



Universidad
Francisco de
Vitoria

UFV Madrid

Práctica de asignatura 1

Hilos, Procesos y Ficheros

Enunciado

Acabáis de ser contratados por una gran consultora para trabajar en su departamento de Big Data. Se espera que ayudéis a solucionar un problema que la compañía tiene con uno de sus procesos de información, el cual les está bloqueando la entrega de un proyecto.

La base del Big Data está en la capacidad de poder realizar pequeñas operaciones atómicas a larga escala. Estas operaciones permiten procesar largos ficheros de datos con el fin de encontrar información útil para la realización de análisis más detallados. Para eso, es necesario poder ejecutar de forma paralela y sobre ficheros de grandes dimensiones operaciones de comparación, suma o resta, entre otras.

Vuestro trabajo es:

- 1) Descargar los datos de la siguiente URL:
<http://files.grouplens.org/datasets/movielens/ml-latest-small.zip>. Aquí encontraréis un fichero llamado **movies.csv**, el cual contiene datos de un catálogo de película. El fichero es un dataset con la categorización y reviews de 3.600 películas. Los géneros de películas listados pueden ser:

- | | |
|---------------|-------------|
| • Action | • Film-Noir |
| • Adventure | • Horror |
| • Animation | • Musical |
| • Children's | • Mystery |
| • Comedy | • Romance |
| • Crime | • Sci-Fi |
| • Documentary | • Thriller |
| • Drama | • War |
| • Fantasy | • Western |

- 2) Crear un programa que sea capaz de contar las películas de cada género que se encuentran en el fichero (emplead hilos independientes por categoría).
- 3) Guardar los resultados finales en un fichero designado **ResumenPelícula.txt** (Para garantizar la integridad de los datos, la escritura en el fichero resumen deberá ser protegida con **Mutex** o con **Semáforos**).
- 4) Para optimizar el proceso, se debe medir el tiempo que le ha tomado a cada hilo llevar a cabo el conteo.

Características de la práctica

- a) El estudiante deber diseñar y desarrollar, en C y sobre UNIX (Linux), un programa que cumpla con las tareas anteriores (1, 2, 3 y 4).
- b) Se debe evitar el uso de sentencia `goto`, `break`, `continue` y `exit` dentro de los bucles.
- c) El código debe compilar sin errores y funcionar con la menor configuración local posible.



Universidad
Francisco de
Vitoria

UFV Madrid

Práctica de asignatura 1

Hilos, Procesos y Ficheros

- d) El código debe ir correctamente documentado y/o comentado, explicando qué hace cada función desarrollada.
- e) La solución debe estar optimizada, sin redundancia de código, bucles innecesarios o condiciones mal definidas (if repetidos por todo el código).
- f) Se valora el uso de una interfaz que le indique al usuario de qué trata la aplicación.
- g) Se espera que ponga en práctica los temas Procesos, hilos, concurrencias y semáforos.

Documentación a entregar

- a) Memoria de la práctica en PDF.
- b) Fuentes del programa desarrollador.

Bibliografía

- El lenguaje de programación C: diseño e implementación de programas. Félix García, Jesús Carretero, Javier Fernández y Alejandro Calderón. Prentice-Hall, 2002.
- The UNIX System S.R. Bourne Addison-Wesley, 1983.
- Advanced UNIX Programming M.J. Rochkind Prentice-Hall, 1985.
- Sistemas Operativos: Una visión aplicada Jesús Carretero, Félix García, Pedro de Miguel y Fernando Pérez. McGraw-Hill, 2001.