今日学习心得

1. 正则表达式

正则表达式，又称正规表示式、正规表示法、正规表达式、规则表达式、常规表示法（英语：Regular Expression，在代码中常简写为 regex、regexp 或 RE），计算机科学的一个概念。正则表达式使用单个字符串来描述、匹配一系列符合某个句法规则的字符串。在很多文本编辑器里，正则表达式通常被用来检索、替换那些符合某个模式的文本。

许多程序设计语言都支持利用正则表达式进行字符串操作。例如，在 Perl 中就内建了一个功能强大的正则表达式引擎。正则表达式这个概念最初是由 UNIX 中的工具软件（例如sed和grep）普及开的。正则表达式通常缩写成“regex”，单数有 regexp、regex，复数有 regexps、regexes、regexen。

2. 基本语法

一个正则表达式通常被称为一个模式（pattern），为用来描述或者匹配一系列符合某个句法规则的字符串。

选择

|竖直分隔符表示选择，例如"boy|girl"可以匹配"boy"或者"girl"

数量限定

数量限定除了我们举例用的\*,还有+加号,?问号,如果在一个模式中不加数量限定符则表示出现一次且仅出现一次：

+表示前面的字符必须出现至少一次(1次或多次)，例如，"goo+gle",可以匹配"gooogle","goooogle"等；

?表示前面的字符最多出现一次(0次或1次)，例如，"colou?r",可以匹配"color"或者"colour";

\*星号代表前面的字符可以不出现，也可以出现一次或者多次（0次、或1次、或多次），例如，“0\*42”可以匹配42、042、0042、00042等。

范围和优先级

()圆括号可以用来定义模式字符串的范围和优先级，这可以简单的理解为是否将括号内的模式串作为一个整体。例如，"gr(a|e)y"等价于"gray|grey"，（这里体现了优先级，竖直分隔符用于选择a或者e而不是gra和ey），"(grand)?father"匹配father和grandfather（这里体验了范围，?将圆括号内容作为一个整体匹配）。

3. grep模式匹配命令

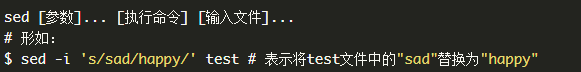
grep命令用于打印输出文本中匹配的模式串，它使用正则表达式作为模式匹配的条件。grep支持三种正则表达式引擎，分别用三个参数指定：

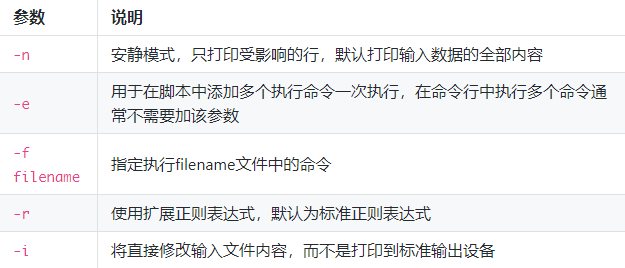


4. sed流编辑器

sed工具在 man 手册里面的全名为"sed - stream editor for filtering and transforming text "，意即，用于过滤和转换文本的流编辑器。

sed 命令基本格式：





5. awk文本处理语言

AWK是一种优良的文本处理工具，Linux及Unix环境中现有的功能最强大的数据处理引擎之一.其名称得自于它的创始人Alfred Aho（阿尔佛雷德·艾侯）、Peter Jay Weinberger（彼得·温伯格）和Brian Wilson Kernighan（布莱恩·柯林汉)姓氏的首个字母.AWK程序设计语言，三位创建者已将它正式定义为“样式扫描和处理语言”。它允许您创建简短的程序，这些程序读取输入文件、为数据排序、处理数据、对输入执行计算以及生成报表，还有无数其他的功能。最简单地说，AWK是一种用于处理文本的编程语言工具。

awk命令基本格式：



其中-F参数用于预先指定前面提到的字段分隔符（还有其他指定字段的方式） ，-v用于预先为awk程序指定变量，-f参数用于指定awk命令要执行的程序文件，或者在不加-f参数的情况下直接将程序语句放在这里，最后为awk需要处理的文本输入，且可以同时输入多个文本文件。