# 正则表达式

 正则表达式（regular expression）就是用一个“字符串”来描述一个特征，然后去验证另一个“字符串”是否符合这个特征。

# **常见匹配规则**

|  |  |
| --- | --- |
| 规则 | 说明 |
| . | 任意字符 |
| \d | 数字(0-9) |
| \D | 除数字外的其它字符 |
| \w | 字符word(a-z, A-Z, 0-9, \_) |
| \W | 非字符外的其它字符 |
| \s | white space(空格, tab, 换行符) |
| \S | 除white space以为的字符 |
|  |  |
| \b | 边界,如\bHa\b表示Ha两边必须为空格或者回车才可匹配中 |
| \B | 非边界 |
| ^ | 字符或字符串开始 |
| $ | 字符或字符串结束 |
|  |  |
| [] | 数组，如[89]表示匹配到8或者9，[a-z]表示匹配单个字母 |
| [^ ] | 匹配不在数组中的数字 |
| | | Or |
| () | 组，如(com|edu)表示匹配字符com或者edu |
|  |  |
| \* | 0 or More |
| ? | 0 or One |
| + | 1 or More |
| {3} | 连续匹配3个，如\d{3}表示匹配连续的三个数字 |
| {3, 5} | 连续匹配数字个数，最小3，最大5 |

[]中用于匹配多个**单字符**的集合，()用于匹配多个**单或多字符**的集合。

# **常用匹配**

|  |  |  |
| --- | --- | --- |
| 分类 | 规则 | 举例 |
| 网址 | https?://(www\.)?(\w+)(\.\w+) | [https://www.google.com](Https://www.google.com) |
| 邮箱 | [a-zA-Z0-9\_.+-]+@[a-zA-Z0-9-]+\.[a-zA-Z0-9-.]+ | [menghui8592@yeah.net](mailto:Menghui8592@yeah.net) |
| HTML tags | <[^>]+> |  |