República Bolivariana de Venezuela.

Universidad de Carabobo.

Facultado Experimental de Ciencias y Tecnología.

Departamento de Computación.

Arquitectura del Computador.

SIMULADOR DE MEMORIA CACHE.

Estudiante:

Gustavo A. Morillo M.

C.I.: V-26.160.169

DEFINICIONES DE INTERES.

* **Principio de Localidad:** Es un principio con el que se busca determinar que bloques de memoria se debería seleccionar, esto con el propósito de aumentar la probabilidad de aciertos en las búsquedas de los distintos niveles de la jerarquía de memoria. Este principio se constituye de 2 ideas: La localidad temporal (si se requiere una dirección de memoria una vez es posible que en el futuro se vuelva a requerir) y la Localidad espacial (si se requiere una dirección de memoria una vez, es posible que en un futuro se requiera esta o las direcciones de memoria adyacentes).
* **Jerarquía de memoria:** Es la manera en que se estructura el sistema de memoria del computador moderno, el propósito es el de acelerar el acceso a la información almacenada y que ese almacenamiento se incremente, esto se logra implementando diversos componentes (se consigue la ilusión de tener un único componente rápido pero que a su vez posee una gran capacidad de almacenamiento) ubicándolos en varios sitios del computador, siendo los más cercanos al procesador los más rápidos pero de menor capacidad de almacenamiento y los más lejanos a este de mayor lentitud pero mayor capacidad de almacenamiento. En la actualidad podemos ordenar la jerarquía de memoria de la manera siguiente:

1. **Nivel 1:** Memoria Caché(En este proyecto nos enfocaremos exclusivamente en esta)
2. **Nivel 2:** Memoria Ram
3. **Nivel 3:** Discos duros
4. **Nivel 4:** Cintas magnéticas
5. **Nivel 5:** Disco Duro extraíble
6. **Nivel 6:** Almacenamiento en red

* **Memoria Cache:** Es un componente de hardware o software que guarda datos para que las solicitudes futuras de esos datos se puedan atender con mayor rapidez; los datos almacenados en una caché pueden ser el resultado de un cálculo anterior o el duplicado de datos almacenados en otro lugar, generalmente, da velocidad de acceso más rápido.
* **Aciertos de la cache:** Un acierto ocurre cuando la dirección de los datos solicitados por el procesador se encuentran en la memoria cache.
* **Fallos de la cache:** Un fallo ocurre cuando un bloque de memoria de los datos solicitados por el procesador no se encuentran en la memoria cache, cuando esto ocurre esta accede un nivel más abajo (memoria RAM) en la jerarquía de memoria y obtiene los bloques de memoria relacionados con ese bloque de memoria

Correspondencia Directa.

Es una forma de organizar la memoria cache, esta organización puede ser implementada de 2 maneras:

1. La manera típica, en la cual cada bloque de memoria puede ser almacenado únicamente en una sola posición de memoria en la cache. Dependiendo de la base (binaria, decimal, entre otros) de la dirección de bloque hay una forma de determinar a qué posición de memoria pertenece dicho bloque; Para direcciones en base binaria los n-bits menos significativos (más a la derecha) determinan la posición de cache correspondiente, para direcciones en base decimal basa con calcular el modulo entre la dirección del bloque y el número de posiciones de cache ((dirección de bloque) % (número de posiciones de cache)).
2. Siendo la correspondencia directa una implementación de la correspondencia completamente asociativa (una organización de la cache que permite insertar en cualquier posición de la cache un bloque, ordenándolas en n-vías de m-posiciones) de 1 sola vía, permitiendo así insertar en cualquier posición de cache al bloque.

Correspondencia Asociativa por Conjuntos.

Forma de organizar la cache, esta organización es una combinación de la correspondencia directa y la correspondencia completamente asociativa ya que divide a la memoria cache en n-conjuntos de m-vías cada uno, aquí, un bloque de memoria podrá almacenarse en cualquier vía de un único conjunto posible, se determina por medio del módulo de la dirección del bloque entre el número de conjuntos ((dirección del bloque) % (número de conjuntos), similar al criterio en la correspondencia directa) y una vez indicado el conjunto se puede insertar o encontrar en cualquier vía del conjunto a dicho bloque (mismo criterio de la correspondencia completamente asociativa).

Política de Reemplazo del Menos usado recientemente (Least Recently used)

Las políticas de reemplazo son formas de manejar los fallos de la cache cuando esta tiene todas sus posiciones de memoria ocupadas, en este caso se debe reemplazar el contenido de una posición de memoria por el nuevo contenido a ubicar, con esta finalidad se han definido diversos algoritmos de los cuales el Reemplazo Aleatorio y LRU (Menos usado recientemente por sus siglas en inglés) serán implementados en este proyecto. La Política de Reemplazo LRU consiste en determinar cuál ha sido la posición de cache menos utilizada recientemente y utilizarla para insertar el nuevo bloque de memoria, cabe destacar que esa posición luego de habérsele insertado el nuevo bloque de memoria pasa a ser la más usada recientemente. Este reemplazo no puede ser usado únicamente en el primer tipo de correspondencia directa descrito previamente, del resto puede ser implementado en las demás organizaciones de cache descritas anteriormente, adaptándose a los criterios de cada una.

Política de Reemplazo Aleatorio

Consiste en seleccionar de manera aleatoria una posición de memoria dentro de la cache cuando ocurre un fallo de cache, no puede ser usado para el primer tipo de correspondencia directa descrito anteriormente, del resto puede ser implementado para las demás organizaciones de cache descritas, adaptándose a los criterios de cada una. Cabe destacar que este algoritmo agiliza el proceso de reemplazo ya que no requiere ninguna condición adicional en comparación al Reemplazo LRU.

