

Trabajo en clase 4

(INFO189 - Sistemas Operativos)

Modificar el programa ya realizado, los requisitos que debe cumplir son:

- 1) agregar nueva variables de configuración al archivo .env de variables de ambiente, las variables deben ser: EXTENSION, PATH_FILES_IN, PATH_FILES_OUT, AMOUNT_THREADS, INVERTED_INDEX_FILE

```
EXTENSION=txt
PATH_FILES_IN=/home/lvc/in
PATH_FILES_OUT=/home/home/lvc/out
AMOUNT_THREADS=2
INVERTED_INDEX_FILE=/home/lvc/file.idx
```

donde:

EXTENSION, el tipo de archivos a leer, debe hacer un filtro que solo permita leer este tipo de archivos

PATH_FILES_IN, representa la carpeta donde se encuentran N archivos de texto sobre los cuales debe aplicar conteo de palabras

PATH_FILES_OUT, representa la carpeta de destino, donde por cada archivo sobre el cual realiza conteo de palabras debe establecer un archivo de salida con el mismo nombre

AMOUNT_THREADS, representa la cantidad de threads con los que trabajará su sistema y donde cada thread debe ejecutar un conteo de palabras de un archivo en específico

INVERTED_INDEX_FILE, archivo de salida final del índice invertido

- 2) Debe agregar una nueva opción de menú llamada "PREPARAR DATOS PARA ÍNDICE INVERTIDO". Debe realizar un proceso externo (en c++) que reciba como argumentos de ingreso: EXTENSION, PATH_FILES_IN, PATH_FILES_OUT y AMOUNT_THREADS

la llamada a este proceso externo debe ser así:

```
./app EXTENSION PATH_FILES_IN PATH_FILES_OUT AMOUNT_THREADS
```

Ejemplo:

```
./app "txt" "/home/lvc/in" "/home/lvc/out" "2"
```

- 3) Los path PATH_FILES_IN y PATH_FILES_OUT no pueden ser iguales
- 4) El proceso externo debe utilizar threads según se indique en la variable AMOUNT_THREADS
- 5) La variable AMOUNT_THREADS como máximo puede ser 10.
- 6) El proceso externo deberá realizar el conteo de palabras de todos los archivos con la extensión EXTENSION, los archivos a procesar estarán en el path PATH_FILES_IN,
- 7) Debe considerar al menos 20 archivos de entrada a procesar
- 8) Los nombres de los archivos de entrada deben ser file001, file002, file003,
- 9) El resultado del conteo de palabras debería quedar almacenado en PATH_FILES_OUT, el nombre del archivo debe tener el mismo nombre del archivo de entrada
- 10) El formato del resultado de salida debe ser:

palabra1, cantidad
palabra2, cantidad
palabra3, cantidad

.
.
.
.

- 11) Debe imprimir el resultado de procesamiento de cada thread con el siguiente formato

proceso pid=PID - archivo procesado ARCHIVO-ENTRADA por el thread=THREAD

Ejemplo:

proceso pid=101 - archivo procesado /home/lvc/in/file001.txt por el thread=0

proceso pid=101 - archivo procesado /home/lvc/in/file002.txt por el thread=1

proceso pid=101 - archivo procesado /home/lvc/in/file003.txt por el thread=0

.
.
.

- 12) Debe agregar nueva opción de menú "CREAR ÍNDICE INVERTIDO", esta opción debe gatillar un proceso externo (en c++), el cual reciba como argumentos de ingreso los valores de las variables INVERTED_INDEX_FILE y PATH_FILES_OUT

Ejemplo ./app "/home/lvc/file.idx" "/home/lvc/out"

13) Esta opción de menú solo debe poder ejecutarse si se ejecutó previamente la opción “PREPARAR DATOS PARA ÍNDICE INVERTIDO”

14) Este proceso externo debe realizar un totalizado de todos los archivos generados por “PREPARAR DATOS PARA ÍNDICE INVERTIDO”, el formato a obtener es:

```
palabra1:(archivo-salida1,cantidad);(archivo-salida2,cantidad);.....
palabra2:(archivo-salida3,cantidad);(archivo-salida5,cantidad);.....
palabra3:(archivo-salida1,cantidad);(archivo-salida3,cantidad);.....
.
.
.
.
.
```

Donde:

palabra1: palabra encontrada
en los () debe fijar el nombre del archivo donde encontró la palabra y la cantidad de veces que la encontró, debe recorrer todos los archivos presentes en la ruta PATH_FILES_OUT

Ejemplo:

```
hola:(file001.txt;3);(file004.txt;13);(file006.txt;31);(file007.txt;30)
casa:(file001.txt;4);(file014.txt;133);(file006.txt;131);(file017.txt;80);(file030.txt;10)
perro:(file001.txt;35);(file002.txt;3)
pc:(file005.txt;3)
celular:(file0014.txt;300);(file0015.txt;3);(file0016.txt;1);
.
.
.
.
.
```

15) Cuando el proceso externo finalice debe imprimir por pantalla

El proceso pid=PID generó el archivo=ARCHIVO

Ejemplo:

El proceso pid=675 generó el archivo=/home/lvc/file.idx