DEPARTAMENTO
DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

TP1

Sistemas Operativos

Integrante	LU	Correo electrónico
Lucas Puterman	830/13	Lucasputerman@gmail.com
Ivan Vercinsky	141/15	ivan9074@gmail.com
Alonso Tomás	396/16	tomasalonso96@gmail.com

Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Ciudad Universitaria - (Pabellón I/Planta Baja)

Intendente Güiraldes 2160 - C1428EGA

Ciudad Autónoma de Buenos Aires - Rep. Argentina

Tel/Fax: (54 11) 4576-3359

http://www.fcen.uba.ar

Índice

1. Introducción **3**

2. Desarrollo **4**
2.1. ListaAtomica 4
2.2. ConcurrentHashMap 4
2.2.1. ConcurrentHashMap() 4
2.2.2. ConcurrentHashMap(std::string arch) 4
2.2.3. ConcurrentHashMap(unsigned int nt, std::list<std::string>archs) 4
2.2.4. process_file(std::string arch) 4
2.2.5. add_and_inc(std::string key, int amount) 5
2.2.6. ConcurrentHashMap::member(std::string key) 5
2.2.7. ConcurrentHashMap::maximum(unsigned int nt) 5
2.3. Decisiones de implementacion 5

3. Resultados **6**
3.1. Tests 6
3.1.1. Tiempos de ejecucion 6

4. Conclusión **9**

1. Introducción

El motivo de este trabajo practico es implementar un diccionario que funciona sobre una tabla de hash. Esta implementacion sera de alta eficiencia y permitira accesos simultaneos manteniendo la consistencia de los datos, es decir, sera concurrente.

El diccionario solo toma strings como claves y se guarda el string con un valor entero que indica la cantidad de inserciones de dicho string, la funcion de hash utilizada consiste en simplemente tomar la primera letra del string a guardar. para las colisiones se utiliza una lista enlazada en cada letra del diccionarion(los resultados de la funcion de hash), donde las palabras nuevas se van agregando al final con su valor inicial de 1.

A continuacion se marcan las ideas iniciales y las decisiones de implementacion tomadas a lo largo del tp y finalmente se muestra una breve experimentacion para probar el funcionamiento correcto del codigo y mediciones de tiempo respecto a la funcion maximo la cual fue implementada de forma concurrente y tanbien de forma no concurrente.

2. Desarrollo

2.1. ListaAtomica

La $ListaAtomica$ es una implementación de lista usando variables atómicas. En particular, expone el método $push_front$que se encarga de lidiar con llamadas concurrentes, de esta forma, actualiza la cabeza de la lista que es atómica utilizando $compare\_exchange\_strong$.

A su vez, sus elementos se leen usando la función $load$ para así obtener el dato correcto.

2.2. ConcurrentHashMap

La estructura $ConcurrentHashMap$ cuenta con un arreglo de listas atómicas representando cada una de las letras del alfabeto. Además cuenta con un arreglo de mutex con un mutex por letra.

se implementaron los siguientes constructores y funciones:

2.2.1. ConcurrentHashMap()

```
ConcurrentHashMap()
        inicializar 26 listas atómicas.
```

2.2.2. ConcurrentHashMap(std::string arch)

Crea el ConcurrentHashMap y carga en su estructura el archivo pasado de forma no concurrente.

```
ConcurrentHashMap(string arch)
        Inicializar 26 listas atómicas.
        process_file(arch)
```

2.2.3. ConcurrentHashMap(unsigned int nt, std::list<std::string>archs)

Crea un ConcurrentHashMap de forma concurrente corriendo nt threads en paralelo que procesan los archivos.

```
 ConcurrentHashMap(int nt, list<string> archs)
        -Inicializar 26 listas atómicas.
        -declarar atomico para el indice de las listas
        -lanzar nt threads
        -esperar la finalización de los threads
```

Para esto se utiliza una variable atómica inicializada en -1 que va a representar el indice en la lista de archivos. Cada uno de los threads, pide un índice y mientras luego procesa el íesimo archivo

```
        Thread:
                i = indice.get_and_inc
                mientras i sea menor a la cantidad de archivos
                        process_file(archivos[i])
                        i = indice.get_and_inc
```

2.2.4. process_file(std::string arch)

Procesa las palabras de cada archivo y las agrega al ConcurrentHashMap llamando a la función add_and_inc.

