

Unix中选择合适的字符串处理方式Linux认证考试 PDF转换可能丢失图片或格式，建议阅读原文

https://www.100test.com/kao_ti2020/586/2021_2022_Unix_E4_B8_AD_E9_80_89_c103_586591.htm

在Unix操作系统中，说实话其字符串的处理功能实在不强。如没有现成的length(获取字符串的长度)等函数。要实现对字符串的操作，往往需要采用expr命令结合相关的通配符来实现一些特定的功能。虽然在Unix操作系统中，对于字符串的操作比较麻烦。但是系统工程师只要知道各种处理方式特点与缺点，那么对症下药，仍然可以找到一些简单的方式来处理系统中的字符串问题。

一、对用户输入的字符串长度进行验证。系统工程师有时候往往需要对用户输入的字符串长度进行验证。如在设计一个登陆程序时，为了安全起见，可能需要限制用户密码输入的最小位数，如必须是六位以上的密码等等。此时，其实就是统计用户输入的字符串长度问题。如果用户输入的密码长度不够，就提醒用户.如果够的话，就直接保存用户输入的信息。所以，简单的说这可以利用if语句或者while循环判断语句来实现。而这个条件则可以利用expr关键字来实现。如果用判断用户输入的字符串长度，可以利用如下的格式来写。Expr “\$userpasswd” : ‘.*’ 即可。在编写这个命令时，需要注意一些基本的规则。如这个expr命令往往需要使用两个表达式，而且中间需要使用英文状态小的冒号进行分割。而且这个两个表达式的顺序也有严格的限制。冒号左边是需要处理的字符串(需要统计长度的字符串或者变量)，而冒号右边则是普通的正则表达式。其次，需要处理的字符串需要使用双引号括起来。这里跟普通命令中字符串的处理方式有所不同。

在普通的命令中，变量不需要使用双引号引起来。如果引起起来的话，则系统反而会将其当作常量来处理。但是如果在if条件语句中将利用这个命令来统计字符串长度时，则需要将变量使用双引号引起来。这个字符串处理上的差异需要注意，不少系统工程师在这里吃过亏。第三，有时候正则表达式也需要利用单引号来引起来。如上面的 `'.*'` 则表示所有匹配的字符。如果不使用单引号的话，系统可能会提示语法错误。

二、截取字符串。

有时候系统管理员需要截取字符串的自串来完成特定的任务。如现在系统工程师需要对某个文件根据系统的时间来进行命名，其命名的规则为年份(最后两位)加月份的形式，即0906表示这个文件是09年6月份创建的。因为每个月创建一个独立的日志文件，为此按这个格式对其进行命名算是比较合理的。但是在系统中实现起来有一定的困难。然后通过date命令加上一些可选项参数可以截取日期的年份、月份等信息。可是利用date命令截取的年份是四位数字(如2009)，而不是两位数字。这跟系统工程师的需求还有一定的距离。那么如果一定要采用年份的两位数字来对文件或者文件夹进行命名，该如何处理呢?此时就需要用到expr命令的另一个功能，即截取字符串的功能。一般来说，expr命令主要有三个功能，分别为求字符串的长度、截取字符串、求一个字符在一个字符串中的位置。当然如果这些功能跟其他命令结合使用的话，可以实现一些比较复杂的需求。

在Unix系统中，如果需要截取字符串的话，笔者首先建议的是采用一些自带的命令。如date %w则表示显示今天星期几。虽然expr命令也可以实现类似的功能，但是初期起来比较麻烦。为此对于截取字符串笔者建议，能够使用系统自带功能

来实现的，那么就使用系统自带的命令。如果没有的话，那么就只有用过expr命令来实现。也就是说，expr是一个通用的命令，所有的字符串截取工作其都可以完成。只是代码书写起来没有那些系统自带命令那么简单。如果现在系统工程师需要截取年份的后两位数字，此时该如何处理呢？在系统转义字符的帮助下，可以完成这个功能。如下所示：`$expr`
“2009”：‘`..\(..\)`’ 在实现这个功能的时候，笔者使用了\这个转义字符。这个正则表达式的含义就表示只考虑前面这个字符串的最后两个字符，而忽略其他的字符。若需要利用上面这个规则来进行文件命令的话，只需要先利用这个命令截取年份的最后两位，然后再把月份加上即可。上面的字符串2009系统工程师也可以使用变量来代替。系统工程师需要注意的是这个正则表达式的书写方法。在使用expr命令来对字符串进行处理，其难点就在于这个正则表达式的书写。通常情况下，只要合理编写这个正则表达式，则其字符串的处理功能，并不会比数据库系统中字符串处理函数来的差，甚至有过之而不及。现在的主要问题是，由于经验有限，系统工程师往往很难根据用户需求来定义这个正则表达式。就连笔者有时候也会在这里卡住。如果系统工程师想要提高自己正则表达式的书写能力，那么多看看其他系统工程师书写的脚本，往往能够在最短时间内提高自己正则表达式的书写水平。

三、修改文件的扩展名。如现在操作系统中某个目录下面有很多的sh文件(Unix系统中的配置文件，其实跟txt相同)。现在系统工程师需要将这些sh文件的扩展名改掉，全部改为txt。虽然可以通过一个个的手工修改来实现，但是处理起来的难度比较大。如果要实现批量更改文件扩展名的话，则

可以借助for等循环语句来实现。不过循环语句内部的操作有两种实现方式。一是采用expr命令。其实对文件名进行操作，说到底就是对字符串的操作。因为文件名字也是一种字符串。如系统工程师可以先利用expr命令截取字符串的功能，将扩展名前面的字符串截取出来，然后将其保存在一个变量中，如filenameshort。然后再利用系统的重命名或者移动文件的命令，对系统进行重命名，如\$filenameshortdoc。这就可以实现更改扩展名的需求。不过笔者并不建议采用这个expr命令来完成这个需求。因为当文件名的长度不固定时，这个正则表达式比价难写。要将文件名字符串截取出来(去掉扩展名的名字)，有一定的难度。在这里，笔者介绍另外一个命令basename命令。其实这这也是一个字符串相关的处理命令。其基本格式为basename text1 text2。他的基本功能就是从第一个字符串中去掉第二个字符串的内容。如执行命令basename setup.sh sh，则表示将setup.sh这个字符串中的sh字符串去掉，变为setup。那么此就可以利用这个命令来进行成批修改扩展名。其具体的实现方式跟上面类似。只是在截取文件名字符串的时候，采用了basename命令，而不是expr命令。而basename命令本身就带有截取字符串的功能，不需要采用正则表达式。所以实现起来相对简单一点。从这个案例中，笔者也想再次强调一下一个字符串处理的基本原则。即无论是截取字符串还是对字符串进行的其他操作，如果可以通过系统命令 可选项参数来实现的，就采用这种实现方式。如果没有的话，需要考虑能否采用basename等不需要正则表达式的命令来实现。只有在以上两种方式都不可行的情况下，才考虑使用exper命令来实现。这主要是因为正则表达式的设计

实在比较麻烦，虽然其功能比较强大。利用exper来统计一下字符串的长度还好，但是如果要实现一些复杂功能，如截取字符串或者确定字符的位置，就需要系统工程师有比较深厚的功底了。更多优质资料尽在百考试题论坛 百考试题在线题库 linux认证更多详细资料 100Test 下载频道开通，各类考试题目直接下载。详细请访问 www.100test.com