

Linux 上 12 个高效的文本过滤命令

2017-2-17 09:00

编译自：<http://www.tecmint.com/linux-file-operations-commands/>
原创：LCTT <https://linux.cn/article-8216-1.html>

作者：Aaron Kili
译者：yangmingming



在这篇文章中，我们将会看一些 Linux 中的过滤器命令行工具。过滤器是一个程序，它从标准输入读取数据，在数据上执行操作，然后把结果写到标准输出。

因此，它可以用来以强大的方式处理信息，例如重新结构化输出以生成有用的报告，修改文件里面的文本，以及其他很多系统管理任务。



下面是 Linux 上的一些有用的文件或者文本过滤器。

1、awk 命令

awk ^[1] 是一个卓越的模式扫描和处理语言，它可被用于在 Linux 下构造有用的过滤器。你可以通过阅读我们的 **awk 系列 1 到 13 部分** ^[2] 来开始使用它。

另外，也可以通过阅读 **awk** 的 man 手册来获取更多的信息和使用选项。

```
$ man awk
```

2、 sed 命令

sed ^[3] 是一款过滤和转换文本的强大的流编辑器。我们已经写了两篇关于 sed 的有用的文章，你可以通过这儿来了解：

- [如何使用 GNU sed 命令在 Linux 下创建、编辑和处理文件](#) ^[4]
- [日常 Linux 系统管理员任务使用的 15 个有用的 sed 命令小贴士和技巧](#) ^[5]

sed 的 man 手册已经添加控制选项和说明：

```
$ man sed
```

3、 grep、 egrep、 fgrep、 rgrep 命令行

这些过滤器输出匹配指定模式的行。它们从一个文件或者标准输入读取行，并且输出所有匹配的行，默认输出到标准输出。

注意：主程序是 **grep** ^[6]，这些变体与 [使用特定的选项的 grep](#) ^[7] 相同，如下所示（为了向后兼容性，它们依旧在使用）：

```
$ egrep = grep -E
$ fgrep = grep -F
$ rgrep = grep -r
```

下面是一些基本的 grep 命令：

```
tecmin@TecMint ~ $ grep "aaronkilik" /etc/passwd
aaronkilik:x:1001:1001::/home/aaronkilik:
tecmin@TecMint ~ $ cat /etc/passwd | grep "aronkilik"
aaronkilik:x:1001:1001::/home/aaronkilik:
```

在 [Linux 下的 grep、egrep 和 fgrep 的差异？^{\[8\]}](#) 中，你可以了解更多。

4、 head 命令

head 用于显示文件前面的部分，默认情况下它输出前 **10** 行。你可以使用 `-n` 标志来指定显示的行数：

```
tecmin@TecMint ~ $ head /var/log/auth.log

Jan  2 10:45:01 TecMint CRON[3383]: pam_unix(cron:session): session opened for user root by
(uid=0)

Jan  2 10:45:01 TecMint CRON[3383]: pam_unix(cron:session): session closed for user root

Jan  2 10:51:34 TecMint sudo:  tecmint : TTY=unknown ; PWD=/home/tecmin ; USER=root ;
COMMAND=/usr/lib/linuxmint/mintUpdate/checkAPT.py

Jan  2 10:51:34 TecMint sudo: pam_unix(sudo:session): session opened for user root by (uid=0)

Jan  2 10:51:39 TecMint sudo: pam_unix(sudo:session): session closed for user root

Jan  2 10:55:01 TecMint CRON[4099]: pam_unix(cron:session): session opened for user root by
(uid=0)

Jan  2 10:55:01 TecMint CRON[4099]: pam_unix(cron:session): session closed for user root

Jan  2 11:05:01 TecMint CRON[4138]: pam_unix(cron:session): session opened for user root by
(uid=0)

Jan  2 11:05:01 TecMint CRON[4138]: pam_unix(cron:session): session closed for user root

Jan  2 11:09:01 TecMint CRON[4146]: pam_unix(cron:session): session opened for user root by
(uid=0)

tecmin@TecMint ~ $ head -n 5 /var/log/auth.log

Jan  2 10:45:01 TecMint CRON[3383]: pam_unix(cron:session): session opened for user root by
(uid=0)

Jan  2 10:45:01 TecMint CRON[3383]: pam_unix(cron:session): session closed for user root

Jan  2 10:51:34 TecMint sudo:  tecmint : TTY=unknown ; PWD=/home/tecmin ; USER=root ;
COMMAND=/usr/lib/linuxmint/mintUpdate/checkAPT.py

Jan  2 10:51:34 TecMint sudo: pam_unix(sudo:session): session opened for user root by (uid=0)

Jan  2 10:51:39 TecMint sudo: pam_unix(sudo:session): session closed for user root
```

