



GREP

Section: Manuel de l'utilisateur Linux (1)

NOM

grep, egrep, fgrep - Afficher les lignes correspondant à un motif donné.

SYNOPSIS

```
grep [options] MOTIF [FICHIER. . .]
```

```
grep [options] [-e MOTIF | -f FICHIER] [FICHIER. . .]
```

DESCRIPTION

Grep recherche dans les *FICHIERS* indiqués (ou depuis l'entrée standard si aucun fichier n'est fourni, ou si le nom - est mentionné) les lignes correspondant à un certain *MOTIF*. Par défaut, **grep** affiche les lignes qui correspondent au motif.

De plus, il existe deux variantes du programme : **egrep** et **fgrep**. **Egrep** est identique à **grep -E**. **Fgrep** est identique à **grep -F**.

OPTIONS

-A NOMBRE, **-after-context=NOMBRE**

Afficher *NOMBRE* lignes de contexte supplémentaires après les

lignes qui concordent. Une ligne contenant -- est mise entre les groupes contigus de correspondances.

-a, -text

Traiter un fichier binaire comme s'il s'agissait de texte ; cette option est équivalente à l'option **-binary-files=text**.

-B NOMBRE, -before-context=NOMBRE

Afficher *NOMBRE* lignes de contexte supplémentaires avant les lignes qui concordent. Une ligne contenant -- est mise entre les groupes contigus de correspondances.

-c NOMBRE, -context=NOMBRE

Afficher *NOMBRE* lignes de contexte. Une ligne contenant -- est mise entre les groupes contigus de correspondances.

-b, -byte-offset

Avant chaque ligne, afficher sa position relative (en octets) au sein du fichier.

-binary-files=TYPE

Si les premiers octets d'un fichier indiquent qu'il contient des données binaires, supposer que le fichier est de type *TYPE*. Par défaut, *TYPE* est **binary** (binaire), et soit **grep** affiche un message d'une ligne disant que le fichier binaire concorde, soit il n'affiche rien s'il n'y a pas concordance. Si *TYPE* est **without-match** (sans concordance), **grep** suppose qu'un fichier binaire ne concorde jamais. C'est équivalent à l'option **-i**. Si *TYPE* est **text** (texte), **grep** traite un fichier binaire comme du texte. C'est équivalent à l'option **-a**. *Attention* : **grep -binary-files=text** peut afficher des données binaires indésirables, qui peuvent avoir des conséquences fâcheuses si la sortie est envoyée vers un terminal, et si le pilote de celui-ci interprète cet affichage comme étant des commandes.

-colour[=QUAND], -color[=QUAND]

Entourer la chaîne correspondante avec les marqueurs trouvés dans la variable d'environnement **GREP_COLOR**. Les valeurs possibles de *QUAND* sont **never** (jamais), **always** (toujours) ou **auto** (automatique).

-c - -count

Ne pas afficher les résultats normaux. À la place, afficher un compte des lignes correspondantes pour chaque fichier. Avec l'option **-v**, **- -invert-match** (voir plus bas), afficher le nombre de lignes ne correspondant pas au motif.

-D ACTION, - -devices=ACTION

Si le fichier est un périphérique, une FIFO ou une socket, utiliser *ACTION* dessus. Par défaut, *ACTION* est **read** (lecture), ce qui signifie que les périphériques sont lus comme des fichiers normaux. Si *ACTION* est **skip** (escamotage), les périphériques ne sont pas pris en compte et aucun message n'est affiché.

-d ACTION, - -directories=ACTION

Si le fichier est un répertoire, utiliser *ACTION*. Par défaut, *ACTION* est **read** (lecture), ce qui signifie que les répertoires sont lus comme des fichiers normaux. Si *ACTION* est **skip** (escamotage), les répertoires ne sont pas pris en compte et aucun message n'est affiché. Si *ACTION* est **recurse** (récursivité), **grep** lit tous les fichiers présents dans chaque répertoire, récursivement. C'est équivalent à l'option **-r**.

-E, - -extended-regexp

Interpréter le *MOTIF* comme une expression rationnelle étendue (voir plus bas).

-e MOTIF, - -regexp=MOTIF

Utiliser le *MOTIF* indiqué. Ceci permet de protéger les motifs commençant par **-**.

-F, - -fixed-strings

Interpréter le *MOTIF* comme une liste de chaînes de caractères, séparées par des sauts de ligne. Chacune d'entre elles doit être recherchée.

-P, - -perl-regexp

Interpréter le *MOTIF* comme étant une expression rationnelle de Perl.

