

UNIVERSIDAD DE SONORA

Limpieza y preparación de datos usando Emacs

Autor:
Raúl MONTES

Profesor:
Dr. Carlos LIZÁRRAGA

7 de febrero de 2017



Resumen

El trabajo consistió en la preparación y limpieza de datos recolectados de la red de sondeos de norteamérica. Para la actividad usamos un script que descargamos de la plataforma, el cual contenía las instrucciones de descargar los datos de sondeo de una estación a elegir, con ayuda de Emacs se editó el código para cambiar el lugar y la fecha de donde se requerían los datos. Lo siguiente fue usar la terminal para crear un documento de texto con todos los datos recolectados y poder manejarlos más fácilmente, después con ayuda de Gnuplot se hicieron dos gráficas, una temperatura-altura y otra de presión-altura usando los datos del 1 de febrero de 2017 a las 12Z. Por último se filtraron los datos del año 2016 con ayuda de la terminal para determinar cuántos días al mes se habían hecho mediciones y a que horas habían sido realizadas para posteriormente elaborar una tabla con los datos.

Introduccion

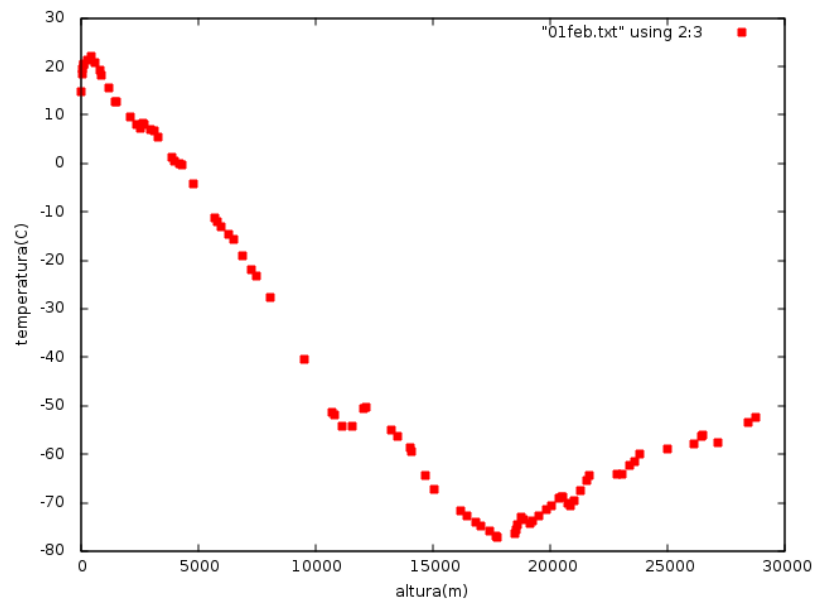
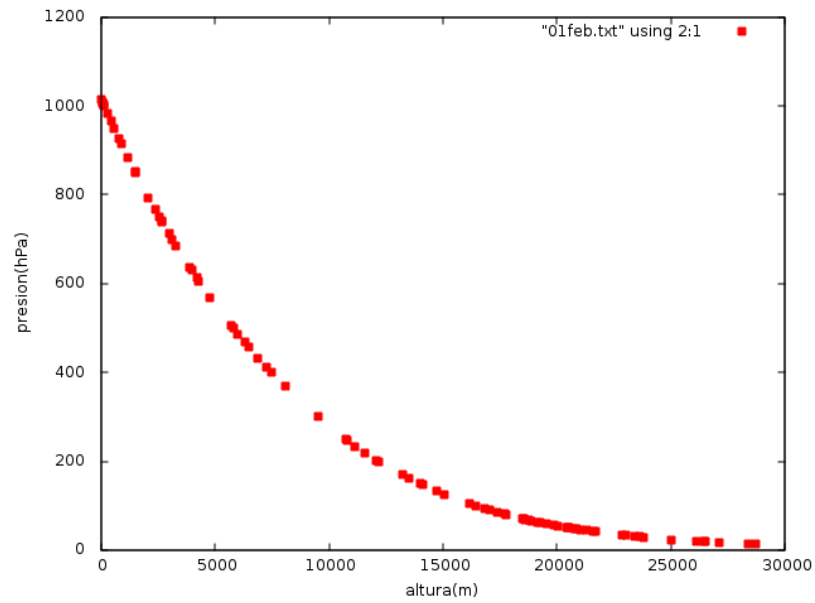
A Veces nos parece muy fácil manejar datos usando por ejemplo copy-paste, todo el tiempo juntamos datos de esta manera para hacer un mejor manejo de ellos, pero, debemos preguntarnos si es eficiente hacer esto cuando se tiene una cantidad realmente grande de datos, la respuesta es no. En esta actividad veremos que hay métodos muy eficientes para el manejo, preparación y limpieza de datos. No es un procedimiento muy complicado pero es necesario un conocimiento básico del entorno de la terminal de Linux, gráficas en gnuplot y editar textos en Emacs, ya que fueron los recursos usados para poder llevar a cabo la práctica.