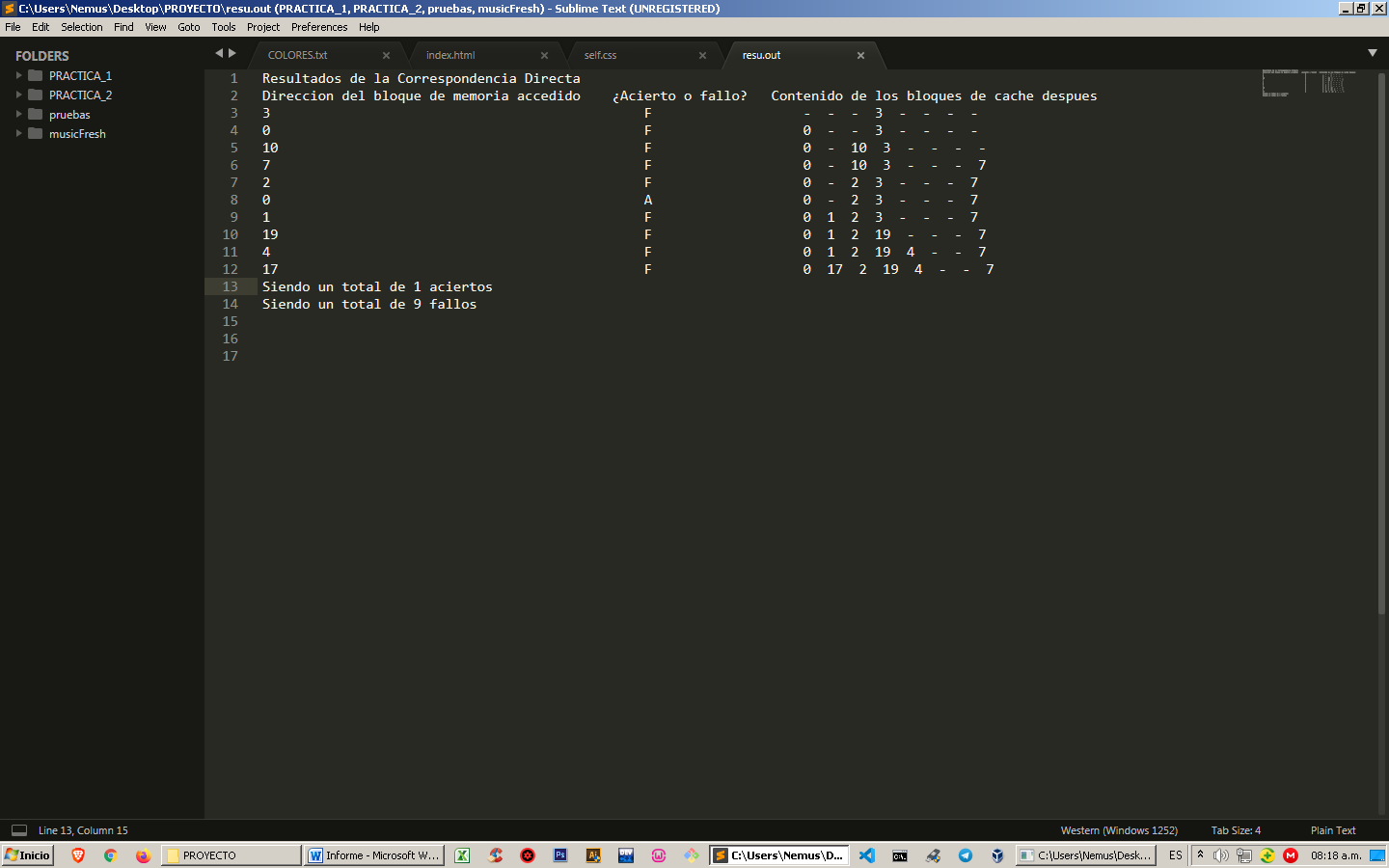
Desarrollo del simulador de memoria cache.

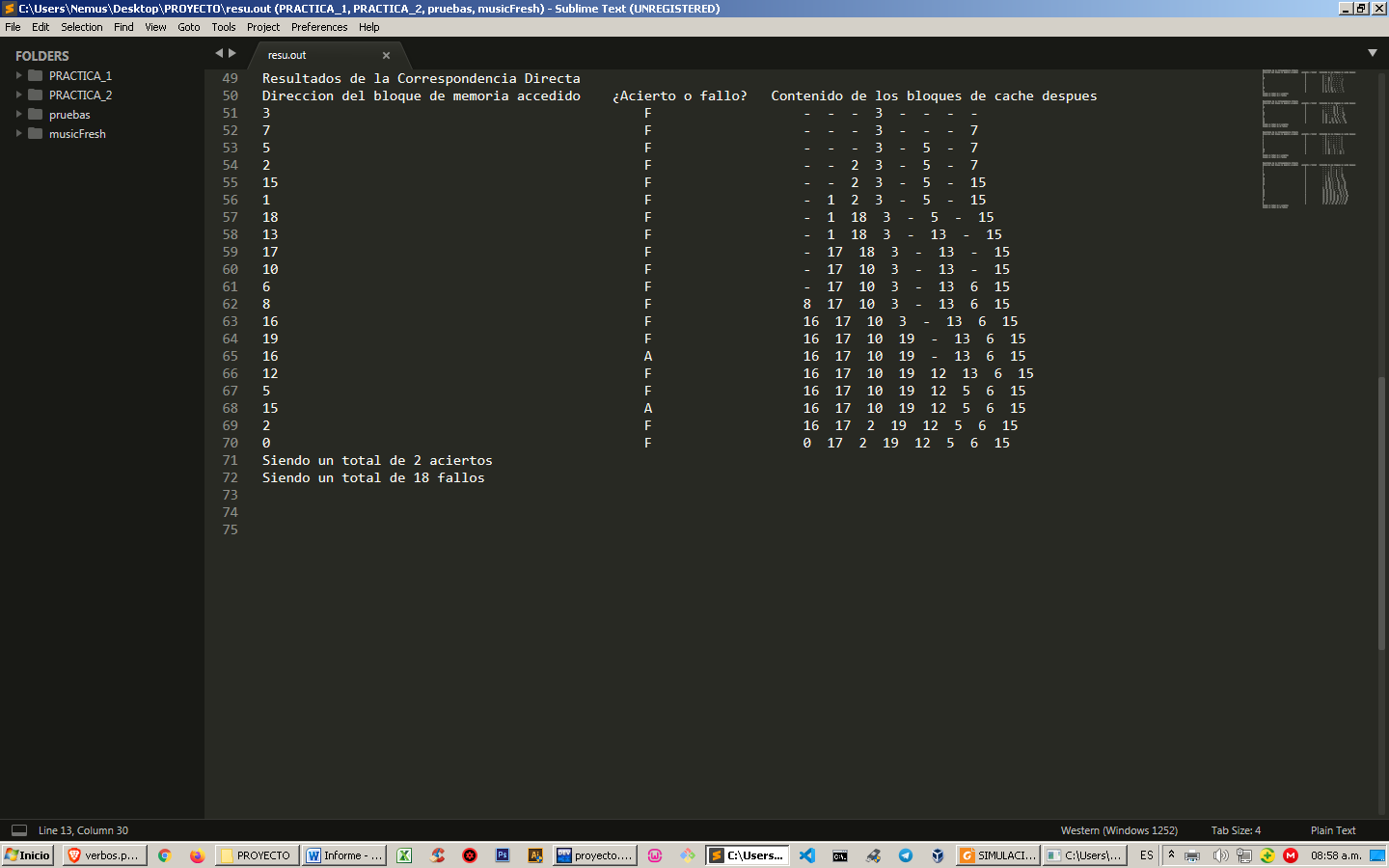
En el archivo adjunto proyecto.cpp se desarrolla un simulador de memoria cache, cache que se organiza como correspondencia directa del primer tipo descrito y como correspondencia asociativa por conjuntos, en este simulador se implementa tanto el Reemplazo Aleatorio como el Reemplazo LRU solamente para la correspondencia asociativa por conjuntos ya que en la correspondencia directa simulada un bloque solamente puede insertarse en una única posición de memoria de la cache. Debajo se listan las funciones y procedimientos implementados en el archivo.