学习如何 [使用带有 tail 和 cat 命令的 head 命令^{\[9\]}](#)，以便在 Linux 下更有效的使用。

5、tail 命令

tail 输出一个文件的后面的部分（默认 10 行）。使用 `-n` 选项来指定显示的行数。

下面的命令将会输出指定文件的最后 5 行：

```
tecmin@TecMint ~ $ tail -n 5 /var/log/auth.log

Jan  6 13:01:27 TecMint sshd[1269]: Server listening on 0.0.0.0 port 22.

Jan  6 13:01:27 TecMint sshd[1269]: Server listening on :: port 22.

Jan  6 13:01:27 TecMint sshd[1269]: Received SIGHUP; restarting.

Jan  6 13:01:27 TecMint sshd[1269]: Server listening on 0.0.0.0 port 22.

Jan  6 13:01:27 TecMint sshd[1269]: Server listening on :: port 22.
```

另外，**tail** 有一个特殊的选项 `-f`，可以 [实时查看一个文件的变化](#) ^[10]（尤其是日志文件）。

下面的命令将会使你能够监控指定文件的变化：

```
tecmin@TecMint ~ $ tail -f /var/log/auth.log

Jan  6 12:58:01 TecMint sshd[1269]: Server listening on :: port 22.

Jan  6 12:58:11 TecMint sshd[1269]: Received SIGHUP; restarting.

Jan  6 12:58:12 TecMint sshd[1269]: Server listening on 0.0.0.0 port 22.

Jan  6 12:58:12 TecMint sshd[1269]: Server listening on :: port 22.

Jan  6 13:01:27 TecMint sshd[1269]: Received SIGHUP; restarting.

Jan  6 13:01:27 TecMint sshd[1269]: Server listening on 0.0.0.0 port 22.

Jan  6 13:01:27 TecMint sshd[1269]: Server listening on :: port 22.

Jan  6 13:01:27 TecMint sshd[1269]: Received SIGHUP; restarting.

Jan  6 13:01:27 TecMint sshd[1269]: Server listening on 0.0.0.0 port 22.

Jan  6 13:01:27 TecMint sshd[1269]: Server listening on :: port 22.
```

阅读 tail 的 man 手册，获取使用选项和说明的完整内容：

```
$ man tail
```

6、 sort 命令

sort ^[11] 用于将文本文件或标准输入的行进行排序。

下面是一个名为 domain.list 的文件的内容：

```
tecmin@TecMint ~ $ cat domains.list  
  
tecmin.com  
  
tecmin.com  
  
news.tecmin.com  
  
news.tecmin.com  
  
linuxsay.com  
  
linuxsay.com  
  
windowsmint.com  
  
windowsmint.com
```

你可以像这样运行一个简单的 **sort 命令** ^[12] 来排序文件内容：

```
tecmin@TecMint ~ $ sort domains.list  
  
linuxsay.com  
  
linuxsay.com  
  
news.tecmin.com  
  
news.tecmin.com  
  
tecmin.com  
  
tecmin.com  
  
windowsmint.com  
  
windowsmint.com
```

你可以有多种方式来使用 sort 命令，请参阅以下一些关于 sort 命令的有用的文章。

- [Linux 的 ‘sort’ 命令的14个有用的范例（一）](#) ^[13]
- [Linux 的 'sort'命令的七个有趣实例（二）](#) ^[14]
- [如何基于修改日期和时间来查找和排序文件](#) ^[15]