-f FICHIER, - -file=FICHIER

Lire les motifs dans le *FICHIER* indiqué, un motif par ligne. Un

fichier vide ne contient aucun motif, si bien qu'aucune concordance n'est trouvée.

-G, - -basic-regexp

Interpréter le *MOTIF* comme une expression rationnelle simple (voir plus bas). C'est le comportement par défaut.

-H, - -with-filename

Afficher le nom du fichier pour chaque concordance.

-h, - -no-filename

Ne pas afficher le nom des fichiers dans les résultats lorsque plusieurs fichiers sont parcourus.

- -help

Afficher un court message d'aide.

-I

Traiter un fichier binaire comme s'il ne contenait aucune concordance ; c'est équivalent à l'option **- -binary-files=without-match**.

-i, - -ignore-case

Ignorer les différences majuscules/minuscules aussi bien dans le *MOTIF* que dans les fichiers.

-L, - -files-without-match

Ne pas afficher les résultats normaux. À la place, indiquer le nom des fichiers pour lesquels aucun résultat n'aurait été affiché. La recherche à l'intérieur de chaque fichier s'arrête dès la première concordance.

-l, - -files-with-matches

Ne pas afficher les résultats normaux. À la place, indiquer le nom des fichiers pour lesquels des résultats auraient été affichés. La recherche à l'intérieur de chaque fichier s'arrête dès la première concordance.

-m *NOMBRE*, - -max-count=*NOMBRE*

Arrêter de lire un fichier après avoir lu *NOMBRE* lignes qui correspondent. Si l'entrée est l'entrée standard prise depuis un fichier normal, et si *NOMBRE* lignes qui correspondent sont

affichées, **grep** s'assure avant de s'arrêter que l'entrée standard est positionnée juste après la dernière ligne qui concorde, même s'il y a des lignes de contexte supplémentaires qui viennent ensuite. Ceci permet au processus d'appel de redémarrer une recherche. Quand **grep** s'arrête après *NOMBRE* lignes de concordance, il affiche éventuellement des lignes de contexte. Quand l'option **-c** ou **-count** est utilisée, **grep** n'affiche pas plus que *NOMBRE* lignes. Quand l'option **-v** ou **-invert-match** est aussi utilisée, **grep** s'arrête après avoir affiché *NOMBRE* lignes qui ne concordent pas.

-mmap

Si possible, utiliser la fonction système **mmap(2)** pour lire les données, à la place de la fonction système **read(2)** utilisée par défaut. Dans certaines circonstances, **-mmap** permet de meilleures performances. Cependant, **-mmap** peut avoir un comportement indéfini (y compris la génération d'un *core dump*) si un fichier voit sa taille réduite pendant que **grep** le lit, ou si une erreur d'entrée-sortie survient.

-n, -line-number

Ajouter à chaque ligne de sortie un préfixe contenant son numéro dans le fichier.

-o, -only-matching

N'afficher que la partie d'une ligne qui correspond au *MOTIF*.

-label=TIQUETTE

Afficher les données provenant de l'entrée standard comme si elles provenaient du fichier *TIQUETTE*. C'est particulièrement utile pour des outils comme **zgrep**, par exemple, qui font :
gzip -cd foo.gz |grep --label=truc quelquechose

-line-buffered

Vider la sortie après chaque ligne. Ceci peut détriorer les performances.

-q, -quiet, -silent

Ne pas afficher les résultats normaux. Sortir immédiatement avec un code de retour égal à zéro si aucune concordance n'est trouvée, et même si une erreur est détectée. Voyez

aussi l'option **-s** ou **--no-messages**.

-R, -r, --recursive

Lire tous les fichiers à l'intérieur de chaque répertoire, récursivement. C'est équivalent à l'option **-d recurse**.

-include=MOTIF

Procéder récursivement dans les répertoires, pour les fichiers qui correspondent au *MOTIF*.

--exclude=MOTIF

Procéder récursivement dans les répertoires, pour les fichiers qui ne correspondent pas au *MOTIF*.

-s, --no-messages

Ne pas afficher les messages d'erreur concernant les fichiers inexistantes ou illisibles. Note de portabilité : la différence de GNU **grep**, le **grep** traditionnel n'est pas conforme au standard POSIX.2, car le **grep** traditionnel n'a pas d'option **-q** et son option **-s** agit comme l'option **-q** du **grep** de GNU. Les scripts shell destinés à être utilisés avec ces différentes versions devraient éviter d'utiliser les options **-q** et **-s**, et devraient plutôt rediriger la sortie vers `/dev/null`.

