接下载。详细请访问 www.100test.com