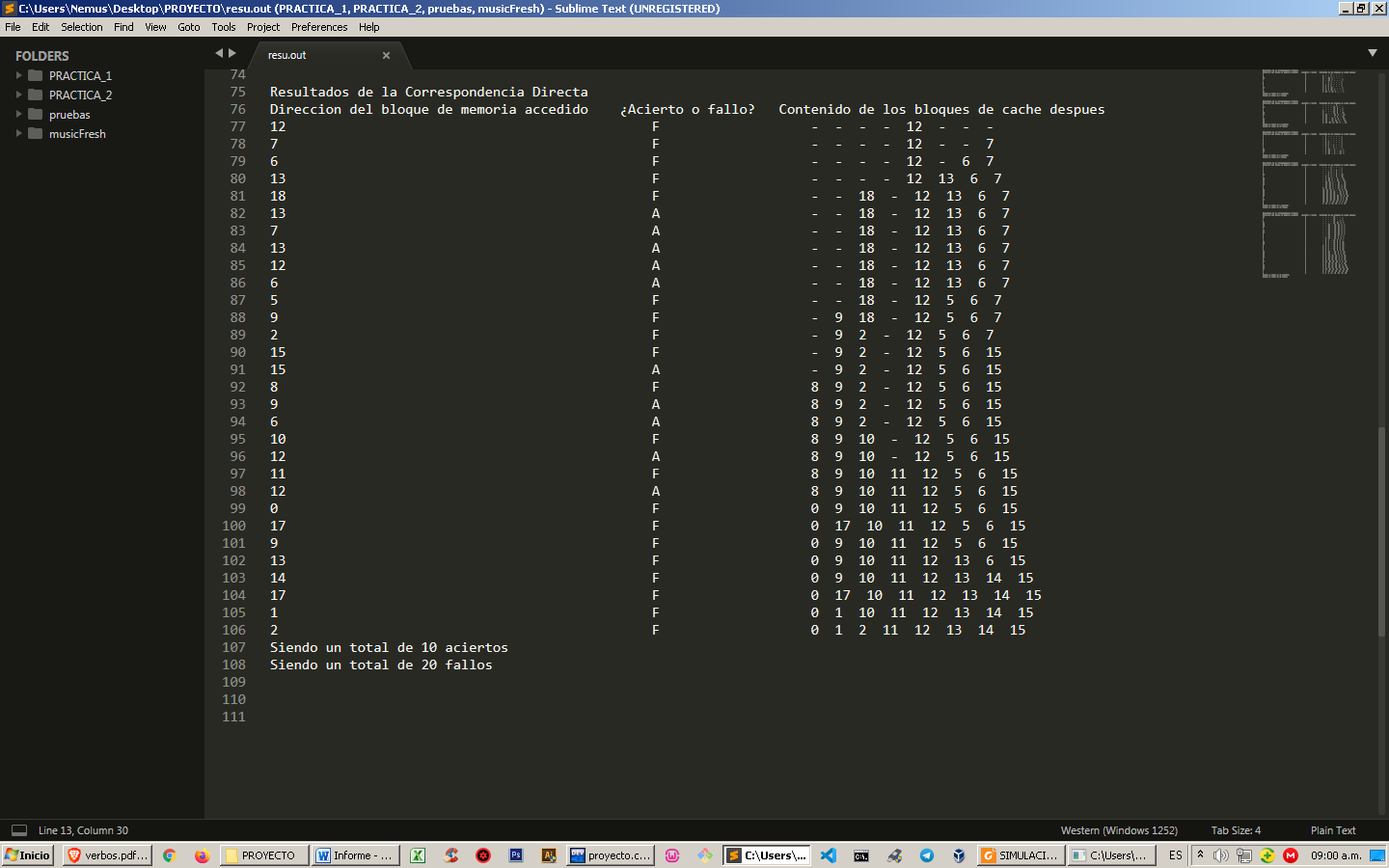
* Void bienvenida(void) ::: Procedimiento que se llama desde la función main() para mostrar en pantalla un menú de opciones donde se podrá seleccionar un tipo de correspondencia o salir, además, se leerán las opciones de configuración para la caché(tamaño de cache, tamaño de línea, valores para especificar cual correspondencia se ejecutara y valores para especificar cual algoritmo de reemplazo usar) y las direcciones de bloque, estableciendo entonces las entradas para la cache, las correspondencias, los algoritmos de reemplazo, el número de entradas, número de líneas de la cache. Una vez el usuario selecciona una opción de correspondencia el programa llama al respectivo procedimiento pasando como parámetros los bloques, número de bloques y el número de líneas.
* Void bloques\_aleatorios(void) ::: El programa debe leer las direcciones de bloque desde un archivo ‘’traz.in’’, lo que hace este procedimiento es escribir en dicho archivo direcciones aleatorias con cada ejecución del programa. Escribe números aleatorios tantas veces como se lo indiquemos con la variable ‘’i’’ y el rango de valores que puede tomar un numero aleatorio está dado por el denominador del módulo entre rand() % 25.
* Int lectura\_bloques(int bloques[]) ::: Esta función lee lo que escribió bloques\_aleatorios() en el archivo ‘’traz.in’’ y lo guarda línea por línea en el arreglo ‘’bloques’’ al mismo tiempo que va contando el número de direcciones leídas, finalmente retorna ese número de bloques y finaliza la función con las direcciones guardadas en ‘’bloques’’.
* Int lectura\_configuracion(int correspondencias[2],int algoritmos[2]) ::: Función que guarda en ‘’correspondencias’’ los valores que representan las correspondencias de directa y asociativa por conjuntos y guarda en ‘’algoritmos’’ los valores que representan los algoritmos de reemplazo de LRU y Aleatorio, finalmente esta función devuelve el número de líneas que tendrá la cache al dividir el tamaño de la cache entre el tamaño que tendrá cada una de sus líneas.
* Void escribir\_en\_archivo(resultados r[],int n\_bloques,int n\_lineas,int aciertos,int fallos,int k) ::: Imprime en el archivo ‘’resu.out’’ los resultados almacenados en ‘’r’’, arreglo cuyos elementos tienen una estructura que por una dirección de bloque indicará si se produjo un acierto o un fallo, el conjunto en el que almacenó el bloque(en caso de asociativa por conjuntos) y la cache resultante luego de la manipulación de esa dirección ( en el caso de asociativa por conjuntos será el conjunto accedido), ya que se implementan 2 tipos de correspondencia con la variable k se indica para cual correspondencia va a ser la salida.
* Void directa(int bloques[],int n\_bloques,int n\_lineas) ::: Ejecuta la correspondencia directa. Una cache de este tipo puede ser vista como un arreglo unidimensional, dicho arreglo se inicializa a -1 todas sus posiciones, luego por cada dirección de bloque se determina a que línea de cache pertenece con el criterio (Dirección de bloque) % (número de líneas) y se verifica si en esta línea ya estaba el bloque o no, en caso de no estarlo se marca como un fallo ‘F’ y se cuenta cómo uno(incrementando ‘fallos’) y se inserta en esa línea, en caso contrario se marca como un acierto ‘A’ y se cuenta como uno(incrementando aciertos). Por último, se guarda en el campo ‘tmp\_cache’ la cache resultante luego de manipular un bloque y revisadas todas las direcciones se escribe en el archivo de salida. La complejidad en tiempo del procedimiento por si solo (si no se tuviera que escribir la cache resultante para la salida) es Lineal (O(n)), en este caso es Cuadrático (O(n2)).