-U, --binary

Traiter les fichiers comme s'ils étaient des fichiers binaires. Par défaut, sous MS-DOS et MS-Windows, **grep** détermine le type de fichier en regardant le contenu des 32 premiers kilo-octets. Si **grep** décide que le fichier est un fichier de texte, il enlève les retours chariot (CR) du contenu du fichier original (afin que les expressions avec **^** et **\$** fonctionnent correctement). L'option **-U** modifie ce comportement, tous les fichiers sont alors lus et traités tels quels. Si le fichier est un fichier de texte avec des paires CR-LF en fin de ligne, certaines expressions rationnelles peuvent échouer. Cette option n'a aucun effet sur des plates-formes autres que MS-DOS et MS-Windows.

-u, --unix-byte-offsets

Afficher la position relative en octets dans le style Unix. Avec cette option, **grep** affiche la position relative en octets comme si le fichier était un fichier de texte de type Unix, c'est-à-dire avec les

caractères CR supprimés. Ceci permet d'avoir un résultat identique à celui retourné par **grep** sur une machine Unix. Cette option n'a d'effet que si elle est utilisée conjointement avec l'option **-b**, et sur une machine MS-DOS ou MS-Windows.

-V, - -version

Afficher le numéro de version de **grep** sur la sortie d'erreur standard. Ce numéro de version devra être inclus dans tous les rapports de bogues (voir plus bas).

-v, - -invert-match

Inverser la mise en concordance, pour sélectionner les lignes ne correspondant pas au motif.

-w, - -word-regexp

Ne sélectionner que les lignes contenant une concordance formant un mot complet. La sous-chaîne correspondante doit donc être soit au début de la ligne, soit être précédée d'un caractère ne pouvant entrer dans la constitution d'un mot. De même elle doit se trouver soit à la fin de la ligne, soit être suivie par un caractère ne pouvant entrer dans la constitution d'un mot. Les caractères composant les mots sont les lettres, les chiffres et le souligné («**_**»).

-x, - -line-regexp

Ne sélectionner que les concordances qui occupent une ligne entière.

-y

Synonyme obsolète pour **-i**.

-Z, - -null

Afficher un octet nul (le caractère ASCII **NUL**) à la place du caractère qui suit d'ordinaire le nom du fichier. Par exemple, **grep -IZ** affiche un octet nul après chaque nom de fichier, à la place du saut de ligne. Cette option permet de rendre la sortie non ambiguë, même quand les noms de fichiers contiennent des caractères inhabituels, comme des sauts de ligne. Cette option peut être utilisée avec des commandes telles que **find -print0**, **perl -0**, **sort -z**, et **xargs -0** pour traiter des fichiers avec des noms quelconques, même ceux contenant des sauts de ligne.

EXPRESSIONS RATIONNELLES

Une expression rationnelle (*regular expression*) est un motif qui permet de d  crire un ensemble de cha  nes. Les expressions rationnelles sont construites comme des op  rations arithm  tiques ; elles utilisent diff  rents op  rateurs pour combiner des expressions plus petites.

Grep comprend deux versions diff  rentes pour la syntaxe des expressions rationnelles :   « simple   » (*basic*) et   «   tendue   » (*extended*). Dans la version GNU de **grep**, il n'y a pas de diff  rence dans les fonctionnalit  s disponibles, quelle que soit la syntaxe utilis  e. Dans d'autres implantations, les expressions rationnelles simples sont moins puissantes. La description ci-dessous correspond aux expressions   tendues, les diff  rences avec les expressions simples   tant r  sum  es ensuite.

Les briques   l  mentaires sont les expressions rationnelles correspondant    un seul caract  re. La plupart des caract  res, y compris les lettres et les chiffres, sont des expressions rationnelles qui concordent avec eux-m  mes. Tout m  ta-caract  re ayant une signification sp  ciale doit   tre prot  g   en le faisant pr  c  der d'une contre-oblique (*backslash*).

Une liste de caract  res, encadr  e par [et] peut   tre mise en correspondance avec n'importe quel caract  re appartenant    la liste. Si le premier caract  re de la liste est l'accent circonflexe ^ alors la mise en correspondance se fait avec n'importe quel caract  re *absent* de la liste. Par exemple, l'expression rationnelle [0123456789] concorde avec n'importe quel chiffre.

Entre ces crochets, un intervalle de caract  res peut   tre indiqu   en donnant le premier et le dernier caract  re, s  par  s par un tiret. Il correspond    n'importe quel caract  re compris entre le premier et le dernier caract  re (ceux-ci inclus), l'ordre des caract  res d  pendant des param  tres r  gionaux (*locale*, en anglais) en cours. Ainsi avec la valeur par d  faut (appel  e   « C   »), [a-d] est   quivalent    [abcd].