7、uniq 命令

uniq ^[16] 命令用于报告或者忽略重复行，它从标准输入过滤行，并且把结果写到标准输出。

在对一个输入流运行 `sort` 之后，你可以使用 `uniq` 删除重复行，如下例所示。

为了显示行出现的数目，使用 `-c` 选项，要在对比时忽略大小写的差异，使用 `-i` 选项：

```
tecmin@TecMint ~ $ cat domains.list
tecmin.com
tecmin.com
news.tecmin.com
news.tecmin.com
linuxsay.com
linuxsay.com
windowsmint.com

tecmin@TecMint ~ $ sort domains.list | uniq -c
2 linuxsay.com
2 news.tecmin.com
2 tecmin.com
1 windowsmint.com
```

通过阅读 `uniq` 的 man 手册来获取进一步的使用信息和选项：

```
$ man uniq
```

8、fmt 命令行

fmt 是一款简单的优化的文本格式化器，它重新格式化指定文件的段落，并且打印结果到标准输出。

以下是从文件 `domain-list.txt` 提取的内容：

```
1.tecmin.com 2.news.tecmin.com 3.linuxsay.com 4.windowsmint.com
```

为了把上面的内容重新格式化成一个标准的清单，运行下面的命令，使用 `-w` 选项定义最大行宽度：

```
tecmin@TecMint ~ $ cat domain-list.txt
1.tecmint.com 2.news.tecmint.com 3.linuxsay.com 4.windowsmint.com
tecmin@TecMint ~ $ fmt -w 1 domain-list.txt
1.tecmint.com
2.news.tecmint.com
3.linuxsay.com
4.windowsmint.com
```

9、 pr 命令

`pr` 命令转换文本文件或者标准输入之后打印出来。例如在 **Debian** 系统上，你可以像下面这样显示所有的安装包：

```
$ dpkg -l
```

为了将要打印的列表在页面和列中组织好，使用以下命令。

```
tecmin@TecMint ~ $ dpkg -l | pr --columns 3 -l 20

2017-01-06 13:19                                         Page 1

Desired=Unknown/Install ii  adduser                  ii  apg
| Status=Not/Inst/Conf- ii  adwaita-icon-theme ii  app-install-data
|/ Err?=(none)/Reinst-r ii  adwaita-icon-theme- ii  apparmor
|// Name                ii  alsa-base                  ii  apt
+++===== ii  alsa-utils                  ii  apt-clone
ii  accountsservice      ii  anacron                    ii  apt-transport-https
ii  acl                  ii  apache2                   ii  apt-utils
ii  acpi-support         ii  apache2-bin                ii  apt-xapian-index
ii  acpid                ii  apache2-data               ii  aptdaemon
ii  add-apt-key           ii  apache2-utils              ii  aptdaemon-data

2017-01-06 13:19                                         Page 2
```

```
ii aptitude          ii avahi-daemon      ii bind9-host
ii aptitude-common   ii avahi-utils        ii binfmt-support
ii apturl            ii aview              ii binutils
ii apturl-common     ii banshee            ii bison
ii archdetect-deb    ii baobab             ii blt
ii aspell            ii base-files         ii blueberry
ii aspell-en         ii base-passwd        ii bluetooth
ii at-spi2-core      ii bash               ii bluez
ii attr             ii bash-completion   ii bluez-cups
ii avahi-autoipd     ii bc                 ii bluez-obexd
.....
```

其中，使用的标志如下：

- `--column` 定义在输出中创建的列数。
- `-l` 指定页面的长度（默认是 66 行）。

10、tr 命令行

这个命令从标准输入转换或者删除字符，然后输出结果到标准输出。

使用 `tr` 的语法如下：

```
$ tr options set1 set2
```

看一下下面的例子，在第一个命令，`set1([:upper:])` 代表指定输入字符的大小写（都是大写字符）。`set2([:lower:])` 代表期望结果字符的大小写。第二个例子意思相似，转义字符 `\n` 表示在新的一行打印输出：

```
tecmint@TecMint ~ $ echo "WWW.TECMINT.COM" | tr [:upper:] [:lower:]
www.tecmint.com

tecmint@TecMint ~ $ echo "news.tecmint.com" | tr [:lower:] [:upper:]
NEWS.TECMINT.COM
```