* Void asociativaConjuntos(int bloques[],int n\_bloques,int n\_lineas) ::: Establece el número de vías por conjunto y calcula cuantos conjuntos dividirán a la cache, esto es dividiendo el (n\_lineas) / (vias). Por último, se llaman a los algoritmos de reemplazo pasando a cada uno como parámetros los bloques, número de estos, el número de conjuntos calculados y el número de vías establecido.
* int busqueda(int cache[],int vias,int x) ::: Es una función que recorre de manera lineal un arreglo ‘’cache’’ de tantas ‘’vías’’ para buscar un valor ‘x’ si lo encuentra deja de recorrer el arreglo y devuelve la vía en que se encuentra ese valor, en este proyecto se implementa una búsqueda lineal debido a que las direcciones de cache no estarán ordenadas, razón por la cual no se implementa la búsqueda binaria. Esta función servirá para ‘’ver’’ dentro de una vía.
* Void reemplazoAleatorio(int bloques[],int n\_bloques,int conjuntos,int vias) ::: Ejecuta La Política de Reemplazo Aleatorio, una cache de correspondencia asociativa por conjuntos puede ser vista como una matriz bidimensional donde los conjuntos serían vistos como filas y las vías serían vistos como columnas. Empezamos determinando a que conjunto pertenece el bloque para posteriormente hacer una búsqueda dentro del conjunto y verificar si se encuentra o no el bloque, en caso de no encontrarse se hace una búsqueda pero para hallar una vía vacía(el reemplazo se realiza cuando no se consiga alguna). La vía a la cual se le hará el reemplazo de su bloque se selecciona con el módulo rand() % (vías). Se considerará un fallo cuando un bloque no se encuentre y se inserte en una vía vacía o cuando se haga el reemplazo. Finalmente se escribe en el archivo de salida los resultados almacenados en ‘’r’’ para cada dirección. La complejidad en tiempo de este procedimiento por si solo (si no se tuviera que escribir el conjunto accedido luego de manipular una dirección) es cuadrática (O(n2)), con la escritura del conjunto resultante esta complejidad en tiempo no cambia.
* int lru(int conjunto[],int vias) ::: Busca el término mayor en un arreglo ‘’conjunto’’ de tantas ‘’vías’’ y devuelve la posición del término.
* Void reemplazoLRU(int bloques[],int n\_bloques,int conjuntos,int vias) :: Ejecuta el Reemplazo LRU, para implementar este algoritmo se declara además de una matriz bidimensional que representará a la cache asociativa por conjuntos, se declara una matriz bidimensional paralela a la cache que en vez de almacenar bloques de memoria almacenará contadores para cada vía de cada conjunto dichos contadores cambiarán su valor cuando no se utilice una vía no vacía para un bloque en la cache o cuando se use una vía no vacía para un bloque. Se inicializan tanto la cache como la cache paralela, se determina el conjunto al que pertenece el bloque y se vuelva el contenido de las vías de ese conjunto tanto en la memoria cache como en la cache paralela en arreglos unidimensionales con los cuales se harán búsquedas para conseguir el bloque o una vía vacía (con el conjunto de la cache) y determinar la vía menos recientemente usada (con el conjunto de contadores paralelo al conjunto de la cache). Cuando se utiliza una vía de un conjunto esa vía y todas las vías vacías del conjunto se ponen a 0 en la cache paralela mientras que las vías no vacías que no fueron usadas van incrementando, el propósito es el de ver la vía con el término mayor como la vía que lleva más tiempo sin ser usada. Finalmente almacena los resultados en ‘’r’’ y los manda a escribir en el archivo de salida.

