



UAX

UNIVERSIDAD ALFONSO X EL SABIO

Búsqueda de patrones en Bash

GRADO EN BIOMEDICINA

Bioinformática

idelhgar@uax.es

2023

Grep

- Este comando permite buscar un patrón en archivos de texto.
- Como regla general, los patrones se ponen entre comillas para evitar que la Shell los confunda con sus propios caracteres especiales
- **grep [opciones] pattern [archivo]**
- **grep -help**
- **grep --version**

Versión	Descripción
<code>-G o --basic-regexp</code>	Interpreta los patrones como regex básicas (BREs). Es la versión estándar de la herramienta
<code>-E o --extended-regexp</code>	Interpreta los patrones como regex extendidas (EREs)
<code>-F o --fixed-strings</code>	Interpreta los patrones como string fijos, no como regex
<code>-P o --perl-regexp</code>	Interpretar los patrones como expresiones regulares compatibles con Perl (PCREs)

En la implementación GNU grep no hay diferencia funcional entre las sintaxis de expresión regular básica y extendida. **La única diferencia es que en las expresiones regulares básicas los metacaracteres ? + { | () se interpretan como caracteres literales y por tanto se deben escapar con \.**

Flag	Descripción
<code>-e "patron(es)"</code> <code>-o</code> <code>--regexp="patron(es)"</code>	Busca los patrones especificados como argumento y los caracteres <code>\n</code> separan un patrón del siguiente
<code>-f archivo</code> <code>-o</code> <code>--file=archivo</code>	Busca los patrones de un archivo, uno por línea
<code>-i</code> <code>-o</code> <code>--ignore-case</code>	Hace que la regex no sea sensible a mayúsculas o minúsculas. La regex "abc" coincide con el patrón "AbC"
<code>-v</code> <code>-o</code> <code>--invert-match</code>	Invierte el sentido de coincidencia, para seleccionar líneas que no coinciden con el patrón
<code>-w</code> <code>-o</code> <code>--word-regexp</code>	Selecciona solo las líneas que contengan coincidencias que formen palabras completas
<code>-x</code> <code>-o</code> <code>--line-regexp</code>	Seleccione solo los patrones que coincidan exactamente con toda la línea

Opción	Descripción
<code>-c O --count</code>	Se imprime el número de líneas que coinciden con un patrón para cada archivo de entrada. Con la opción <code>-v</code> se cuentan las líneas que no coinciden
<code>--color</code>	Se colorean los patrones, líneas coincidentes o de contexto, nombres de archivo, y números de línea
<code>-L O --files-without-match</code>	Se imprime el nombre de los archivos de entrada que no coinciden con el patrón de búsqueda
<code>-l O --files-with-matches</code>	Se imprime el nombre de los archivos de entrada que coinciden con el patrón de búsqueda
<code>-m num O --max-count=num</code>	La búsqueda del patrón se detiene luego de que se cumpla con el número de coincidencias especificadas en num
<code>-o O --only-matching</code>	Se imprime solo las partes coincidentes con el patrón de búsqueda de las líneas, con cada parte separada por una nueva línea
<code>-q O --quiet</code>	No se imprime nada en la terminal
<code>-s O --no-messages</code>	Suprimir los mensajes de error de los resultados que se imprimen en la terminal

Opclón	Descrlpción
<code>-H o --with- filename</code>	Imprime el nombre de archivo por cada coincidencia. Este es el valor predeterminado cuando hay más de un archivo para buscar
<code>-h o --no- filename</code>	Suprima el prefijo de los nombres de archivo. Este es el valor predeterminado cuando solo hay un archivo para buscar
<code>-n o --line- number</code>	Imprime el número de línea de cada coincidencia antes del patrón
<code>-T o --initial- tab</code>	Se verifica de que el primer carácter del contenido de la línea tiene en una tabulación

Opción	Descripción
<code>-A num O --after- context=num</code>	Imprime el número de líneas que se ubican después de las coincidencias con el patrón de búsqueda
<code>-B num O --before- context=num</code>	Imprime el número de líneas que se ubican antes de las coincidencias con el patrón de búsqueda
<code>-C num O --context=num</code>	Imprime el número de líneas que se ubican antes y después de las coincidencias con el patrón de búsqueda