11、 more 命令

more ^[17] 命令是一个有用的文件过滤器，最初为查看证书而建。它一页页显示文件内容，用户可以通过按回车来显示更多的信息。

你可以像这样使用它来显示大文件：

```
tecmin@TecMint ~ $ dmesg | more

[ 0.000000] Initializing cgroup subsys cpuset
[ 0.000000] Initializing cgroup subsys cpu
[ 0.000000] Initializing cgroup subsys cpuacct
[ 0.000000] Linux version 4.4.0-21-generic (build@lgw01-21) (gcc version 5.3.1 20160413
(Ubuntu 5.3.1-14ubuntu2) ) #37-Ubuntu SMP Mon Apr 18 18:33:37 UTC 2016 (Ubuntu 4.4.0-21.37-
generic
4.4.6)
[ 0.000000] Command line: BOOT_IMAGE=/boot/vmlinuz-4.4.0-21-generic root=UUID=bb29dda3-
bdaa-4b39-86cf-4a6dc9634a1b ro quiet splash vt.handoff=7
[ 0.000000] KERNEL supported cpus:
[ 0.000000]   Intel GenuineIntel
[ 0.000000]   AMD AuthenticAMD
[ 0.000000]   Centaur CentaurHauls
[ 0.000000] x86/fpu: xstate_offset[2]: 576, xstate_sizes[2]: 256
[ 0.000000] x86/fpu: Supporting XSAVE feature 0x01: 'x87 floating point registers'
[ 0.000000] x86/fpu: Supporting XSAVE feature 0x02: 'SSE registers'
[ 0.000000] x86/fpu: Supporting XSAVE feature 0x04: 'AVX registers'
[ 0.000000] x86/fpu: Enabled xstate features 0x7, context size is 832 bytes, using
'standard' format.
[ 0.000000] x86/fpu: Using 'eager' FPU context switches.
[ 0.000000] e820: BIOS-provided physical RAM map:
[ 0.000000] BIOS-e820: [mem 0x0000000000000000-0x00000000000009d3ff] usable
[ 0.000000] BIOS-e820: [mem 0x00000000000009d400-0x00000000000009ffff] reserved
[ 0.000000] BIOS-e820: [mem 0x000000000000e0000-0x000000000000ffffff] reserved
[ 0.000000] BIOS-e820: [mem 0x00000000000100000-0x00000000000a56affff] usable
```

```
[ 0.000000] BIOS-e820: [mem 0x00000000a56b0000-0x00000000a5eaffff] reserved
[ 0.000000] BIOS-e820: [mem 0x00000000a5eb0000-0x00000000aaabefff] usable
--More--
```