obtención de datos

Los datos empleados provienen de un sitio web perteneciente a la universidad de Wyoming, el cual contiene información acerca de la atmosfera de diferentes lugares. En el sitio se pueden obtener los datos según la región, estación, y día del año. Para realizar la práctica fue necesario primero elegir una estación atmosférica disponible en el sitio web. Después se procedió a editar un script proporcionado por el profesor, mediante Emacs se cambiaron los datos de la fecha y la estación para posteriormente correr el script en la terminal de Linux.

gráficas

Con los datos disponibles en la página del 1 de febrero de 2017 a las 12Z horas se creó un archivo en Emacs y con la terminal usando gnuplot se crearon dos gráficas, una de presión-altura y otra de temperatura-altura.



observaciones del 2016

Una vez descargados los archivos con ayuda del script, lo siguiente fue unirlos en un solo documento de texto. Después con ayuda de la terminal, se filtraron los datos según mes y hora, con lo que se construyó la siguiente tabla.

mes	00Z	12Z
Enero	0	27
Febrero	0	3
Marzo	0	0
Abril	0	0
Mayo	0	0
Junio	0	0
Julio	0	11
Agosto	0	29
Septiembre	1	23
Octubre	1	7
Noviembre	0	0
Diciembre	7	8

Table 1: observaciones realizadas a las 00Z y 12Z horas.

Preparación de datos

Para preparar los datos y poder hacer correcto uso de la información fue necesario estar familiarizado con el entorno de Emacs, Gnuplot y Linux. Fueron de gran utilidad funciones como "queri replace" para modificar el script así como ctrl-x y ctrl-s para salvarlo. Al graficar fue necesario conocer el entorno de gnuplot para poder ajustar los encabezados e identificar las columnas a graficar. Probablemente lo más importante en esta actividad fue el uso de los comandos "grep" y "wc" para filtrar datos y mostrar los datos disponibles en pantalla, fue así como pudieron analizarse las observaciones hechas en 2016 para posteriormente vaciarse la tabla 1.

Texto usado en Emacs(sin modificar)

```
# Descarga por mes. Cambiar año de consulta. Ajustar el numero de estacion.
#!/bin/bash
# Despues de editar: chmod 755 script1.sh
# Para ejecutar: ./script1.sh

IFS=":"
LISTM31="01:03:05:07:08:10:12"
#LISTM31="01:03:05:07"
LISTM30="04:06:09:11"
#LISTM30="04:06"
LISTM28="02"

# Script para bajar info por mes. Año 2016, dentro del URL:  YEAR=2015
# Months 31 days

for i in $LISTM31 ; do
    /usr/bin/wget "http://weather.uwyo.edu/cgi-bin/sounding?region=naconf&TYPE=TEXT%3
    /bin/sleep 5

done
# Months 30 days

for i in $LISTM30 ; do
    /usr/bin/wget "http://weather.uwyo.edu/cgi-bin/sounding?region=naconf&TYPE=TEXT%3
    /bin/sleep 5

done
# Feb. 28 days

for i in $LISTM28 ; do
    /usr/bin/wget "http://weather.uwyo.edu/cgi-bin/sounding?region=naconf&TYPE=TEXT%3
    /bin/sleep 5

done
```

Bibliografía

- [1] , WWW.WEATHER.UWYO.EDU *Empalme, Son. Observations*, 2017.
- [2] , PROGFORTTRAN.PBWORKS.COM *01Linux*, 2017
- [3] , WWW.MINISCONLATEX.BLOGSPOT.MX *Tablas con latex*, 2017.