

Proyecto 2: Programación Concurrente con Hilos y Procesos (15%)

I. Objetivos

- Utilizar las primitivas de hilos (*threads*) y procesos para el manejo de la concurrencia.
- Desarrollar un mecanismo de comunicación de memoria compartida para la comunicación entre los hilos (*mutex* y semáforos) y procesos (*pipes* y señales)
- Manejar primitivas que permitan medir el tiempo de ejecución de los procesos/hilos.
- Manejar primitivas para el manejo de archivos binarios y texto.

II. Enunciado

Ud. deberá implantar en C dos versiones (una usando procesos y otra usando hilos) de un programa que analiza información del operativo Censos 2011 de Uruguay. El territorio uruguayo se divide en *Departamentos*, dentro de los cuales las poblaciones se denominan *Localidades*. La información recolectada en el operativo consiste en un archivo por Departamento, que contiene la cantidad de hombres y mujeres por área y localidad, según edades por decenas (de 0 a 9, de 10 a 19, ..., 80 a 89, 90 o más).

II.1 Formato de Archivo

Los archivos se encuentran en formato CSV (*Comma-Separated Variables*), que implica la utilización de comas para separar los valores de cada fila. La primera fila contiene los nombres de las variables, y todas las filas tienen las variables en ese orden. Así, por ejemplo, si un archivo contiene:

```
Localidad y Sexo,Total,0 a 9,10 a 19,20 a 29,30 a 39,40 a 49,50 a 59,60 a 69,70 a 79,80 a 89,90 o más
Alejandro Gallinal,1357,235,264,170,172,163,131,100,86,32,4
Hombres,687,126,138,85,79,84,68,50,43,12,2
Mujeres,670,109,126,85,93,79,63,50,43,20,2
Casupá,2402,285,382,260,298,277,270,241,237,128,24
Hombres,1136,135,192,129,143,134,129,118,96,52,8
Mujeres,1266,150,190,131,155,143,141,123,141,76,16
Florida,33639,4864,5512,4410,4550,4039,3646,2979,2235,1193,211
Hombres,15997,2519,2774,2193,2209,1874,1710,1346,909,401,62
Mujeres,17642,2345,2738,2217,2341,2165,1936,1633,1326,792,149
```

Se puede leer como la siguiente tabla:

Localidad y Sexo	Total	0 a 9	10 a 19	20 a 29	30 a 39	40 a 49	50 a 59	60 a 69	70 a 79	80 a 89	90 o más
Alejandro Gallinal	1357	235	264	170	172	163	131	100	86	32	4
Hombres	687	126	138	85	79	84	68	50	43	12	2
Mujeres	670	109	126	85	93	79	63	50	43	20	2
Casupá	2402	285	382	260	298	277	270	241	237	128	24
Hombres	1136	135	192	129	143	134	129	118	96	52	8
Mujeres	1266	150	190	131	155	143	141	123	141	76	16
Florida	33639	4864	5512	4410	4550	4039	3646	2979	2235	1193	211
Hombres	15997	2519	2774	2193	2209	1874	1710	1346	909	401	62
Mujeres	17642	2345	2738	2217	2341	2165	1936	1633	1326	792	149

Esto indica, por ejemplo, que el número de hombres en la Localidad de Casupá entre 20 y 29 años de edad fue de 129 hombres al momento de realizarse el censo.

Observe que la información de cada localidad ocupa tres líneas. Así, por cada grupo de tres líneas:

- la primera línea muestra el número total de personas (hombres y mujeres) de cada grupo etario para esa Localidad (**líneas en azul de la tabla anterior**)
- la segunda mostrando el número de hombres
- la tercera mostrando el número de mujeres

Todos los archivos poseen las mismas 12 variables. Los datos son terminados por una fila en blanco.

II.2 Requerimientos del programa

Se desea obtener el promedio de edad de hombres y de mujeres de todo el país. Para eso, deberá crear dos versiones de su programa: una con procesos y otra con hilos.

En cada versión del programa, el proceso/hilo principal creará tantos procesos/hilos como se indique en la orden de ejecución, al estilo de abanico. Cada proceso/hilo hijo, se encargará de procesar un archivo por vez: totalizará la cantidad de hombres y mujeres, calculará la edad promedio de los hombres y la edad promedio de las mujeres, pasará dicha información al proceso/hilo padre (esta información es por Departamento). Al terminar de procesar un archivo, el procesos/hilo estará listo para procesar otro archivo, si aún quedan por procesar. Una vez que estén todos los archivos procesados, el proceso/hilo padre, calculará a partir de los resultados parciales de sus procesos/hilos hijos el total de hombres, total de mujeres y edad promedio de ambos, para todo el país, e imprime los resultados por pantalla. La línea de orden de ejecución se realizará de la siguiente forma:

\$ ***censo-p*** *NumProcesos ArchivoListaDepartamentos* (para la versión usando procesos)

\$ ***censo-t*** *NumHilos ArchivoListaDepartamentos* (para la versión usando hilos)

Donde:

NumProcesos/NumHilos: es la cantidad de procesos e hilos que se crearán en cada versión respectiva.

