

Proyecto 2: Cálculo de Frecuencias

Martes 2 de Octubre

I. DESCRIPCIÓN GENERAL

Este programa acumulará las frecuencias de cada carácter ASCII que se encuentren en todos los archivos de una lista dada. Estos datos se guardan en un archivo cuyo formato será definido por cada grupo. Es muy importante el manejo de argumentos de la línea de comando para este proyecto. Toda la programación debe hacerse en Lenguaje C sobre Linux.

II. FRECUENCIA DE CARÁCTERES ASCII

El código ASCII extendido usa 8 bits (un byte) para cada alfanumérico y códigos de control representados. Esto da 256 posibilidades. Independiente de la naturaleza de cualquier archivo, este se puede entender como una hilera de bytes.

Nuestro objetivo es contar cuantas veces aparece cada uno de los 256 posibles caracteres, acumulando estas cuentas en una tabla que al final será grabada en un archivo de salida indicado por el usuario. El formato interno de dicho archivo será definido por cada grupo (debe incluir los caracteres con frecuencia 0).

III. ARGUMENTOS

Su programa (identificado como `<prog>`) se invocará de dos posibles maneras:

- `<prog> <file-frec> <file-list>`:
Si el archivo de frecuencias `<file-frec>` no existe, este debe ser creado, pero si existiera debe ser cargado primero hacia una tabla de 256 entradas. Después se procesarán todos los archivos mencionados en `<file-list>` y se acumularán las estadísticas en la tabla mencionada previamente, para que al final se grabe la nueva información en `<file-frec>`. Al final se desplegará en modo texto el contenido del archivo `<file-frec>` (detalles más adelante).
- `<prog> <file-frec>`:
Ya que no hay lista de archivos a procesar, en este caso sólo se desplegará el contenido del archivo `<file-frec>` en la forma descrita más adelante. Si `<file-frec>` no existe se da un mensaje apropiado.

Ejemplos:

- `<prog> W B C D`
- `<prog> X *.c`
- `<prog> Y B *.c *.tex`
- `<prog> Z *`
- `<prog> W`

IV. SALIDA

Al terminar de procesar la lista de archivos que se le haya dado se desplegará el contenido actual del archivo de frecuencias en consola en modo texto. Sin embargo, se espera que el despliegue esté formateado apropiadamente con calidad profesional. Las columnas deben estar etiquetadas apropiadamente. Se mostrarán las 256 frecuencias.

En cada fila habrá 3 columnas que serán:

1. Código hexadecimal del carácter.
2. Símbolo (sólo si es desplegable, en caso contrario se deja en blanco).
3. Frecuencia acumulada.

V. REQUISITOS INDISPENSABLES

La ausencia de uno solo de los siguientes requisitos vuelve al proyecto “no revisable” y recibe un 0 de calificación inmediata:

- La colaboración entre grupos se considera fraude académico.
- Todo el código debe estar escrito en C (no C++).
- La única interacción válida con el programa es por medio de argumentos de línea de comando.
- El proyecto debe compilar y ejecutar en Linux. Todo debe estar **integrado**, explicaciones del tipo “*todo está bien pero no pudimos pegarlo*”¹ provocan la cancelación automática de la revisión.
- La presentación debe ser de mucha calidad.
- No debe dar “Segmentation Fault” bajo ninguna circunstancia.
- Hacer la demostración en una máquina que levante Linux de manera real (puede ser dual), es decir no usar máquinas virtuales.

VI. FECHA DE ENTREGA

Revisiones a las 11:30 am el **Martes 2 de Octubre** en la oficina del profesor. Mande además un `.tgz` con todo lo necesario (fuentes, `makefile`, `readme`, etc.) a `torresrojas.cursos.05@gmail.com`. Ponga como subject: [AA] - Proyecto 2 - Fulano - Mengano - Sutano, donde Fulano, Mengano y Sutano son los miembros del grupo.

¹esto incluye los supuestos casos cuando alguien del grupo de trabajo no hizo su parte – el profesor no está interesado en sus problemas de organización.