

Datos Geoespaciales - 513376

R. Cifuentes Lobos ¹ Franco Retamal ¹ N. Soto Delis ¹

¹Departamento de Geofísica
Universidad de Concepción
Concepción, Chile

Introducción a la programación en Bash

Temario

Pipes (Tuberías)

Descripción del comando.

- Es posible ejecutar varios comandos en una sola línea, esto por medio de “pipes”(“|”).
 - Al usar un pipe, se toma la salida del primer comando para ingresarla en el comando que sigue al pipe.
-
- Sintaxis: *comando 1 | comando 2 | ... | comando "n"*.

sort

Descripción del comando.

- Ordena los contenidos de una columna.
- Admite cadenas de texto y/o números.
- Admite entrada directa de archivo o concatenación por tubería.
- Útil para eliminar duplicados.

sort

- Sintaxis: `sort -opción 1 -opción 2 ... -opción n`

Opciones

- `-ki`: Aplica el comando a la *i*-ésima columna.
- `-n (-g)`: Ordenamiento numérico.
- `-r`: Revierte orden de ordenamiento.
- `-u`: Elimina duplicados.

sed

Descripción del comando.

- Deriva de *stream editor*.
- Realiza transformaciones básicas a cadenas de texto.
- Trabaja línea por línea.
- Admite entrada directa de archivo o concatenación por tubería.

sed

- Sintaxis: `sed -opción 's/string inicial/nueva string/g'`

grep

Descripción del comando.

- Deriva de *global regular expression print*.
- Busca expresiones regulares dentro de un archivo de entrada.
- Acepta cadenas de texto y números.

grep

- Sintaxis: `grep -opción "patrón"`

Opciones

- -n: Muestra el número de fila en la que se localiza la expresión.
- -i: Realiza la búsqueda sin discriminar mayúsculas y minúsculas.
- -r: Realiza la búsqueda recursivamente (en subdirectorios y archivos localizados en subdirectorios).
- -e: Permite la búsqueda de más de un patrón.
- . : Similar a wildcard.
- Para buscar caracteres especiales es necesario utilizar un `\`

Descripción del comando.

- Lenguaje de programación para el procesamiento de datos.
- Integrado con bash puede ser una herramienta muy poderosa.
- Permite el uso de funciones y variables dentro de su sintaxis.
- Permite la concatenación de comandos a través de tuberías.

Nociones básicas.

- `awk` trabaja con columnas de texto.
 - Para acceder a una columna se utiliza el símbolo `$`
 - Las acciones se realizan columna a columna.
 - Por defecto, `awk` diferencia una columna de otra a través de un espacio (Esto se puede cambiar con la opción `-F`).
- Comando `print`:
 - Muestra en la terminal la acción realizada a una columna.
 - Puede mostrar el número de filas (NR) o columnas (NF).

Sintaxis y ejemplos.

■ print

- more texto | awk '{print \$1}'
- more texto | awk '{print 2*\$1}'
- more texto | awk '{print "lat"\$2, "lon"\$3}'
- more texto | awk '{print \$1, \$3, \$2, \$4}'
- more texto | awk '{print \$0}'
- more texto | awk '{print NR}'

Herramientas de formateo

printf

- Permite cambiar el modo de visualización de texto.
- Permite cambiar la cantidad de espacios disponibles para un string.
- Permite cambiar el tipo de texto.

substr

- Permite elegir sólo partes de columnas.

Sintaxis y ejemplos de printf

- `cat catalogo.txt | awk '{printf "%10s\n", $5}'`
- `cat catalogo.txt | awk '{printf "%10.3s\n", $5}'`
- `cat catalogo.txt | awk '{printf "%5i\n", $5}'`
- `head | awk 'printf "%12.3e\n", $5'`

Sintaxis y ejemplos de substr

- `substr($i,y,z)`: Columna i-ésima, desde el carácter y-ésimo, z caracteres
- `cat catalogo.txt | awk '{print $3, substr($3,1,2)}'`

Operaciones matemáticas.

- Se puede realizar operaciones a columnas y entre columnas.
- Se utilizan los mismos operadores matemáticos que en bash, python, etc.
- `BEGIN{sum=0}{sum+= i }END{print sum}`

if

- Sintaxis: ... | awk '{if (sentencia) acción}'
- ||: o lógico &&: y lógico
- Ejemplos:
 - cat catalogo.txt | awk '{if (\$9=="M,") print \$9}'
 - cat catalogo.txt | awk '{if (\$10==5) print \$0}'

Ejercicios

■ Archivo catalogo.txt

- Este archivo contiene un catálogo de la sismicidad global, de la red global de IRIS, para las primeras semanas después del terremoto magnitud 8.8 en Chile el 2010. Las columnas son (catálogo, día, hora (UTC), latitud, longitud, profundidad, código de la región, código de la zona sísmica, tipo de magnitud, magnitud).

- Al final de cada columna hay una coma, cómo puede eliminar esa coma?
- Liste las 5 magnitudes más bajas del catálogo.
- Liste sólo los eventos del catálogo QED/NEIC
- Dónde fue el terremoto más profundo en el catálogo global?
- Cuál es el número de réplicas en las regiones 134, 135 el día 28 feb 2010?
- Calcule el promedio de las magnitudes de los terremotos del catálogo FINGER/NEIC.

Ejercicios

- Archivo `Boya_Meteo_10Ago.csv`
- El siguiente archivo tiene una extensión `csv` (comma separated value), en que las columnas están separadas por comas en lugar de espacios. Esto puede inicialmente ser un problema para su tratamiento en `awk`.
 - Despliegue el archivo en la terminal y describa qué variables contiene.
 - Busque cómo pueden ser listadas columnas específicas en este tipo de texto con la opción `-F` de `awk`.
 - Cómo puede solucionar el problema de las comas con el comando `sed`?
 - Cambie los `NaN` del archivo por `0`
 - Muestre las cifras en notación científica.