2.2.5. add_and_inc(std::string key, int amount)

Esta función se fija si la palabra procesada ya se encuentra en la estructura y aumenta su valor. En caso de no estar creada, la crea. Antes de comenzar lockea el mutex de la lista.

```
-lockea el mutex correspondiente a la letra de la palabra.
-recorre la lista correspondiente para ver si existe la palabra.
-si existe, incrementa su valor en amount.
-si no existe la palabra, crea el par para la palabra y lo agrega a la lista.
-libera el mutex correspondiente a la letra de la palabra.
```

2.2.6. ConcurrentHashMap::member(std::string key)

Devuelve un booleano indicando si la palabra indicada está presente en la estructura, recorriendo la lista correspondiente a la palabra.

2.2.7. ConcurrentHashMap::maximum(unsigned int nt)

Calcula cual es la palabra que aparece con más repeticiones en la estructura, corriendo concurrentemente en nt threads.

```
-declara una ListaAtomica de pares y un atomico que va a ser el indice en los threads
-lockea los mutex de todas las letras
(para que nadie pueda agregar palabras mientras se calcula un máximo)
-lanza nt threads, cada uno agregara sus máximos a la lista atómica.
-espera la finalización de los nt threads.
-recorre la lista de máximos y devuelve el mayor
```

Cada uno de los threads funciona de la siguiente manera:

```
Thread:
        i = indice.get_and_inc
        mientras i sea menor a la cantidad de letras
                bucar maximo en la lista de la letra correspondiente
                agregar maximo a la lista atómica de maximos
                i = indice.get_and_inc
```

2.3. Decisiones de implementacion

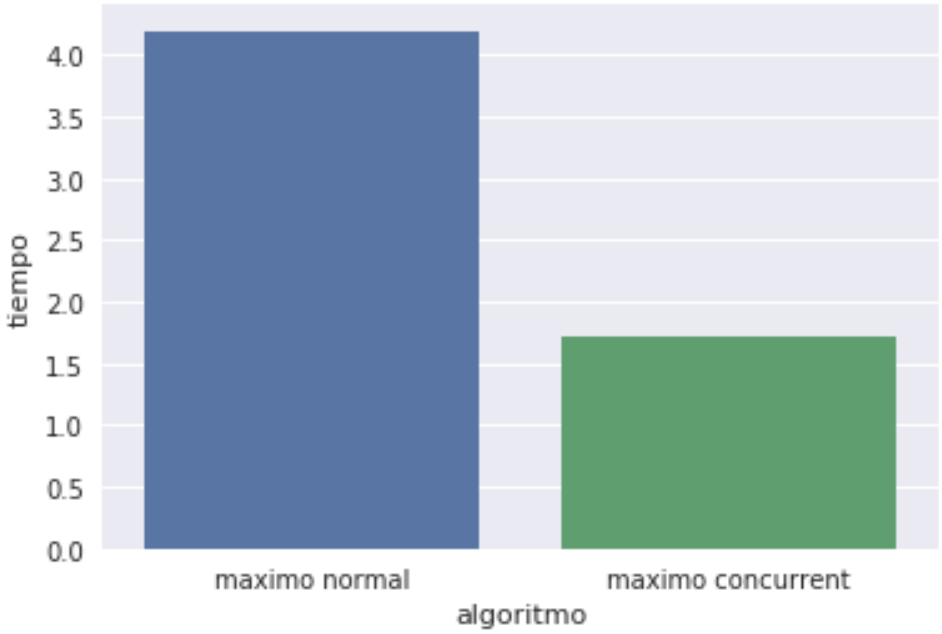
3. Resultados

3.1. Tests

El codigo implementado pasa todos los tests, se realizo un nuevo test el cual utiliza archivos generados aleatoriamente y compara los resultados de las dos funciones de maximos implementadas.

