

GNU Grep: Print lines matching a pattern

version 2.14, 15 May 2012

grep 手册 (grep, 是一个规则匹配引擎.)

Copyright © 1999-2002, 2005, 2008-2012 Free Software Foundation, Inc.

遵循 GNU 自由文档许可条款, 用户可以复制, 发布, 修改该文档内容, 1.3 版 或 最新版由自由软件组织(FSF)发布; 没有固定段落, 没有封面文字, 没有封底文字.

“GNU Free Documentation License” 部分包含证书副本.

目录

1	简介	1
2	调用 grep	2
2.1	命令行选项.....	2
2.1.1	常规程序信息.....	2
2.1.2	匹配控制.....	2
2.1.3	常规输出控制.....	3
2.1.4	输出行前缀控制.....	4
2.1.5	上下文控制.....	6
2.1.6	文件和目录的选择.....	6
2.1.7	其他选项.....	8
2.2	环境变量.....	8
2.3	退出状态.....	11
2.4	grep 议程.....	12
3	正则表达式	13
3.1	基本结构.....	13
3.2	字符集和括号表达式.....	13
3.3	反斜线字符和特殊表达式.....	15
3.4	锚位.....	16
3.5	反向引用和子表达式.....	16
3.6	基本正则与扩展正则对比.....	16
4	用法	17
5	bug 反馈	20
5.1	已知 Bugs.....	20
6	版权	21
6.1	GNU 免费文档证书.....	21
	Index	30

1. 简介

grep 可根据提供的匹配模式列表,查询文件中的匹配行.发现匹配行后,行内容会被标准输出(默认),如果使用了其他参数选项,可产生其他格式的输出生。

grep 用于匹配文本,它对输入行的长度没有限制,除非受限于可用内存大小,并且它可以匹配行内任意字符。

2. 调用 grep

grep 命令行方法如下：

grep options pattern input_file_names

可以不指定或指定多个 option 选项。如果不采用 '-e pattern' 或 '-f file'，那么 pattern 就是我们所见的内容。同样的，我们可以不指定或指定多个 input_file_names 选项。

2.1 命令行选项

grep 有很多参数：一些源于 POSIX，一些是 GNU 的扩展。长参数名，通常是 GNU 的扩展，这些参数甚至来源于 POSIX。POSIX 标准的短文件名，是为了方便进行 POSIX 移植而设计的。部分参数还提供老版本及特殊情况的兼容。

2.1.1 常规程序信息

--help	打印命令行帮助信息，bug 报告地址，并退出程序。
-V --version	打印 grep 版本号，bug 报告中应该包含此版本号。

2.1.2 匹配控制

-e pattern --regexp=pattern	可指定多个查询规则，或使用 '-' 保护匹配规则
-f file --file=file	从文件中获取匹配规则，一行表示一个规则，空文件不包含规则，因此不匹配任何内容
-i -y --ignore-case	忽略匹配规则和输入文件中的大小写
-v --invert-match	获取未匹配的行
-w --word-regexp	获取包含整个词的匹配行，词由字母，数字，下划线组成
-x --line-regexp	整行匹配

2.1.3 常规输出控制

-c --count	输出文件的匹配行数，而非匹配内容的常规输出.与'-v'参数一起使用，可统计未匹配的行数。
--color[=WHEN] --colour[=WHEN]	颜色值可通过环境变量 GREP_COLORS 定义，默认是' ms=01;31:mc=01;31:sl=:cx=:fn=35:ln=32:bn=32:se=36 '， WHEN 值可设置为 never, always, auto 。
-L --files-without-match	仅输出未匹配的文件名
-l --files-with-matches	仅输出匹配文件名
-m num --max-count=num	匹配 num 行后，停止读取文件。当与'-c' 或 '--count'一起使用时，输出的匹配行计数不会大于 num ，与'-v'或'--invert-match'一起使用时，会输入 num 次匹配过程中的未匹配项。
-o --only-matching	打印行内匹配的内容(非空)
-q --quite --silent	不做标准输出。如果发现匹配项，立即退出，即使中间出现错误。
-s --no-messages	针对不存在或不可读的文件，不输出错误信息