Opción	Descripción
<code>-r</code> o <code>--recursive</code>	Para cada operación que se aplique al directorio, lea y procese todos los archivos en ese directorio, de forma recursiva
<code>--exclude=glob</code>	Omitir cualquier archivo de línea de comandos con un sufijo de nombre que coincida con el patrón global, utilizando coincidencias con <i>wildcards</i>
<code>--include=glob</code>	Buscar solo archivos de línea de comandos con un sufijo de nombre que coincida con el patrón global, utilizando coincidencias con <i>wildcards</i>

Ejercicio - Patrones

En Jupyter Notebook

Obtén la versión de grep

Obtén el manual de grep

En el fichero poema.txt buscar lo siguiente:

Computadora y computadora, y guardar el resultado en un archivo

Buscar a o e

Buscar la palabra se

Contar las líneas

Contar las interrogaciones ?

Sed

- Este comando nos permite buscar y reemplazar un texto. También permite insertar y eliminar texto de un documento
- **sed 's/texto_a_buscar/texto_a_reemplazar' <archivo_a_reemplazar >archivo_nuevo**
- **sed --help**
- **sed --versión**
- **<https://geekland.eu/uso-del-comando-sed-en-linux-y-unix-con-ejemplos/>**

Ejercicio - Patrones

En Jupyter Notebook

Obtén la versión de sed

Obtén el manual de sed

En el fichero poema.txt buscar lo siguiente:

Reemplaza o por O y crea un fichero de salida

Reemplaza q por Q por entrada de terminal “que es esto”

Reemplaza el contenido sin generar otro fichero

Reemplaza solo en determinadas líneas

Encadena sustituciones

Cut

- Corta partes de líneas de archivos e imprime el resultado
- **cut opciones archivo**
- **cut --versión**
- **cut --help**
- **Las letras acentuadas cuentan como dos caracteres!!!**

Opción	Descripción
-f LIST O --fields=LIST	Se selecciona especificando un campo/columna, conjunto de campos, o rango de campos. También imprime cualquier línea que no contenga un carácter delimitador, a menos que se especifique la opción -s
-b LIST O --bytes=LIST	Se selecciona especificando un byte, conjunto de bytes, o rango de bytes
-c LIST O --characters=LIST	Se selecciona especificando un carácter, conjunto de caracteres, o rango de caracteres
-d O --delimiter	Se especifica un delimitador que se utilizará en lugar del "TAB", que es el predeterminado
--complement	Cuando se usa esta opción, el corte se realiza en todos los bytes, caracteres o campos excepto el/los seleccionados
-s O --only-delimited	De forma predeterminada, cut imprime las líneas que no contienen ningún carácter delimitador. Cuando se usa esta opción, no se imprimen líneas que no contienen delimitadores
--output-delimiter	El comportamiento predeterminado de cut es utilizar el delimitador de entrada como delimitador de salida. Esta opción permite especificar un delimitador de salida diferente

Forma	Descripción
N	El campo, byte, o carácter N, comenzando por 1
N-	Desde el campo, byte, o carácter N, hasta el final de la línea
N-M	Desde el campo, byte, o carácter N al M
-M	Desde el primer campo, byte, o carácter hasta el M

Ejercicio - Patrones

En Jupyter Notebook

Obtén la versión de cut

Obtén el manual de cut

Ficheros grades.txt y students.txt

Combinarlos usando paste

Extraer la segunda columna

Extraer rangos de columnas

Ver como añadir separadores

Sort

- Reordena las líneas de un archivo de forma alfabética o numérica
- **sort opciones archivo**
- **sort --versión**
- **sort --help**

Opción	Descripción
<code>-b o --ignore-leading-blanks</code>	Ignorar los espacios en blanco iniciales
<code>-g o --general-numeric-sort</code>	Comparar según el valor numérico general
<code>-r o --reverse</code>	Invertir el resultado de las comparaciones
<code>-u o --unique</code>	Permite eliminar las líneas repetidas
<code>-c o --check</code>	Compruebe si el archivo de entrada ha sido ordenado
<code>-k POS1[,POS2] o --key=POS1[,POS2]</code>	Ordena considerando campos o columnas específicas. POS toma la forma <code>F[.C]</code> , donde F es el número de campo y C la posición del carácter en el campo; ambos empiezan en 1

Ejercicio - Patrones

En Jupyter Notebook

Obtén la versión de sort

Obtén el manual de sort

Ordenar ficheros

Ver cómo ordenar por columna y eliminar duplicados