esto tambien cumple la funcion de medir sus tiempos de ejecucion varias veces, este test fue modificado para trabajar con muchos archivos o con pocos y comparar el desempeño de ambos algoritmos, mas adelante se hablara de estas mediciones.

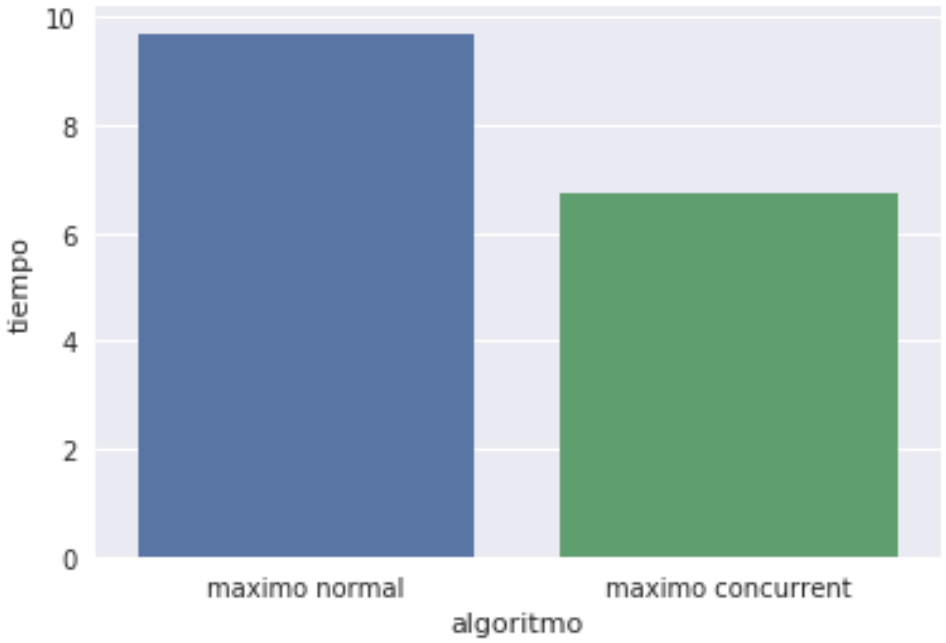
3.1.1. Tiempos de ejecucion



en este grafico se puede ver la comparacion entre las dos implementaciones de maximo funcionando para 5 archivos con 5 threads, se puede ver claramente que la implementacion concurrente es mas del doble de rapida.

	normal	concurrent
count	994.0	994.0
mean	4.196799688128771	1.7193847233400394
std	2.810343303108625	1.5340264700982498
min	1.76498	0.866309
25%	2.5708175	1.08561
50%	3.1239	1.223315
75%	4.4532825	1.4992675
max	20.3669	14.2455

en esta tabla se pueden ver los valores mas estadisticos de la comparacion, se corrieron aproximadamente mil veces los algoritmos para este caso.



Aqui se puede ver la comparacion de ambos para 11 archivos con 1 solo thread. la diferencia es menor pero igualmente es claro que la implementacion concurrente es mejor

	normal	concurrent
count	491.0	491.0
mean	9.699297087576372	6.731969287169046
std	4.632436986573542	3.4697132874875187
min	5.53769	4.58423
25%	7.19579	5.25685
50%	8.17784	5.73923
75%	9.847895000000001	6.5640350000000005
max	33.5729	27.7283

en esta tabla se pueden ver los valores mas estadisticos de la comparacion, en este caso los algorimos se corrieron aproximadamente quinientas veces.

4. Conclusión