

利用正则表达式进行中文排版的实例教程

作者：有故事的人 来源：范文网 www.wtabcd.cn/fanwen/

本文原地址：<https://www.wtabcd.cn/fanwen/zuowen/fdba17dca5d0f3447a74e6291bfa6504.html>

范文网，为你加油喝彩！

目录

前言〇、常见匹配类一、段落排版类1、非段落换行排版（要求段落之间空一行）2、插入空行3、删除空行4、行首批量添加（替换项为2个空格时可实现首行缩进）5、行尾批量添加6、半角引号替换为全角引号7、半角单引号替换全角8、固定宽度加硬回车（下例为宽度50）二、删除特定字符串类1、删除行前空格2、删除行尾空格3、删除包含指定关键字的行4、删除以指定关键字开头的行5、删除html注释6、删除html标签7、删除html标签内的全部三、中文汉字处理类1、中文和英文之间加空格2、英文和中文之间加空格3、中文和数字之间加空格4、数字和中文之间加空格5、上述替换整合6、删除硬回车四、章节处理类1、序号变为章节2、章节之间加换行3、章节样式替换五、巧妙利用正则表达式排版总结

前言

中文排版是字处理的一个基本功能，目前有些文字排版软件，例如“排版助手”等等，可以完成一些通用的中文排版功能，但是对于一些复杂的个性化排版功能就无法实现，实际上，我们通过在一些常用的支持正则的代码编辑软件里，使用正则表达式，就可以完成很多复杂的排版功能，下面将举一些常用的例子。

代码编辑软件可使用notepad++或editplus，两者的正则表达式对于换行有一点点差别，一个用\r\n，另一个用\n。下面以editplus为例，讲述一下操作方法，先用editplus打开一个中文文件，然后点“替换”，点击“使用正则表达式”，然后在“查找项”和“替换项”里输入以下字符串，即可实现特定的排版功能。

此外，在“替换”界面中点“常用项”，可以把经常使用的规则保存为功能列表，这样，使用某个功能的时候，直接调用即可获取相关代码。

〇、常见匹配类

匹配中文字符的正则表达式：[\u4e00-\u9fa5]

匹配双字节字符(包括汉字在内)：[^\\x00-\\xff]

匹配特殊字符：((?=\\x21-\\x7e)+)[^a-zA-Z0-9])

匹配所有的半角空白字符：[\f\r\t\n]

匹配隐藏字符：[\u0000-\u001f]

一、段落排版类

1、非段落换行排版（要求段落之间空一行）

查找项：(^[\n])\n

替换项：\1

2、插入空行

查找项：\n

替换项：\n\n

3、删除空行

查找项：^\s*\n

替换项：空

4、行首批量添加（替换项为2个空格时可实现首行缩进）

查找项：^

替换项：我是行首aaa

5、行尾批量添加

查找项：\n

替换项：bbb我是行尾\n

6、半角引号替换为全角引号

查找项：\" ([^\"]\')*\")\"

替换项：“\1”

7、半角单引号替换全角

查找项：\([^\n\r]*\)\n' 陈奕迅路一直都在

替换项：'\1'

8、固定宽度加硬回车（下例为宽度50）

查找项：(.{1,50})

替换项：\1\n

二、删除特定字符串类

1、删除行前空格

查找项：^[\s]*

替换项：空

2、删除行尾空格

查找项：[\s]*\$

替换项：空

3、删除包含指定关键字的行

查找项：^.*(指定关键字).*\$

替换项：空

4、删除以指定关键字开头的行

查找项：^(\指定关键字)[^\n]+\\n

替换项：空

5、删除html注释

查找项：<!--(.*)-->

替换项：空

6、删除html标签

查找项：<[^>]+>

替换项：空

7、删除html标签内的全部

查找项：`<共青团员申请书;(.*>.*<\\1>|<(.*>|<`

替换项：空

三、中文汉字处理类

1、中文和英文之间加空格

查找项：`([\u4e00-\u9fa5]+)([a-zA-Z]+)`

替换项：`\1 \2`

2、英文和中文之间加空格

查找项：`([a-zA-Z]+)([\u4e00-\u9fa5]+)`

替换项：`\1 \2`

3、中文和数字之间加空格

查找项：`([\u4e00-\u9fa5]+)([0-9]+)`

替换项：`\1 \2`

4、数字和中文之间加空格

查找项：`([0-9]+)([\u4e00-\u9fa5]+)`

替换项：`\1 \2`

5、上述替换整合

查找项：`([\u4e00-\u9fa5]+)([a-zA-Z]+)|([a-zA-Z]+)([\u4e00-\u9fa5]+)|([0-9]+)([\u4e00-\u9fa5]+)|([\u4e00-\u9fa5]+)([0-9]+)`

6、删除硬回车

查找项：`([0-9]+)([\u4e00-\u9fa5]+)`

替换项：`\1 \2`

四、章节处理类

1、序号变为章节

查找项 : \n([\u4e00-\u9fa5]{1,2})\n

替换项 : \n第\1章\n

2、章节之间加换行

查找项 : 第(.*章

替换项 : \n第\1章

3、章节样式替换

查找项 : <p>(第.{1,5}章.**)</p>

替换项 : <牛顿第一定律说课稿;h1>\1</h1>

五、巧妙利用正则表达式排版

中文和英文之间加空格arc是什么

([\u4e00-\u9fa5]+)([a-zA-Z]+)

然后使用

\$1 \$2

中文和数字之间加空格

([0-9]+)([\u4e00-\u9fa5]+)

然后使用

\$1 \$2

一下替换

([\u4e00-\u9fa5]+)([a-zA-Z]+)|([a-zA-Z]+)([\u4e00-\u9fa5]+)|([0-9]+)([\u4e00-\u9fa5]情话段子5+)|([\u4e00-\u9fa5]+)([0-9]+)

总结

到此这篇关于利用正则表达式进行中文排版的文章就介绍到这了,更多相关正则表达式中文排版内容请搜索www.88751.com以前的文章或继续浏览下面的相关文章希望大家以后多多支持www.887551.com !

更多作文请访问 https://www.wtabcd.cn/fanwen/list/92_0.html

文章生成doc功能，由[范文网](#)开发