Resultados de los casos de prueba

**Primer caso de Prueba:** En este ejemplo se utilizan 10, 20, 30 direcciones de memoria, para la correspondencia directa de una memoria cache de 8 líneas…

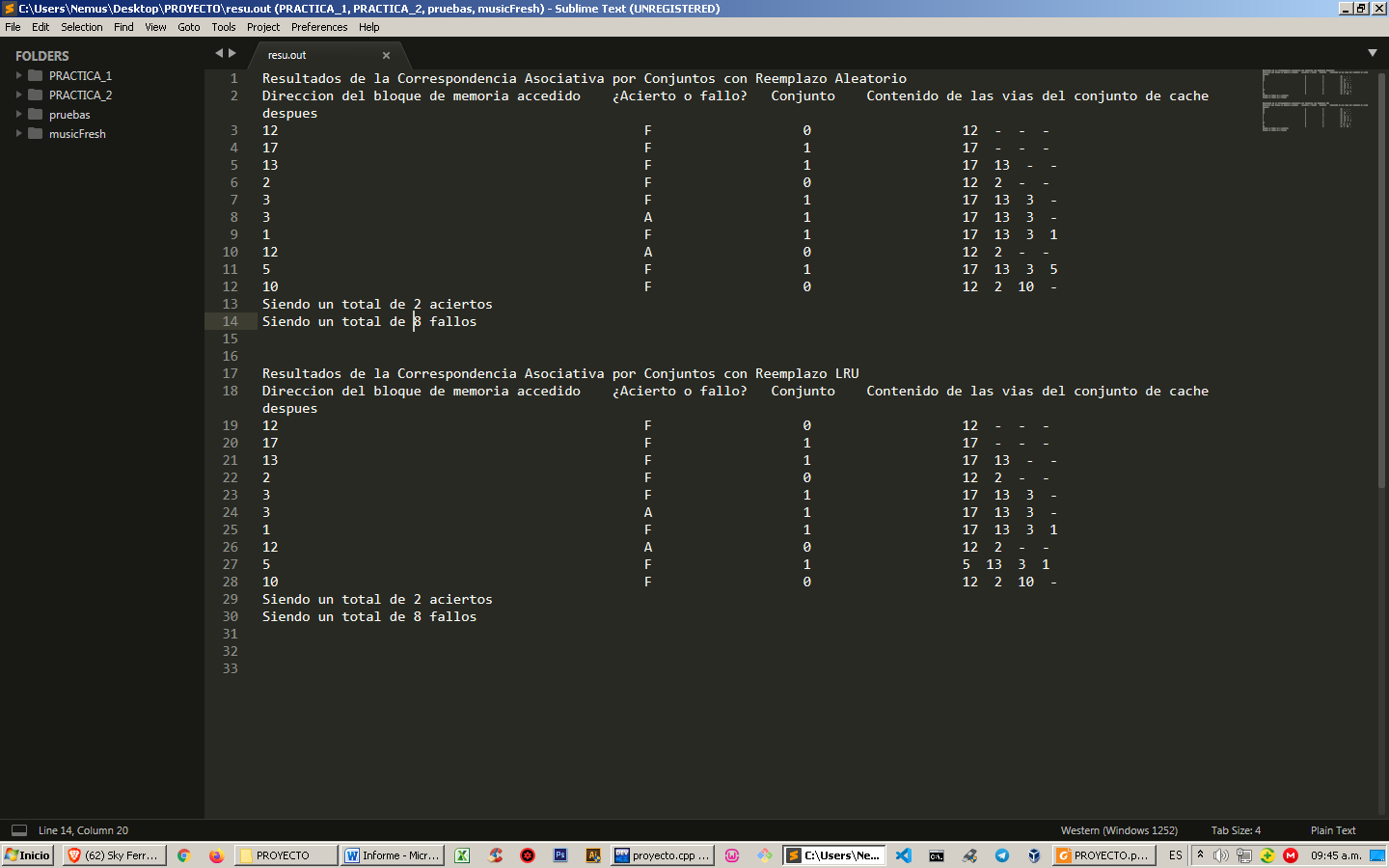


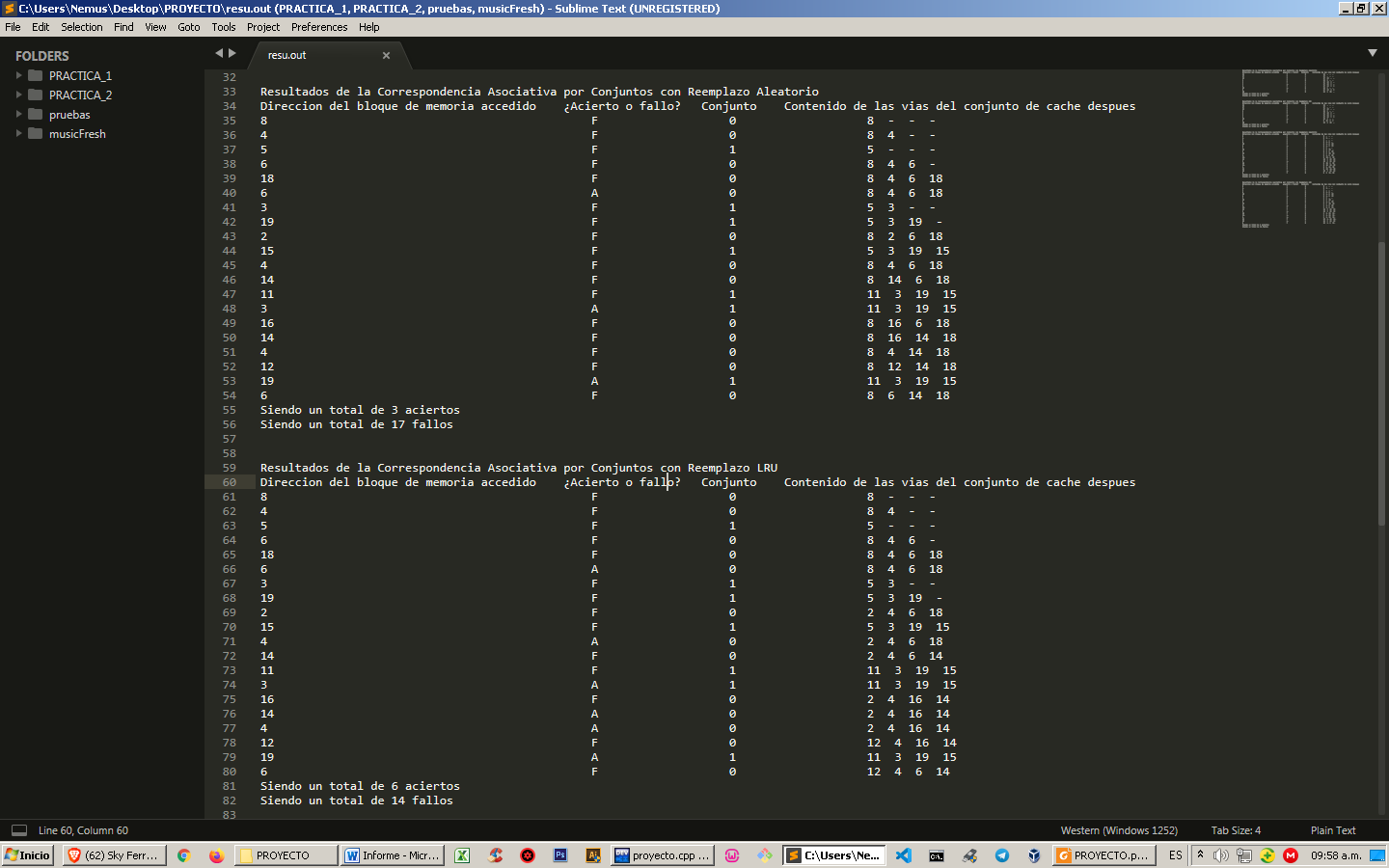


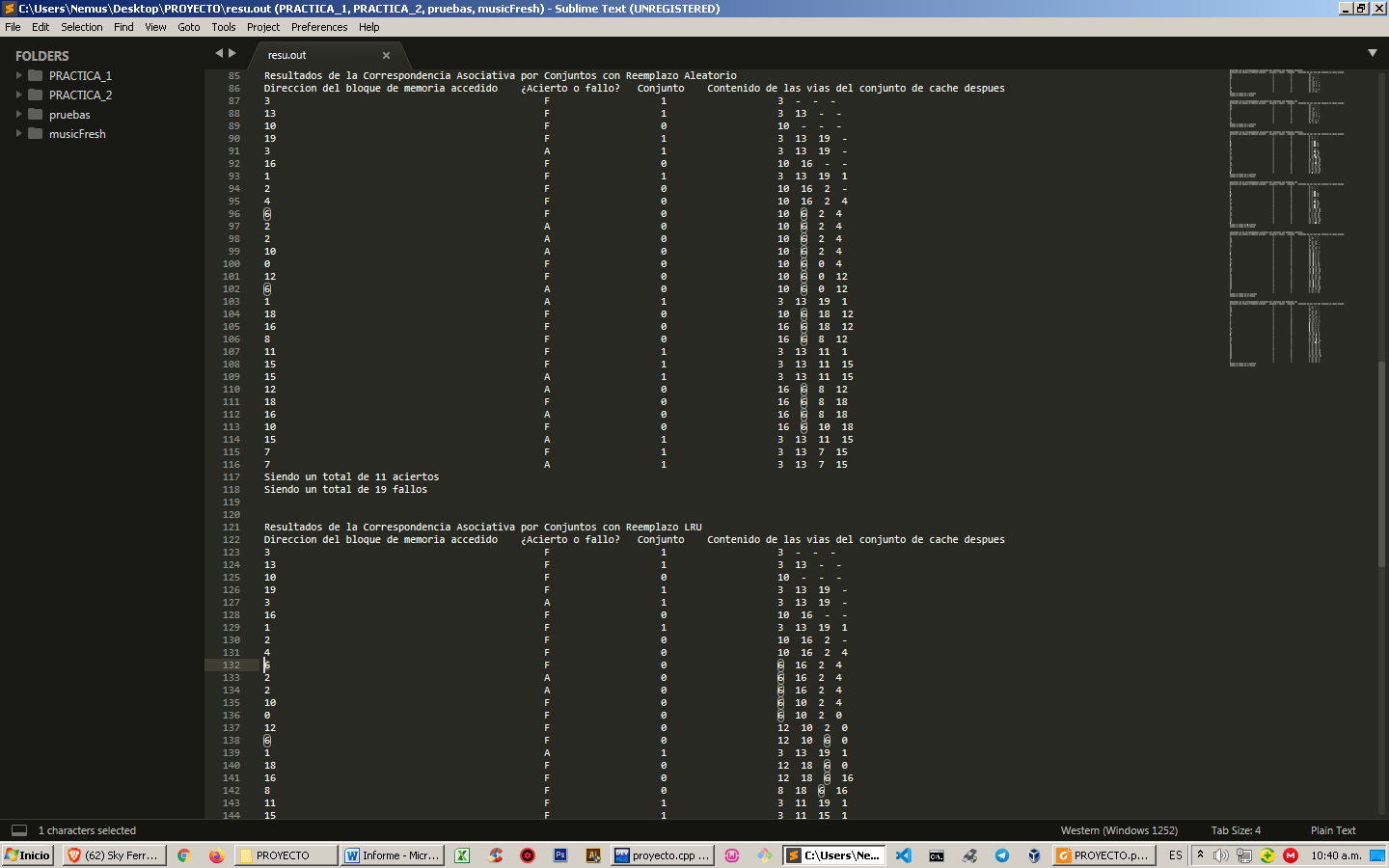


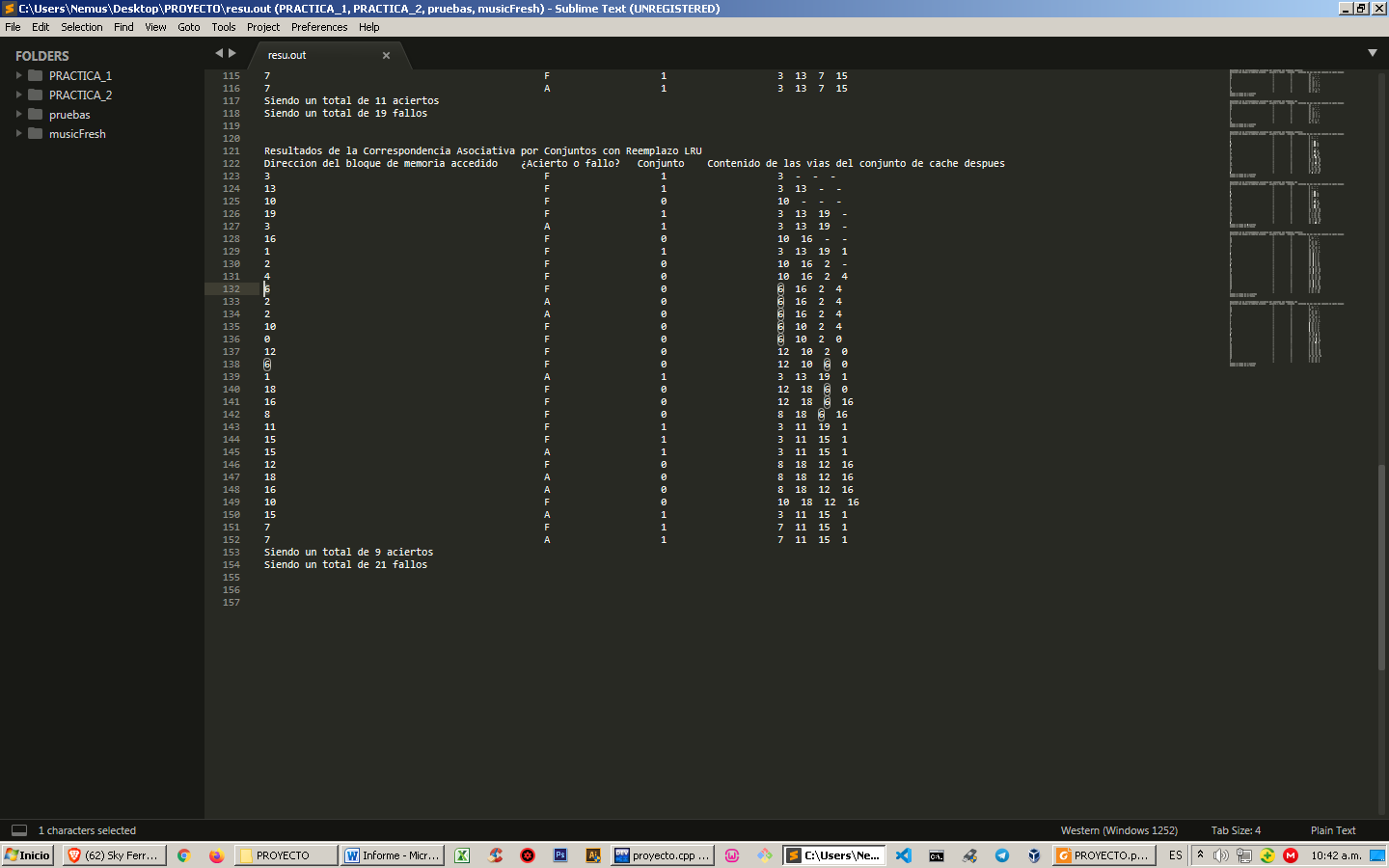
Para 10 direcciones de memoria hubo un fallo del 90% y un acierto del 10%, para 20 direcciones de memoria hubo un fallo del 90% y un acierto del 10%, para 30 direcciones de memoria hubo un fallo del 66.6% y acierto del 33.3%. Algo a destacar con la correspondencia directa es que al tener un solo lugar donde insertar un bloque reemplazar un bloque en una posición de memoria se corre el riesgo de que este se encuentre se vuelva a solicitar en un futuro, provocando así un mayor número de fallos. Otra observación sería el aumento de aciertos cuando es una cantidad más grande de direcciones solicitadas.

