

Trabajo Práctico: *Threading*

Sistemas Operativos - Primer cuatrimestre de 2021

Fecha límite de entrega: viernes 28 de mayo de 2021 a las 23:59

1. Introducción

En este trabajo práctico buscaremos profundizar en una de las facetas principales que aparece al estudiar los sistemas operativos: la gestión de la concurrencia. Practicaremos cómo razonar sobre la ejecución concurrente de programas y las técnicas para gestionar la contención sobre los recursos y evitar que se produzcan condiciones de carrera.

Nos centraremos especialmente en el uso de *threads*, una herramienta provista por los sistemas operativos que nos permite disponer de varios hilos de ejecución concurrentes dentro de un mismo programa. En particular, emplearemos la interfaz para *threads* que forma parte de la librería estándar de C++.

Realizaremos la implementación de una estructura de datos que será denominada **HashMapConcurrente**. Se trata de una tabla de *hash* abierta, que gestiona las colisiones usando listas enlazadas. Su interfaz de uso es la de un *map* o diccionario, cuyas claves serán *strings* y sus valores, enteros no negativos. La idea es poder aplicar esta estructura para procesar archivos de texto contabilizando la cantidad de apariciones de palabras (las claves serán las palabras y los valores, su cantidad de apariciones).

2. Pautas generales

La cátedra les brinda un esqueleto de código sobre el cual trabajar, que será descripto a lo largo de los ejercicios. El mismo incluye un Makefile que podrán usar para compilar el proyecto. También se les proporciona un conjunto de *tests unitarios* para probar los métodos que deben implementar;¹ estos tests se brindan **únicamente** a modo de guía y **no** evalúan aspectos de concurrencia, por lo que podrían no detectar condiciones de carrera u otros problemas de sus implementaciones. Bajo **ningún** punto de vista estos tests constituyen una garantía de que sus soluciones sean correctas.

El código que escriban deberá seguir buenas prácticas de programación, como usar nombres descriptivos para variables y funciones e incluir comentarios donde consideren necesario realizar aclaraciones. Es fundamental que el mismo esté libre de condiciones de carrera y *deadlocks*.

Se espera que el código que entreguen esté acompañado de un informe que describa el proceso de resolución del trabajo práctico. El informe deberá ser claro y prolijo, y lo más conciso posible —evaluaremos la calidad de su contenido, por sobre su cantidad—.

¹Pueden compilar y ejecutar los tests con el comando `make test`.

En cada punto del enunciado se mencionan algunas cuestiones que deben ser abordadas en el informe. Si en el texto del informe incluyen referencias al código, les pedimos, por favor, que las mismas sean claras.

3. Ejercicios

1. En el archivo `ListaAtomica.hpp` se brinda una implementación parcial de una lista enlazada que utilizaremos para almacenar los elementos de cada *bucket* de la tabla de *hash*. Se pide:
 - EN EL CÓDIGO, completen la implementación del método `insertar(T valor)`, que agrega un nuevo nodo al comienzo de la lista enlazada. La operación de inserción debe ser atómica.
 - EN EL INFORME, respondan brevemente y justifiquen: ¿Qué significa que la lista sea atómica? Si un programa utiliza esta lista atómica ¿queda protegido de incurrir en condiciones de carrera? ¿Cómo hace su implementación de `insertar` para cumplir la propiedad de atomicidad?
2. En el archivo `HashMapConcurrente.cpp` se brinda una implementación parcial de la clase `HashMapConcurrente`. Por simplicidad usaremos, a modo de pseudo-función de *hash*, la primera letra de cada palabra (ver el método `hashIndex`). Por lo tanto, nuestra tabla de *hash* tendrá 26 *buckets*, uno para cada letra del abecedario (sin contar la ñ), cada uno de ellos representado por una `ListaAtomica`. Los *strings* utilizados como clave solo constarán de caracteres en el rango a-z (es decir, no habrá strings con mayúsculas, números o signos de puntuación). Se pide:
 - EN EL CÓDIGO, completen las implementaciones de:
 - a) `void incrementar(string clave)`. Si `clave` existe en la tabla (en la lista del bucket correspondiente), se debe incrementar su valor en uno. Si no existe, se debe crear el par `<clave, 1>` y agregarlo a la tabla. Se debe garantizar que solo haya contención en caso de colisión de *hash*; es decir, si dos o más *threads* intentan incrementar concurrentemente claves que no colisionan, deben poder hacerlo sin inconvenientes.
 - b) `vector<string> claves()`. Devuelve todas las claves existentes en la tabla. Esta operación debe ser no bloqueante y libre de espera (*wait-free*) no hace falta garantizar el correcto funcionamiento en caso de que se ejecute incrementar en simultáneo con claves.
 - c) `unsigned int valor(string clave)`. Devuelve el valor de clave en la tabla, o 0 si la clave no existe en la tabla. Esta operación debe ser no bloqueante y libre de espera (*wait-free*), no hace falta garantizar que el resultado sea consistente si se ejecuta en simultaneo con incrementar.

