# W2正则表达式深入

计算机语言：抓素掌握语法——思考要解决什么问题——尝试 边写边用边精细化。一开始不要纠结细节

## 查找替换1

普通检索word，通配符word

➀→[1]：^f [^&]

## 查找替换2

* 具体字词 docx
* 符号 .\\*

## 模式与匹配

### 普通匹配

abcdefg中是否有bcd 逐个比对

### 复杂1

孔子|孟子|[庄莊]子

[]的优先级高于|

### 复杂2

《.+》

. 任意字符

+ 依次或多次，越多越好

### 复杂3

《[^》]》+

[^》] 排除型符号，[^]

## 元字符

\^&.|?\*+()[]

字符：?\*+

### 量词

一般是贪婪的，尽量多

#### .

匹配任意字符，一般不包括换行符/n，不建议匹配换行符

梁.王

#### []

或

ab：ab整体

[ab]:a或b

[a-z]

#### [^]

排除型字符组，要求原处有一个什么符号但非所排除的那个符号，没有没有符号

#### \d

digital阿拉伯数字

汉字？——

### 数词

#### ？

0或1个

孟子对?曰

#### +

1或多个

#### \*

0或多个

+不是必要的，+和\*有等价的写法

a+等价于a|a\*

### 行首尾位置标记

#### ^

行开始，匹配后面的符号

#### $

行结束，匹配前面的符号

匹配空行：^$——可用于删除空行

^《：$

### 选择分支

\1 \2





### 综合应用

修改引文格式

朱熹《朱子》，中华书局，1983年，第1页。

→朱熹：《朱子》，中华书局1983年版，1页。

${3:a:b}

匹配到3，就替换成a；匹配不到就替换成b

支持的引擎有限，且复杂

### tip

能用字符组就少用.号

数量上知道非0次就少用\*

能用位置加以标记就尽量用^$

## 转义字符\

### \+元字符=非元字符

\\* \. \[\] \\

### \+某些非元字符=元字符

\1 \z \Z \Q \E

### \+某些非元字符=控制字符

\t(tab)，\n(换行符)

mac系统换行符：\n

win系统换行符：\r\n

### \+某些非元字符=该字符本身（没有意义的\）

\y=y

## 编程语言中的字符转义

“\n \t \\”

re.match(“\\”,”\\”)❌

\\ 会先转义成一个\，

修改：

re.match(“\\\\”,”\\\\”) #\\\\转义后等价于\\

re.match(r“\\”,”\\”) #r 不转义