

文本处理命令

内容提要

1. 掌握常用的文本显示命令
2. 掌握常用的文本处理命令

cat

- 功能说明：用于从文件头到文件尾方向滚屏显示文本文件内容
- 命令格式：cat [参数] [<文件> ...]
- 常用参数：
 - **-n**：由 1 开始对所有输出的行进行编号
 - **-b**：和 -n 相似，只不过对于空行不编号
 - **-s**：当遇到有连续两行以上的空行时，使用一个空行代替
- 使用举例：

```
$ cat file
$ cat -n file
```

1. 系统还提供了一个 **tac** 命令，用于从文件尾到文件头显示文件内容。
2. 系统还提供了一个 **rev** 命令，与 **tac** 不同，它并不反转行序，而是把每行的内容反转。
3. 可以使用 **cat** 命令连接多个文本文件，如：

```
$ cat file1 file2 > files
```

more

- 功能说明：从文件头到文件尾分屏显示文本文件内容
- 命令格式：more [参数] [<文件> ...]
- 常用参数：
 - **-d**：显示提示信息“[Press space to continue, 'q' to quit.]”
 - **-s**：当遇到有连续两行以上的空行时，使用一个空行代替
 - **+num**：从第 num 行开始显示
- 使用举例：

```
$ more file
$ more +10 file
```

1. 系统还提供了一个 **less** 命令，用于双向显示分屏显示文本文件内容。
2. **less** 的功能比 **more** 丰富的多，可以使用如下命令查看其内置功能

```
$ less --help
```

head

- 功能说明：显示文本文件的前部的若干行
- 命令格式：head [参数] [<文件> ...]
- 常用参数：
 - **-n**：显示前 n 行，不指定此参数显示前 10 行
- 使用举例：

```
$ head file
$ head -5 file
```

tail

- 功能说明：显示文本文件的后部的若干行
- 命令格式：tail [参数] [<文件> ...]
- 常用参数：
 - **-n**：显示后 n 行，不指定此参数显示后 10 行
 - **+n**：从第 n 行显示到文件尾
 - **-F**：用于跟踪显示不断增长的文件结尾内容（通常用于显示日志文件）
- 使用举例：

```
$ tail file
$ tail -5 file
$ tail +5 file
$ tail -F /var/log/messages
```

cut

- 功能说明：纵向切割出文本指定的部分并写到标准输出
- 命令格式：cut [参数] [<文件> ...]
- 常用参数：
 - **-b<LIST>**：只列出<LIST>指定的字节
 - **-c<LIST>**：只列出<LIST>指定的字符
 - **-f<LIST>**：只列出<LIST>指定的字段；并打印所有不包含分界符的行，除非 **-s** 选项被指定
 - **-s**：不打印没有包含分界符的行
 - **-d<DELIM>**：DELIM 是分界符，使用指定<DELIM>代替制表符作为区域分界
 - **--complement**：补足选中的字节、字符或字段的占位
 - **--output-delimiter=<STRING>**：使用指定<STRING>作为输出分界符默认时采用输入的分界符

LIST 的语法

选用 **-b**、**-c** 或 **-f** 中一个或若干个选项时，每个<LIST>都由一个范围域或是由逗号分隔开的多个范围域组成。被选中的输入会被按照和读入时相同的次序写到屏幕，每个输入只会被输出一次。每个范围域可以是以下中的任何一种：

- **N** : 第N个字节、字符或字段, 从 1 开始计数
- **N-** : 从第N个字节、字符或字段, 直到行尾
- **N-M** : 从第N个到第M个已包含的字符、字符或字段
- **-M** : 从第一个到第M个字节、字符或字段
- 使用举例:

```
$ cut -b-10 file
$ cut -c5- file
$ cut -c5-10,15-20 file
$ cut -f1,3,5 file
$ cut -f2-4 file
$ cut -f2-4 -d' ' file
$ cut -f1,2-4,6 -d' ' -s file
```

paste

- 功能说明: 纵向合并多个文本并写到标准输出
- 命令格式: `paste [参数] [<文件> ...]`
- 常用参数:
 - **-d<DELIM>** : DELIM是分界符, 使用指定<DELIM>代替制表符作为区域分界
 - **-s** : 不使用平行的行目输出模式, 而是每个文件占用一行
- 使用举例:

```
$ paste file1 file2
$ paste -s file1 file2
$ paste -d' ' file1 file2
```

