



北京邮电大学

Beijing University of Posts and Telecommunications



元字符和集合

蒋砚军 北京邮电大学计算机学院

► 正则表达式的概念



■ 正则表达式Regular Expressions应用范围

- ◆ 字符串匹配操作和替换操作
- ◆ 举例：Linux中的vi more grep yacc lex awk sed
- ◆ 其他：Visual Studio，Word等文本编辑器

■ 正则表达式的功能

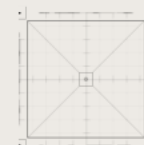
- ◆ 描述一个字符串模式

■ 注意

- ◆ 正则表达式规则与文件名通配符规则不同

- 正则表达式规则用于文本处理的场合
- 文件名匹配规则用于文件处理的场合

- ◆ 不同软件对正则表达式的定义会有差异



► 正则表达式的特殊字符（元字符）



■ 6个元字符

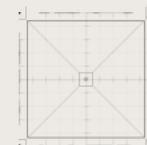
◆ . * [\ ^ \$

■ 其它字符与其自身匹配

■ 转义

◆ 用反斜线可以取消特殊字符的特殊含义。

◆ 如：正则表达end\\.只与字符串end.匹配



► 单字符正则表达式



■ 长的正则表达式由单字符正则表达式构成的

■ 非特殊字符与其自身匹配

◆ 如：正则表达式a与字符串a匹配，b与b，/与/

■ 转义字符（\）

◆ \. * \\$ ^ \[\\\

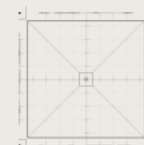
正则表达式*与字符串*匹配，与字符串*不匹配

转义字符后除以上六种之外的不该出现其他字符，例如：不该出现\u，这样的组合被视为undefined（未定义的），后出的软件有可能会有特殊的解释

■ 圆点（.）



◆ 匹配任意单字符

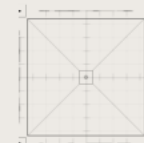


► 单字符正则表达式：定义集合(1)



■ 基本用法

- ◆ 在一对方括号之间的字符为集合的内容，
 - 如：单字符正则表达式[abcd]与a或b,c,d匹配
- ◆ 圆点,星号, 反斜线在方括号内时, 代表它们自己
 - 如:[*.]可匹配3个单字符



► 单字符正则表达式：定义集合(2)

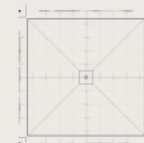


■ 用减号-定义一个区间

- ◆ 如[a-d] [A-Z] [a-zA-Z0-9]
- ◆ [][] 集合含左右中括号两个字符
- ◆ 减号在最后，则失去表示区间的意义
 - [ad-]只与3个字符匹配

■ 用^表示补集

- ◆ ^在开头,则表示与集合内字符之外的任意字符匹配
 - 如：[^a-z]匹配任一非小写字母
 - [^][]匹配任一非中括号字符
- ◆ ^不在开头,则失去其表示补集的意义
 - 如：[a-z^]能匹配27个单字符





北京邮电大学

Beijing University of Posts and Telecommunications



谢谢