Avec beaucoup de paramètres régionaux, les caractères sont triés en suivant l'ordre des dictionnaires, et `[a-d]` n'est alors pas équivalent à `[abcd]`, mais à `[aBbCcDd]`, par exemple. Pour que ces listes aient le comportement usuel de C, vous pouvez positionner la variable d'environnement `LC_ALL` à la valeur `C`.

Enfin, il existe certaines classes de caractères prédéfinies. Leurs noms sont assez explicites : `[:alnum:]`, `[:alpha:]`, `[:cntrl:]`, `[:digit:]` (chiffres), `[:graph:]`, `[:lower:]` (minuscules), `[:print:]` (affichables), `[:punct:]` (ponctuation), `[:space:]` (espace), `[:upper:]` (majuscules), et `[:xdigit:]` (chiffres hexadécimaux). Par exemple, `[:alnum:]` correspond à `[0-9A-Za-z]`, à la différence près que le dernier dépend des paramètres régionaux C et du codage de caractères ASCII, alors que le premier est plus portable. Remarquez que les crochets dans les noms de classes font partie intégrante du nom symbolique, et qu'ils doivent donc être inclus en plus des crochets encadrant la liste. La plupart des méta-caractères perdent leur signification spéciale au sein des listes. Pour inclure un caractère], mettez-le en premier dans la liste. De même, pour inclure un caractère ^, placez-le n'importe où¹ sauf au début de la liste. Enfin, pour inclure un -, placez-le en dernier.

Le point `.` correspond à n'importe quel caractère. Le symbole `\w` est un synonyme de `[:alnum:]` et `\W` un synonyme de `[^[:alnum:]]`.

L'accent circonflexe `^` et le symbole dollar `$` sont des méta-caractères correspondant respectivement à une chaîne vide au début et en fin de ligne. Les symboles `\<` et `\>` correspondent respectivement à une chaîne vide en début et en fin de mot. Le symbole `\b` correspond à une chaîne vide à l'extrémité d'un mot, et `\B` correspond à une chaîne vide ne se trouvant pas à une extrémité de mot.

Une expression rationnelle correspondant à un caractère unique peut être suivie par l'un des opérateurs de répétition suivants :

?

L'opérateur `?` est facultatif et peut être rencontré au plus une fois.

*

L'opérateur $*$ peut être rencontré zéro ou plusieurs fois.

+

L'opérateur $+$ peut être rencontré une ou plusieurs fois.

{*n*}

L'opérateur $\{n\}$ doit être rencontré exactement n fois.

{*n*,}

L'opérateur $\{n,\}$ doit être rencontré n fois ou plus.

{*n*,*m*}

L'opérateur $\{n,m\}$ doit être rencontré au moins n fois, mais au plus m fois.

Deux expressions rationnelles peuvent être juxtaposées ; l'expression résultante correspondra à toute chaîne formée par la juxtaposition de deux sous-chaînes correspondant respectivement aux deux expressions.

Deux expressions rationnelles peuvent être reliées par l'opérateur infixé $|$; l'expression résultante correspondra à toute chaîne concordant avec l'une ou l'autre des deux expressions.

Les opérations ont priorité sur les juxtapositions, qui à leur tour ont priorité sur les alternatives. Une sous-expression peut être entourée par des parenthèses pour modifier ces règles de priorité.

La référence inverse $\backslash n$, où n est un chiffre unique, correspond à la sous-chaîne d'aj mise en correspondance avec la n -ième sous-expression rationnelle entre parenthèses.

Dans les expressions rationnelles simples, les méta-caractères $?, +, \{, |, (, \)$ perdent leur signification spéciale, il faut utiliser à la place leurs versions avec la contre-oblique $\backslash?, \backslash+, \backslash\{, \backslash|, \backslash(, \)$ et $\backslash)$.

La version traditionnelle d'**egrep** ne connaît pas le méta-caractère `{`, et certaines implantations d'**egrep** utilisent `{` à la place, si bien que des scripts shell portables devraient éviter `{` dans les motifs d'**egrep** et utiliser `[{]` pour désigner un caractère `{`.

GNU **egrep** essaie d'émuler l'usage traditionnel en supposant que `{` n'est pas spécial au cas où il rendrait invalide l'expression qu'il commence. Par exemple, la commande shell **egrep** `'{1'` recherche la chaîne composée des deux caractères `{1` au lieu de signaler une erreur de syntaxe dans l'expression rationnelle. POSIX.2 permet ce comportement comme une extension à la norme, mais les scripts portables devraient l'éviter.