12、less 命令

less 是和上面的 **more** 命令相反的一个命令，但是它提供了额外的特性，而且对于大文件，它会更快些。

按照 **more** 命令相同的方式使用它：

```
tecmint@TecMint ~ $ dmesg | less

[ 0.000000] Initializing cgroup subsys cpuset
[ 0.000000] Initializing cgroup subsys cpu
[ 0.000000] Initializing cgroup subsys cpuacct
[ 0.000000] Linux version 4.4.0-21-generic (buildd@lgw01-21) (gcc version 5.3.1 20160413
(Ubuntu 5.3.1-14ubuntu2) ) #37-Ubuntu SMP Mon Apr 18 18:33:37 UTC 2016 (Ubuntu 4.4.0-21.37-
generic
4.4.6)
[ 0.000000] Command line: BOOT_IMAGE=/boot/vmlinuz-4.4.0-21-generic root=UUID=bb29dda3-
bdaa-4b39-86cf-4a6dc9634a1b ro quiet splash vt.handoff=7
[ 0.000000] KERNEL supported cpus:
[ 0.000000]   Intel GenuineIntel
[ 0.000000]   AMD AuthenticAMD
[ 0.000000]   Centaur CentaurHauls
[ 0.000000] x86/fpu: xstate_offset[2]: 576, xstate_sizes[2]: 256
[ 0.000000] x86/fpu: Supporting XSAVE feature 0x01: 'x87 floating point registers'
[ 0.000000] x86/fpu: Supporting XSAVE feature 0x02: 'SSE registers'
[ 0.000000] x86/fpu: Supporting XSAVE feature 0x04: 'AVX registers'
[ 0.000000] x86/fpu: Enabled xstate features 0x7, context size is 832 bytes, using
'standard' format.
[ 0.000000] x86/fpu: Using 'eager' FPU context switches.
[ 0.000000] e820: BIOS-provided physical RAM map:
[ 0.000000] BIOS-e820: [mem 0x0000000000000000-0x00000000000009d3ff] usable
```

```
[ 0.000000] BIOS-e820: [mem 0x0000000000009d400-0x0000000000009ffff] reserved
[ 0.000000] BIOS-e820: [mem 0x000000000000e0000-0x000000000000ffffff] reserved
[ 0.000000] BIOS-e820: [mem 0x00000000000100000-0x000000000a56affff] usable
[ 0.000000] BIOS-e820: [mem 0x000000000a56b0000-0x000000000a5eaffff] reserved
[ 0.000000] BIOS-e820: [mem 0x000000000a5eb0000-0x000000000aaabefff] usable
:
```

学习为什么 Linux 下进行有效的文件浏览， [‘less’ 比 ‘more’ 命令更快](#) ^[18]。

基本上就这些了，如果你还知道其他本文没有提供的 Linux 下 [有用的文本过滤命令行工具](#) ^[19]，可以在下面的评论部分通知我们。

作者简介：Aaron Kili 是一名 Linux 和 F.O.S.S 爱好者、一名未来的 Linux 系统管理员、web 开发者，并且目前是一名 TecMint 上的内容创造者，他喜欢计算机相关的工作，并且坚信知识的分享。

via: <http://www.tecmint.com/linux-file-operations-commands/>

作者：[Aaron Kili](#) ^[20] 译者：[yangmingming](#) ^[21] 校对：[jasminepeng](#) ^[22]

本文由 [LCTT](#) ^[23] 原创编译，[Linux中国](#) ^[24] 荣誉推出

[1]: <https://linux.cn/tag-awk.html>

[2]: <https://linux.cn/article-7586-1.html>

[3]: <https://linux.cn/tag-sed.html>

[4]: <https://linux.cn/article-7161-1.html>

[5]: <http://www.tecmint.com/linux-sed-command-tips-tricks/>

[6]: <https://linux.cn/tag-grep.html>

[7]: <http://www.tecmint.com/linux-grep-commands-character-classes-bracket-expressions/>

[8]: <http://www.tecmint.com/difference-between-grep-egrep-and-fgrep-in-linux/>

[9]: <http://www.tecmint.com/view-contents-of-file-in-linux/>

[10]: <http://www.tecmint.com/fswatch-monitors-files-and-directory-changes-modifications-in-linux/>

[11]: <https://linux.cn/tag-sort.html>

[12]: <https://linux.cn/tag-sort.html>

[13]: <https://linux.cn/article-5372-1.html>

[14]: <https://linux.cn/article-5373-1.html>

[15]: <http://www.tecmint.com/find-and-sort-files-modification-date-and-time-in-linux/>

[16]: <https://linux.cn/tag-uniq.html>

[17]: <https://linux.cn/tag-more.html>

[18]: <http://www.tecmint.com/linux-more-command-and-less-command-examples/>

[19]: <http://www.tecmint.com/tag/linux-tricks/>

[20]: <http://www.tecmint.com/author/aaronkili/>

[21]: <https://github.com/yangmingming>

[22]: <https://github.com/jasminepeng>

[23]: <https://github.com/LCTT/TranslateProject>

[24]: <https://linux.cn/article-8216-1.html?pr>