

Java中正则表达式使用方法详解（三）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_145084.htm 三、应用实例 下面我们来看看Jakarta-ORO库的一些应用实例。

3.1 日志文件处理 任务：

分析一个Web服务器日志文件，确定每一个用户花在网站上的时间。在典型的BEA WebLogic日志文件中，日志记录的格式如下：分析这个日志记录，可以发现，要从这个日志文件提取的内容有两项：IP地址和页面访问时间。你可以用分组符号（圆括号）从日志记录提取出IP地址和时间标记。首先我们来看看IP地址。IP地址有4个字节构成，每一个字节的值在0到255之间，各个字节通过一个句点分隔。因此，IP地址中的每一个字节有至少一个、最多三个数字。图八显示了为IP地址编写的正则表达式：图八：匹配IP地址 IP地址中的句点字符必须进行转义处理（前面加上“\”），因为IP地址中的句点具有它本来的含义，而不是采用正则表达式语法中的特殊含义。句点在正则表达式中的特殊含义本文前面已经介绍。日志记录的时间部分由一对方括号包围。你可以按照如下思路提取出方括号里面的所有内容：首先搜索起始方括号字符（“[”），提取出所有不超过结束方括号字符（“]”）的内容，向前寻找直至找到结束方括号字符。图九显示了这部分的正则表达式。图九：匹配至少一个字符，直至找到“]”现在，把上述两个正则表达式加上分组符号（圆括号）后合并成单个表达式，这样就可以从日志记录提取出IP地址和时间。注意，为了匹配“--”（但不提取它），正则表达式中间加入了“\s-\s-\s”。完整的正则表达式如图十所示

。图十：匹配IP地址和时间标记 现在正则表达式已经编写完毕，接下来可以编写使用正则表达式库的Java代码了。

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