ArchivoListaDepartamentos: es el nombre del archivo que contiene los nombres de los archivos que representan cada Departamento. Cada nombre de archivo estará en una línea separada. Un ejemplo del contenido de este archivo se muestra a continuación:

Artigas
CerroLargo
Florida
Lavalleja
Montevideo

Deberá agregar la extensión “.CSV” al nombre dado.

III. Salida del Programa

La salida del programa la producirá el proceso/hilo padre y consiste únicamente en imprimir los totales globales: número de hombre, número de mujeres, edad promedio de hombres, edad promedio de mujeres..

IV. Detalles de Implementación

- Para la implementación con procesos:

Inicialmente el proceso principal crea todos sus procesos hijos, e indica a cada hijo el nombre del archivo que debe procesar. Cuando haya asignado trabajo a todos sus hijos, esperará por el primero que le envíe información.

Cuando un hijo se libere, le asignará el próximo archivo pendiente por procesar y espera nuevos

resultados parciales de los hijos. Así, hasta que todos los archivos se hayan procesado y todo los hijos hayan terminado. En este último caso, enviará una señal a sus hijos para que terminen.

Cada proceso hijo, al ser creado, esperará por el nombre del archivo que debe procesar. Procesa el archivo, envía la información al padre y espera por un próximo archivo a procesar o por la señal de terminación.

La comunicación entre proceso padre e hijos se podrá realizar a través de pipes no nominales o nominales, pero sólo puede escribir en ellos un proceso a la vez.

- Para la implementación con hilos:

La implementación con hilos se diferencia principalmente en la forma de comunicación que debe hacerse a través de pase de parámetros al momento de crear los hilos y de estructuras de datos compartidas entre los hilos, haciendo uso eficiente de la memoria. Debe evitar que dos hilos escriban en el mismo lugar al mismo tiempo, en particular para tomar el siguiente archivo a procesar.

La otra diferencia es en la forma en que los hilos terminan: no será a través de señales. Si un hilo quiere procesar otro archivo y ya no hay disponibles, entonces termina.

Para ambas implementaciones, deberá tomar el tiempo total de ejecución. En

http://www ldc.usb.ve/~yudith/docencia/ci-3825-taller/Ejemplo_Tiempo.c encontrará un ejemplo del uso de rutinas para tomar tiempos.

IV.2 Cálculo de la edad promedio

Para obtener la edad promedio, debe calcularse la media ponderada de las edades de la siguiente forma

$$\frac{\sum_E (N_E \times m(E))}{\sum_E (N_E)}$$

Donde

- E es el conjunto de grupos etarios
- N es el número de personas en ese grupo etario
- $m(E)$ es el punto medio del grupo etario, calculado usando $(\text{límite inferior} + \text{límite superior}) \div 2$

Así, si hay 1000 personas entre 0 y 9, 2000 entre 10 y 19; la edad promedio es

$$\frac{1000 \times \left(\frac{0+9}{2}\right) + 2000 \times \left(\frac{10+19}{2}\right)}{1000 + 2000} = \frac{1000 \times 4,5 + 2000 \times 14,5}{3000} = 11,17$$

V. Observaciones

- El archivo de entrada y los archivos de los Departamentos se suponen correctos.
- Deben escribir un makefile para la compilación óptima.
- Deben estructurar modularmente el programa usando archivos .h y manejar correctamente los posibles errores de las llamadas al sistema.
- **Recuerde validar las llamadas al sistema con perror() y verificar los argumentos de entrada.**

VI. Entrega

La entrega del proyecto está pautada para el martes de la semana 12. Ese día Ud. deberá:

- Subir a Aula Virtual un archivo tar.gz que contenga únicamente los fuentes (archivos .c y .h) de su proyecto y un archivo Makefile para generar los ejecutables.

- Un informe de no más de 3 páginas explicando
 - el manejo de la memoria compartida
 - parámetros de llamada a los procesos/hilos
 - comunicación de procesos/hilos
 - Comparación de la versión de hilos con la de procesos en función a la memoria utilizada, tiempo de ejecución, facilidad de programación, etc. Agregar al informe una tabla de tiempo, para comparar las dos versiones del proyecto con 2, 4, 10, 20 hilos/procesos. Especifique la arquitectura que usó para realizar las ejecuciones (Intel dual core, dual processor, etc.).

VII. Evaluación

Se asignarán:

- 3 puntos por su informe
- 3 puntos por la creación de hilos y comunicación entre ellos
- 3 puntos por la creación de procesos y comunicación entre ellos
- 2 puntos por su estrategia de exclusión mutua y su implementación adecuada
- 2 puntos por ejecución correcta y completa
- 2 puntos por seguir convenciones de C en UNIX (incluyendo makefile)