sort

- 功能说明: 以行为单位对文件进行排序
- 命令格式: `sort [参数] [<文件> ...]`
- 常用参数:
 - **-b** : 忽略前导的空格
 - **-d** : 只考虑空格, 字母, 数字
 - **-f** : 忽略字母的大小写
 - **-i** : 只考虑可打印字符
 - **-M** : 排序月份, (未知词)<"JAN"< ... <"DEC"
 - **-n** : 根据字符串的数值进行排序
 - **-r** : 逆向排序
 - **-u** : 对相同的行只输出一行
 - **+n** : n 为数字, 对指定的列进行排序, +0 表示第1列, 以空格或制表符作为列的间隔符
- 使用举例:

```
$ sort file
$ sort -bd file
```

```
$ sort -bn file
$ sort -r file
$ sort -u file
$ sort +5 file
$ sort +5 -rb file
$ sort file1 file2
$ sort -br file1 file2
```

1. 本地环境变量会影响排序结果。
2. 如果希望以字节的自然值获得最传统的排序结果，请设定 `LC_ALL=C`。

uniq

- 功能说明：删除文本文件中相邻的重复的行并写到标准输出
- 命令格式：`uniq [参数] [<输入文件> [<输出文件>]]`
- 常用参数：
 - **-c**：在每行前加上表示相应行目出现次数的前缀编号
 - **-d**：只显示重复的行
 - **-i**：略大小写差异
 - **-u**：只显示出现一次的行
 - **-s<N>**：<N> 为数字，对各行前<N>个字符不作比较
 - **-w<N>**：<N> 为数字，对各行第<N>个字符以后的内容不作比较
- 使用举例：

```
$ uniq file
$ uniq -i file
$ uniq -cd file
$ uniq -u file
```

WC

- 功能说明：统计指定文本文件的行数、字数、字符数
- 命令格式：`wc [参数] [<文件> ...]`
- 常用参数：
 - **-c**：输出字节数统计
 - **-l**：输出行数统计
 - **-L**：输出最长一行的长度
 - **-w**：输出单词数统计
- 使用举例：

```
$ wc file
$ wc -l file
$ wc -w file
$ wc -c file
$ wc -L file
```

expand

- 功能说明：将文件中的制表符转换为空格，写到标准输出
- 命令格式：expand [参数] [<文件> ...]
- 常用参数：
 - **-i**：不转换非空格后的制表符
 - **-t<N>**：设定每个制表符为指定<N>的宽度，而不是默认的 8
- 使用举例：

```
$ expand file
$ expand -t4 file
```

unexpand

- 功能说明：将文件中的空格转换为制表符，写到标准输出
- 命令格式：unexpand [参数] [<文件> ...]
- 常用参数：
 - **-a**：转换所有空格字符而不仅仅是字母首部的空格
 - **--first-only**：只转换首部的空格字符序列（覆盖 **-a** 选项）
 - **-t<N>**：设定每个制表符为指定<N>的宽度，而不是默认的 8（激活 **-a** 选项）
- 使用举例：

```
$ unexpand file
$ unexpand -t4 file
```

iconv

- 功能说明：将文件从一种编码转换成另一种编码
- 命令格式：iconv [参数] <输入文件>
- 常用参数：
 - **-f <encoding>**：指定原始文本编码
 - **-t <encoding>**：指定要转换的编码
 - **-l**：列出所有已知编码字符集
 - **-c**：忽略输出中的无效字符
 - **-o <output file>**：指定输出文件，而不是在标准输出上显示
- 使用举例：

```
$ iconv -l
$ iconv -f ISO-8859-1 -t UTF-8 -o outputfile inputfile
$ iconv -f GB2312 -t UTF-8 -o outputfile inputfile
$ iconv -f GBK -t UTF-8 -o outputfile inputfile
$ iconv -f BIG5 -t UTF-8 -o outputfile inputfile
$ iconv -f UTF-8 -t GB2312 -o outputfile inputfile
```

dos2unix

- 功能说明：将**DOS**格式的文本文件转换成**UNIX**格式的文本文件
- 命令格式：dos2unix [参数] <文件> [<输出文件>]
- 常用参数：
 - **-k**：不改变文件的时间戳
 - **-n**：新文件模式，即不改变原文件将转换结果保存到指定的输出文件
- 使用举例：

```
$ dos2unix dosfile
$ dos2unix -n dosfile linuxfile
$ dos2unix -k *
$ dos2unix -k -n dosfile linuxfile
```

1. 系统还提供了一个 **unix2dos** 命令，用于将**UNIX**格式的文本文件转换成**DOS**格式的文本文件。使用方法与 **dos2unix** 命令一样。

- 显示源文件
- 登录