VARIABLES D'ENVIRONNEMENT

Le comportement de **grep** est modifié par les variables d'environnement suivantes : **LC_ALL**, **LC_truc** et **LANG**, dans cet ordre. La variable positionnée en premier détermine le choix des paramètres régionaux. Par exemple, si **LC_ALL** n'est pas positionnée, mais **LC_MESSAGES** vaut **fr_FR**, alors le français est utilisé pour l'affichage des messages. Le défaut C est utilisé si aucune variable d'environnement n'est trouvée, ou si le catalogue des paramètres régionaux n'est pas installé, ou bien si **grep** a été compilé sans le support pour les langues nationales (NLS).

GREP_OPTIONS

Cette variable définit des options qui seront ajoutées avant les options de la ligne de commande. Par exemple, si **GREP_OPTIONS** vaut `« -binary-files=without-match -directories=skip »`, **grep** se comporte comme si les deux options `-binary-files=without-match` et `-directories=skip` avaient été spécifiées avant les options explicites. Différentes options peuvent être séparées par des espaces, et une contre-oblique supprime la signification spéciale du caractère suivant, ce qui permet de spécifier une option contenant un espace ou une contre-oblique.

GREP_COLOR

Spécifie les marqueurs pour la mise en relief.

LC_ALL LC_COLLATE LANG

Ces variables sp cifient le choix des param tres r gionaux pour **LC_COLLATE**, qui d termine l'ordre des caract res utilis s dans des intervalles tels que [a-z].

LC_ALL, LC_CTYPE, LANG

Ces variables sp cifient le choix des param tres r gionaux pour **LC_CTYPE**, qui d termine le codage de caract res utilis s, par exemple pour indiquer quels caract res sont consid r s comme  tant des espaces.

LC_ALL, LC_MESSAGES, LANG

Ces variables sp cifient le choix des param tres r gionaux pour **LC_MESSAGES**, qui d termine la langue utilis e par **grep** pour ses messages. Avec le d faut C, les messages sont en am ricain.

POSIXLY_CORRECT

Si cette variable est positionn e, **grep** se comporte comme indiqu  dans la norme POSIX.2. Sinon, **grep** se comporte plus comme les autres programmes GNU. POSIX.2 requiert que les options qui suivent des noms de fichiers soient consid r es aussi comme des noms de fichiers. Par d faut, ces options sont d plac es avant la liste des op randes et sont trait es comme des options. POSIX.2 requiert aussi que les options non reconnues soient consid r es comme  « ill gales  » ; mais comme elles n'enfreignent pas vraiment la loi, elles sont rapport es comme  tant  « invalides  » par d faut.

POSIXLY_CORRECT d sactive aussi l'option **_N_GNU_nonoption_argv_flags_**, qui est d crite plus bas.

_N_GNU_nonoption_argv_flags_

(ici, *N* est l'identifiant num rique du processus de **grep**). Si le *i*-i me caract re de la valeur de cette variable d'environnement vaut **1**, le *i*-i me op rande de **grep** n'est pas consid r  comme  tant une option, m me s'il semble l' tre. Un shell peut placer cette variable dans l'environnement de chaque commande lanc e, pour sp cifier quels op randes sont le r sultat du remplacement de m ta-caract res et ne doivent donc pas  tre consid r s comme des options. Ce comportement n'est

prÃ©sent qu'avec la bibliothÃ¨que C de GNU, et seulement si `POSIXLY_CORRECT` n'est pas positionnÃ©e.

DIAGNOSTICS

Normalement, le code de retour est 0 si des concordances ont Ã©tÃ© trouvÃ©es, et 1 si aucune concordance n'a Ã©tÃ© faite. Mais le code est 2 si une erreur est survenue, Ã moins que les options `-q`, `-quiet` ou `-silent` ne soient utilisÃ©es et qu'une ligne ne soit trouvÃ©e.

BOGUES

Envoyez les rapports de bogue ([ndt] en anglais !) Ã bug-gnu-utils@gnu.org. Assurez-vous d'inclure le mot Â« grep Â» quelque part dans le sujet du message.

Dans les constructions $\{m,n\}$ de grandes valeurs de rÃ©pÃ©tition peuvent pousser grep Ã utiliser beaucoup de mÃ©moire. De plus, certaines autres expressions rationnelles tordues peuvent prendre un temps trÃ¨s long, et mener Ã un manque de mÃ©moire.

Les rÃ©fÃ©rences inverses sont trÃ¨s lentes et peuvent demander un temps trÃ¨s long.

TRADUCTION

Christophe Blaess, 1997.

Denis Barbier, 2003.