Las tres operaciones deben estar libres de condiciones de carrera y de *deadlock*. Con este fin, podrán agregar a la clase `HashMapConcurrente` las primitivas de sincronización que consideren necesarias, modificando si es preciso el archivo `HashMapConcurrente.hpp`.

- EN EL INFORME, expliquen cómo lograron que la implementación esté libre de condiciones de carrera, justificando sus elecciones de primitivas de sincronización. Expliquen también qué decisiones debieron tomar para evitar generar más contención o espera que las permitidas.

3. La clase `HashMapConcurrente` tiene implementado el método `maximo()`, que devuelve el par `<clave, valor>` con el valor más alto de la tabla. Se pide:

■ EN EL CÓDIGO:

- a) La implementación provista de `maximo` puede ejecutarse concurrentemente con `incrementar`, lo cual podría generar inconsistencias en los resultados de la ejecución. Modifiquen la implementación para que esto ya no pueda suceder.
- b) Implementen el método `pair<string, unsigned int> maximoParalelo(unsigned int cantThreads)`, que realiza lo mismo que `maximo` pero repartiendo el trabajo entre la cantidad de *threads* indicada. Cada *thread* procesará una fila de la tabla a la vez. Se debe maximizar la concurrencia: en ningún momento debe haber *threads* inactivos, y los mismos deben terminar su ejecución si y solo si ya no quedan filas por procesar.

■ EN EL INFORME:

- a) ¿Qué problemas puede ocasionar que `maximo` e `incrementar` se ejecuten concurrentemente? Piensen un escenario de ejecución en que `maximo` devuelva un resultado que no fue, en ningún momento, el máximo de la tabla.
- b) Describan brevemente su implementación de `maximoParalelo`. ¿Cuál fue la estrategia elegida para repartir el trabajo entre los *threads*? ¿Qué recursos son compartidos por los *threads*? ¿Cómo hicieron para proteger estos recursos de condiciones de carrera?

4. En el archivo `CargarArchivos.cpp`, se brinda una implementación parcial de la función `cargarArchivo(HashMapConcurrente hashMap, string filePath)`. La misma lee el archivo indicado mediante `filePath` y carga todas sus palabras en `hashMap`. Se pide:

■ EN EL CÓDIGO:

- a) Completar la implementación de `cargarArchivo`.
- b) Implementar la función `void cargarMultiplesArchivos(HashMapConcurrente hashMap, unsigned int cantThreads, vector<string> filePaths)`. La misma carga todos los archivos indicados por parámetro, repartiendo el trabajo entre `cantThreads` *threads*. Cada *thread* deberá ocuparse de un archivo a la vez, teniendo en cuenta las mismas consideraciones mencionadas para `maximoParalelo`.

■ EN EL INFORME:

- a) ¿Fue necesario tomar algún recaudo especial, desde el punto de vista de la sincronización, al implementar `cargarArchivo`? ¿Por qué?
- b) Describan brevemente las decisiones tomadas para la implementación de `cargarMultiplesArchivos`.

5. En este punto, buscaremos evaluar qué ventajas ofrece, en términos de performance, la ejecución concurrente a la hora de encontrar la palabra con mayor cantidad de apariciones en un conjunto de archivos. Elaboren en primer lugar una hipótesis, basándose en los conocimientos que poseen sobre el tema.

Luego, diseñen y efectúen experimentos que permitan poner a prueba esta hipótesis. Se sugiere medir el tiempo de ejecución requerido por cada una de las dos etapas del

proceso (cargar las palabras en la tabla de *hash* y computar el máximo), variando la cantidad de *threads* utilizada en cada una de ellas. Para hacerlo pueden utilizar la función `clock_gettime` (con la opción `CLOCK_REALTIME`) de la biblioteca `time.h`, que se puede linkear utilizando `-lrt`. Presenten sus resultados en el informe de manera clara, utilizando los recursos que les parezcan adecuados (gráficos, figuras, etc.).

Realicen luego un análisis de los resultados. No olviden tener en cuenta qué factores pueden haber influido en los mismos (realizando ajustes si los consideran necesarios), por ejemplo:

- Las características físicas de la(s) computadora(s) en las que ejecuten los experimentos, especialmente el grado de multiprocesamiento (múltiples núcleos o hilos de ejecución) disponible.
- Cómo varían los tiempos de ejecución, y la incidencia en los mismos de la concurrencia, según el tamaño de los datos procesados.

4. Condiciones de entrega

Deberán entregar un archivo comprimido que contenga:

- El informe, en formato PDF.
- El código correspondiente a su solución. El mismo debe poder compilarse sin problemas ejecutando el comando `make`.
- Todos los archivos, datos, scripts y/o instrucciones necesarias para poder replicar cualquier experimento mencionado en el informe.

La entrega se realizará a través de la página de la materia en el campus virtual. No se aceptarán entregas más allá de la fecha límite, sin excepción.