2.1.4 输出行前缀控制

前缀输出，通常有 文件名, 行号, 字节偏移, 可通过下面参数指定。

-b --byte-offset	在每行输出内容的前面，打印字符偏移
-H --with-filename	在匹配内容前，打印文件名，查询多个文件时，这是默认设置。
-h --no-filename	输出时，不打印文件名，查询单个文件时，这是默认设置
--label=LABEL	使用标签，将标准输入以文件输入的方式表示 gzip -cd foo.gz grep -label=foo -H something
-n --line-number	显示匹配行的行号
-T --initial-tab	tab 对齐
-u --unix-byte-offsets	
-Z --null	文件名后面的一个字符使用一个 0 字节代替(ASCII NUL 字符)

2.1.5 上下文控制

<code>-A num</code> <code>--after-context=num</code>	打印匹配行后的 <code>num</code> 行(<code>num</code> 不包含匹配行)
<code>-B num</code> <code>--before-context=num</code>	打印匹配行前的 <code>num</code> 行(<code>num</code> 不包含匹配行)
<code>-C num</code> <code>-num</code> <code>--context=num</code>	打印匹配行前后各 <code>num</code> 行
<code>--group-separator=string</code>	使用 <code>'-A','-B','-C'</code> 参数时, 使用 <code>string</code> 代替默认的 <code>'--'</code> 分割线
<code>--no-group-separator</code>	不使用分割线

2.1.6 文件和目录的选择

<code>-a</code> <code>--text</code>	将二进制文件作为文本文件处理, 等效于 <code>'--binary-files=text'</code>
<code>--binary-files=type</code>	默认, <code>type</code> 为 <code>binary</code> , 会有二进制文件类型提示, 如果未匹配, 则无提示。若设定 <code>type</code> 为 <code>without-match</code> , 则不匹配二进制文件, 等效于参数 <code>'-l'</code> 。若设定 <code>type</code> 为 <code>text</code> , 则会将二进制作作为文本处理, 等效于参数 <code>'-a'</code> , 此选项会输出不友好的二进制文件内容。
<code>-D action</code> <code>--devices=action</code>	如果输入的文件是一个设备, 文件流, 或 <code>socket</code> , 则可设定此选项, 决定如何处理。若 <code>action</code> 设定为 <code>skip</code> , 则会跳过不处理。默认情况下, 会读取命令行下设备, 若启用 <code>'-R'</code> 或 <code>'-r'</code> 选项, 则会跳过处理。该参数对标准输入的文件无效。
<code>-d action</code> <code>--directories=action</code>	如果指定的是目录, 则可设定 <code>action</code> , 决定如何处理它。默认 <code>action</code> 为 <code>read</code> , 则将目录作为常规文件处理(有些操作系统或文件系统不允许这样做, 会导致 <code>grep</code> 打印每个目录的错误信息, 或跳过处理目录)。若 <code>action</code> 设定为 <code>skip</code> , 则会跳过目录。若 <code>action</code> 设为 <code>recurse</code> , <code>grep</code> 会进行递归处理, 等效于 <code>'-r'</code> 选项。
<code>--exclude=glob</code>	跳过匹配 <code>glob</code> 的文件名, <code>glob</code> 可以是 <code>'*'</code> , <code>'?'</code> , <code>'[...]'</code> , 可使用 <code>\</code> 引用通配符或反斜线字符。
<code>--exclude-from=file</code>	可通过文件指定多个 <code>glob</code> (类似于指定多个 <code>--exclude=glob</code>)
<code>--exclude-dir=dir</code>	递归处理时, 可排除匹配规则的目录
<code>-l</code>	设定二进制文件无匹配项, 等效于 <code>'--binary-files=without-match'</code>
<code>--include=glob</code>	只处理匹配规则的文件, 与 <code>--exclude</code> 相反,

-r --recursive	对目录进行递归处理，等效于' —directories=recurse '，会跟随命令行下的符号链接，不会跟随递归过程中的符号链接。
-R --dereference-recursive	递归处理，跟随递归过程中的符号链接。

2.1.7 其他选项

--line-buffered	输出行缓冲内容，会引起性能问题
--mmap	此选项已放弃使用。
-U --binary	处理二进制方式处理文件。默认，在 MS-DOS 和 Windows 平台下， grep 会根据' —binary-files '选项来判定是文本，还是二进制。如果 grep 认为文件是文本，则会去掉文件原始内容中的 CR 字符(以确保 ^ 和 \$ 正常工作)。如果文本文件行以 CR/LF 结尾，会导致正则表达式匹配失败。此选项仅对 MS-DOS 和 Windows 生效。
-z --null-data	将输入内容作为行的合集，每个集合以空字节作为结束(ASCII NUL 字符)

