



北京邮电大学

Beijing University of Posts and Telecommunications

组合与锚点

蒋砚军 北京邮电大学计算机学院



► 单字符正则表达式的组合(1)



■ 串结

◆ 如abc, [A-Z].[0-9].

■ 星号 (*)

◆ 单字符正则表达式后跟*, 匹配此单字符正则表达式的**0次或任意多次**出现

■ 例:正则表达式12*4

◆ 与字符串1234不匹配, 与1224, 12224, 14匹配

■ 例: 正则表达式[A-Z][0-9]*

此例中*作用的单字符正则表式为[0-9], 代表

[A-Z]

[A-Z][0-9]

[A-Z][0-9][0-9]

[A-Z][0-9][0-9][0-9], 等等

2

与A,A1,C45,D768匹配, 与b64512,T56t不匹配

► 单字符正则表达式的组合(2)



■ 例：正则表达式 `[Cc]hapter *[1-4]`

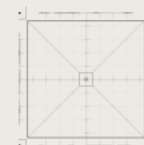
◆ 在*号前有一个空格，允许数字1-4之前有多个或者0个空格。可匹配 Chapter2, chapter 3等等。

■ 例：正则表达式 `a\[i] *= *b\[j] *\[k]`

◆ 匹配字符串 `a[i]=b[j]*c[k]`，容许等号和星号两侧有空格

■ 例：在vi中使用

`:1,$s/[0-9]*/xx/g`

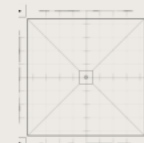
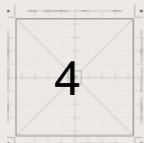


■ \$ 在尾部时有特殊意义，否则与其自身匹配

- ◆ 例：123\$ 匹配文件中行尾的123，不在行尾的123字符不匹配
- ◆ 例：\$123与字符串\$123匹配
- ◆ 例：.\$ 匹配行尾的任意字符

■ ^ 在首部时有特殊意义，否则与其自身匹配

- ◆ 例：^printf匹配行首的printf字符串，不在行首的printf串不匹配
- ◆ 例：Hel^lo与字符串Hel^lo匹配
- ◆ 例：在vi中使用 :10,50s/^ //g
删除10-50行的每行行首的4个空格



► 正则表达式扩展

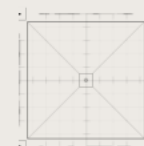


ERE: 扩展的正则表达式(ERE)

PCRE: Perl-compatible regular expression

■ 对基本正则表达式 (BRE) 进行了改进:

- ◆ 表示**分组**: 圆括号()
- ◆ 表示逻辑运算: 表示逻辑“**或**”的符号 |
(xy)* 可匹配空字符串, xy, xyxy, xyxyxy
(pink|green) 与pink或green匹配
- ◆ **重复次数**定义: 与星号地位类似的+和?, 限定重复次数 $\{m,n\}$
 - *号表示它左边的单字符正则表达式的0次或多次重复
 - +号表示1次或多次: $[0-9]^+$ 匹配长度至少为1数字串
 - ?表示0次或一次: a? 匹配零个或一个a
 - 限定重复次数 $\{m,n\}$, 例如: $[1-9][0-9]\{6,8\}$ 7-9位数字, 首位非0



► 正则表达式扩展



◆ 命名的预定义集合

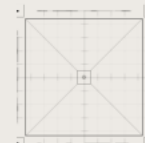
`[[:xdigit:]]` 十六进制数字

`\d` 数字

`\D` 非数字

◆ 比`^`和`$`更灵活的锚点定义

例如：寻找一个数字串，但是要求这个数字串不许出现在“合计”两个字之后





北京邮电大学

Beijing University of Posts and Telecommunications



谢谢