2.2 环境变量

2.3 退出状态

2.4 grep 议程

grep 使用给定的匹配规则，查询输入文件中的匹配行。默认，**grep** 会打印匹配行。如果输入文件为 '-'，这代表使用标准输入作为输入文件内容。如果未指定输入，且命令参数中有递归选项，那么 **grep** 会搜索工作目录 '.'；否则，**grep** 会搜索标准输入。**grep** 有四种变种，由下面参数控制：

-G --basic-regexp	默认选项，匹配规则为常规正则表达式(BRE)
-E --extended-regexp	匹配规则为扩展正则表达式(ERE)
-F --fixed-strings	将匹配规则视为分行的字符串列表
-P --perl-regexp	将匹配规则，视为 perl 正则表达式。这是一个测试功能，对未实现的功能，会给出预警提示。

另外，还有两个程序可用，分别是 **egrep** 和 **fgrep**。**egrep** 等效于 '**grep -E**'。**fgrep** 等效于 '**grep -F**'。

3. 正则表达式

正则表达式是描述字符的规则。正则表达式的构造类似于算数表达式，使用各种运算符，将短的表达式结合在一起。**grep** 可读懂三种正则表达式语法：基本正则表达式(BRE)，扩展正则表达式(ERE)，**perl** 正则表达式。**GNU** 版本的 **grep**，基本正则表达式与扩展正则表达式，在功能上没什么不同。在其他的地方，基本正则表达式则不像扩展正则表达式那样强劲。接下来描述的是扩展正则表达式，与基本正则表达式的不同之处，会在后面进行说明。**perl** 正则表达式提供了许多额外的功能，请参阅 **pcresyntax** 手册文档，这种表达式可能并不适用于每种系统。

3.1 基本结构

正则结构块，由匹配单个字符的表达式组成。正则匹配的大多数字符，包括字母，数字。任何拥有特殊意义的原字符，如果想要引用它，在其前面放置一个反斜线进行转义。

.	匹配任意单个字符
?	最多出现一次
*	0 次以上
+	1 次以上
{n}	n 次
{n,}	n 次或 n 次以上
{,m}	最多出现m 次
{n,m}	最少 n 次，最多 m 次

空的正则表达式，匹配空字符。两个正则表达式可以拼接在一起，结果由匹配各正则表达的结果拼接而成。两个正则表达式间使用'|'操作符。

重复的优先级高于串联操作，反过来又优先于交替。一整个表达式可以用括号括起来，以覆盖这些优先级规则，并形成子表达式。

3.2 字符集和括号表达式

为了获得传统的括号表达式需设定 **LC_ALL** 为 **C**。

[:alnum:]	[:alpha:] 与 [:digit:] 的合集，等同于 [0-9A-Za-z]
[:alpha:]	[:lower:] 与 [:upper:] 的合集，等同于 [A-Za-z]
[:blank:]	空白字符：空格键和 tab 键
[:cntrl:]	控制字符，字符对应的十进制数 000 到 037，还有 177(del)
[:digit:]	0 1 2 3 4 5 6 7 8 9
[:graph:]	可视字符， [:alnum:] 与 [:punct:] 的合集
[:lower:]	小写字母
[:print:]	可打印字符： [:alnum:] ， [:punct:] ，space
[:punct:]	!"#\$%&'()*+,-./:;<=>?@[\] ^ _ ' { } ~
[:space:]	空白符

<code>[[:upper:]]</code>	大小写字母
<code>[[:xdigit:]]</code>	十六进制数字: 0 1 2 3 4 5 6 7 8 9 A B C D E F a b c d e f

<code>]</code>	括号表达式结束符, 如果你想在列表中使用']'字符, 请将其放在第一位。
<code>[.</code>	打开排序符号
<code>.]</code>	关闭排序符号
<code>[=</code>	打开等价类
<code>=]</code>	关闭等价类
<code>[:</code>	开启字符类符号
<code>:]</code>	关闭字符类符号
<code>-</code>	如果该字符不在起始或结尾, 那么表示范围
<code>^</code>	表示字符不在列表中, 如果你想使用这个字符, 请勿将其放在第一位。

3.3 反斜线字符和特殊表达式

<code>\b</code>	匹配单词边缘空字符
<code>\B</code>	匹配非单词边缘的空字符
<code>\<</code>	
<code>\></code>	
<code>\w</code>	
<code>\W</code>	
<code>\s</code>	
<code>\S</code>	

3.4 锚位

`^`与`$`用于匹配字符起始和结束位的空字符。

3.5 反向引用和子表达式

反向引用`\n` (`n`表示一个数字), 可指定第 `n` 个子表达式匹配的内容。例如:`(a)\1` 会匹配 `'aa'`。交替使用时, 如果组并未参与匹配, 那么反向引用会导致整个匹配失败。例如:`'a(.)|b\1'`不会匹配`'ba'`。使用`'-e'`可指定多个表达式, 或者使用`'-f file'`从文件指定。

3.6 基本正则与扩展正则对比

在基本正则表达式中, 元字符`'?', '+', '{', '|', '(', ')'`失去其原意, 通常需要加上反斜线`\?', '\+', '\{', '\|', '\(', '\)'`。

传统的 `egrep` 不支持原字符`'{'`, 一些 `egrep` 使用`'\{'`代替。可移植脚本应该避免在`'grep -E'`模式中使用`'{'`, 推荐使用`'[{'`匹配`'{'`。

global/regular expression/print
g/re/p
Fixed grep
Extended grep

5 Reporting bugs

Email bug reports to bug-grep@gnu.org, a mailing list whose web page is <http://lists.gnu.org/mailman/listinfo/bug-grep>. The Savannah bug tracker for grep is located at <http://savannah.gnu.org/bugs/?group=grep>.

5.1 Known Bugs

Large repetition counts in the ‘{n,m}’ construct may cause grep to use lots of memory. In addition, certain other obscure regular expressions require exponential time and space, and may cause grep to run out of memory.

Back-references are very slow, and may require exponential time.

6 Copying

GNU grep is licensed under the GNU GPL, which makes it free software.

The “free” in “free software” refers to liberty, not price. As some GNU project advocates like to point out, think of “free speech” rather than “free beer”. In short, you have the right (freedom) to run and change grep and distribute it to other people, and—if you want—charge money for doing either. The important restriction is that you have to grant your recipients the same rights and impose the same restrictions.

This general method of licensing software is sometimes called open source.

The GNU project prefers the term “free software” for reasons outlined at

<http://www.gnu.org/philosophy/open-source-misses-the-point.html>.

This manual is free documentation in the same sense. The documentation license is included below. The license for the program is available with the source code, or at <http://www.gnu.org/licenses/gpl.html>.

6.1 GNU Free Documentation License

Version 1.3, 3 November 2008

Copyright c 2000-2002, 2007-2008, 2010-2012 Free Software Foundation, Inc.

<http://fsf.org/>

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document free in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondly, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of “copyleft”, which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software. We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The “Document”,

Index

```
*
* :
+
+ :
-
--after-context :
--basic-regexp :
--before-context :
--binary :
--binary-files :
--byte-offset :
--color :
--colour :
--context :
--count :
--dereference-recursive :
--devices :
--directories :
--exclude :
--exclude-dir :
--exclude-from :
--extended-regexp :
--file :
--files-with-matches :
--files-without-match :
--fixed-strings :
--group-separator :
--help :
--ignore-case :
--include :
--initial-tab :
--invert-match :
--label :
--line-buffered :
--line-number :
--line-regexp :
--max-count :
--mmap :
--no-filename :
--no-messages :
--null :
--null-data :
--only-matching :
--perl-regexp :
--quiet :
--recursive :
--regexp=pattern :
--silent :
--text :
--unix-byte-offsets :

--version :
--with-filename :
--word-regexp :
-a
-A
-b
-B
-c
-C
-d
-D
-e
-E
-f
-F
-G
-h
-H
-i
-l
-L
-m
-n
-num:
-o
-P
-q
-r
-R
-s
-T
-u
-U
-v
-V
-w
-x
-y
-z
-Z
:
?
? :
_N_GNU_nonoption_argv_flags_ environment
variable :
--version :
--with-filename :
--word-regexp :
-a
```