Continuando con el primer caso de prueba se adjuntan los resultados para la asociativa por conjuntos



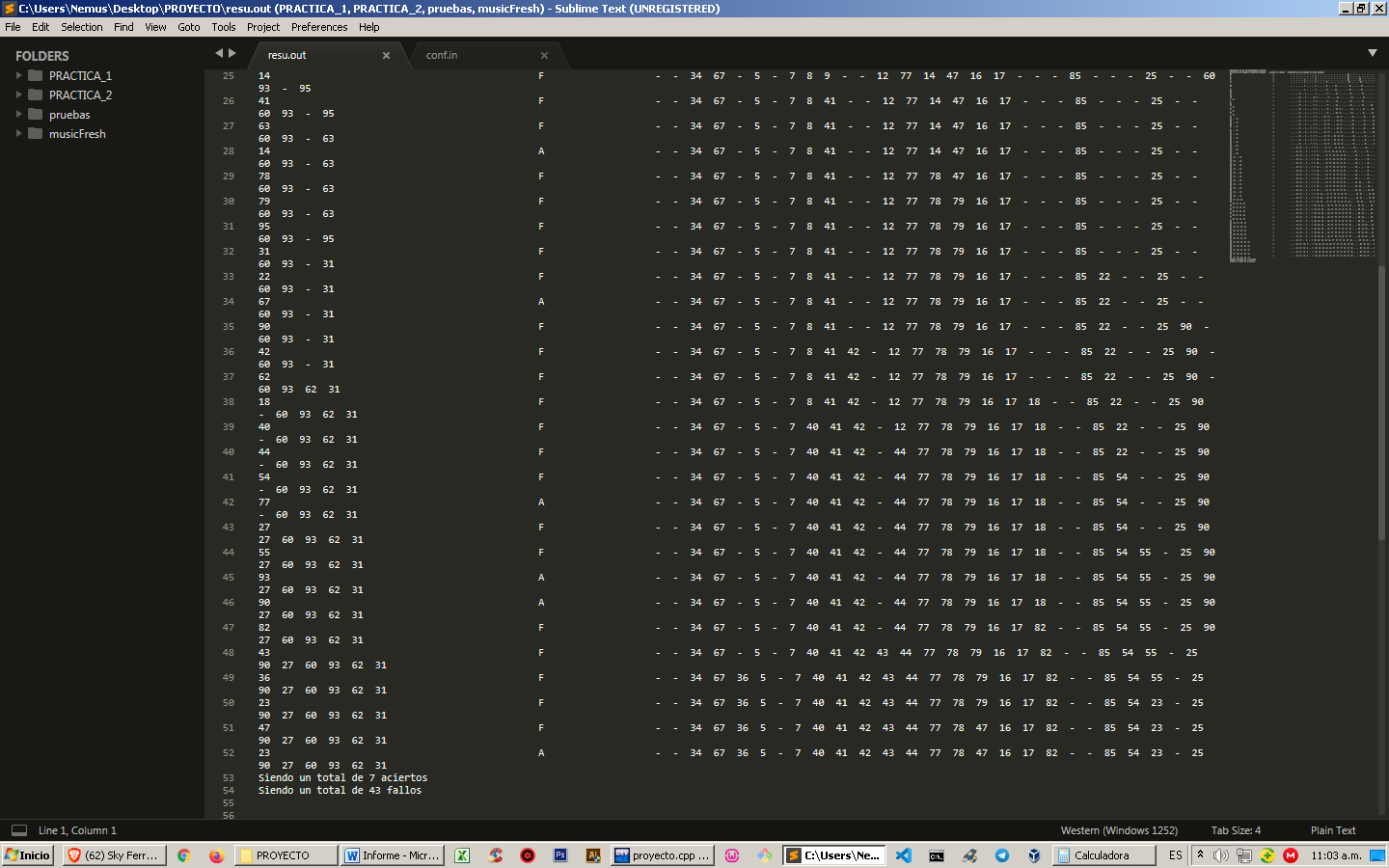


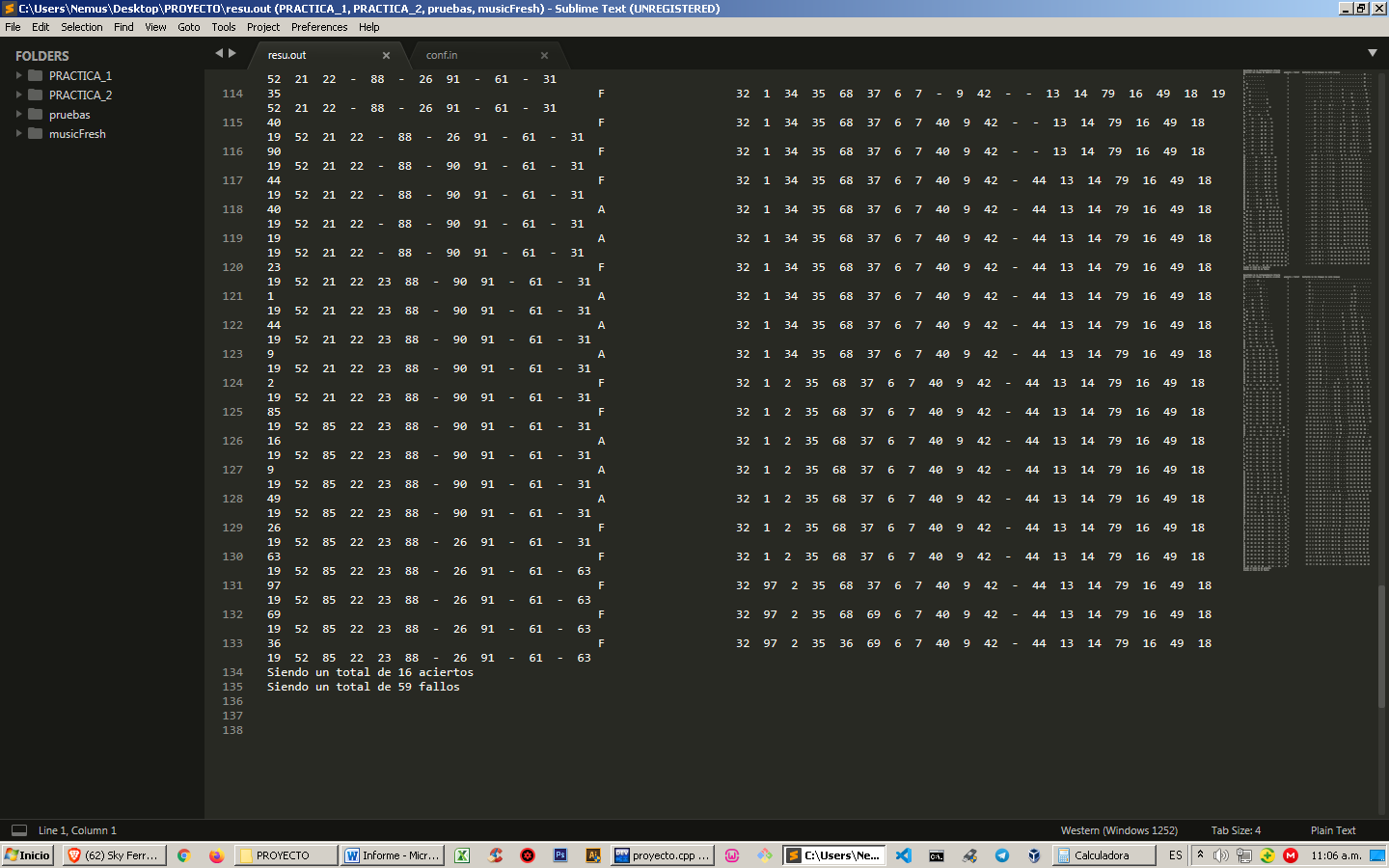


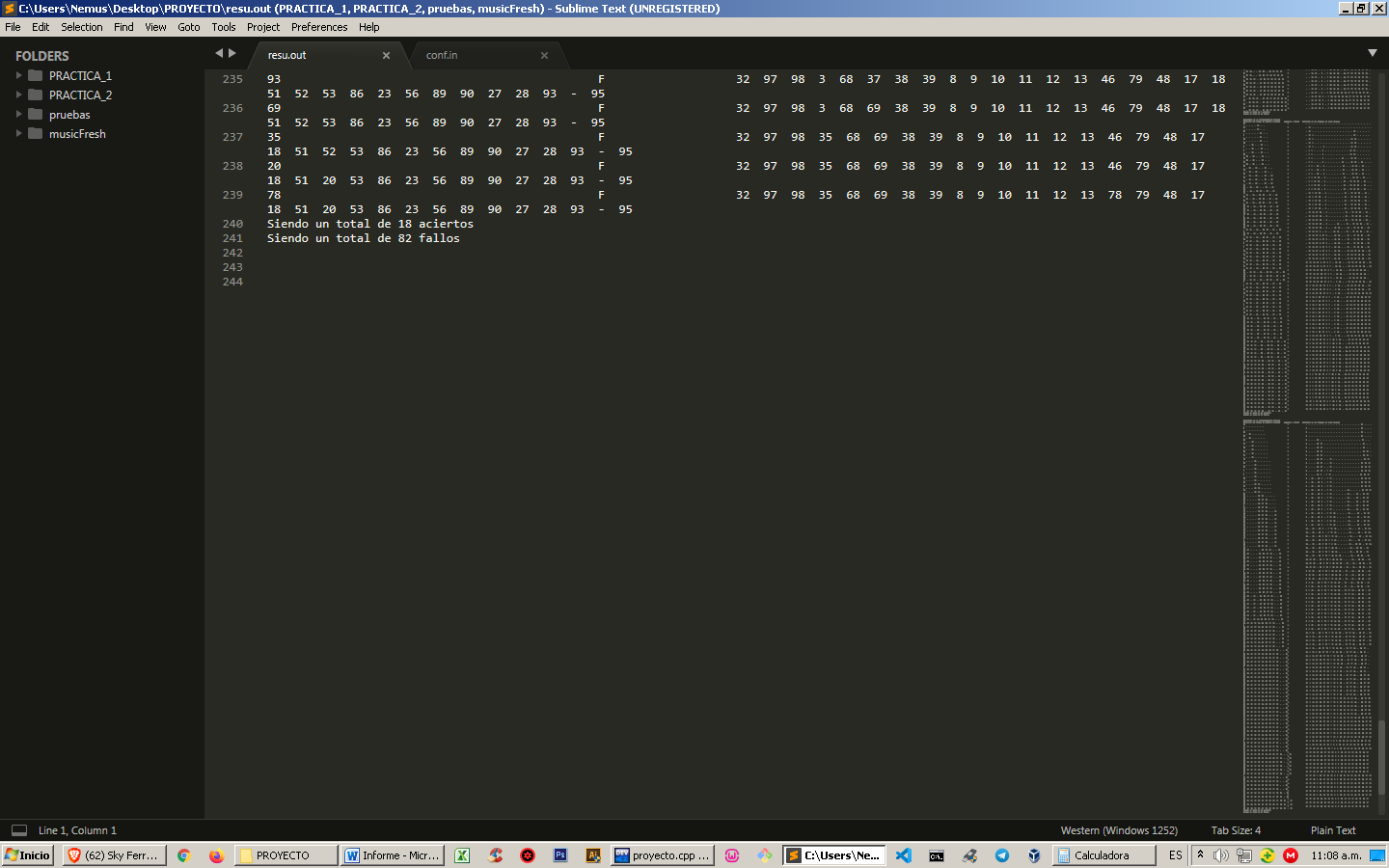


Para 10 direcciones de memoria hubo un fallo del 80% y de acierto del 20% en ambos algoritmos de reemplazo, para 20 direcciones de memoria hubo un fallo del 85% y un acierto del 15% para Reemplazo Aleatorio mientras que un fallo del 80% y un acierto del 20% para LRU, para 30 direcciones de memoria hubo un fallo del 60.3% y un acierto del 29.7% para Reemplazo Aleatorio mientras que un fallo del 70% y un acierto del 30% para LRU. En estas pruebas el algoritmo de Reemplazo Aleatorio resultó ser más eficiente que LRU, en cuanto a tiempo de ejecución LRU es más costoso ya que se debe verificar de entre un conjunto de vías que vía tiene más tiempo sin ser usada… Otra observación sería el provecho de la localidad temporal que hace LRU, al ser este el criterio que utiliza para escoger una vía candidata

**Segundo caso de prueba:** Una cache de 32 líneas a la que se le solicitarán 50, 75, 100 direcciones de memoria, iniciando con la correspondencia directa…

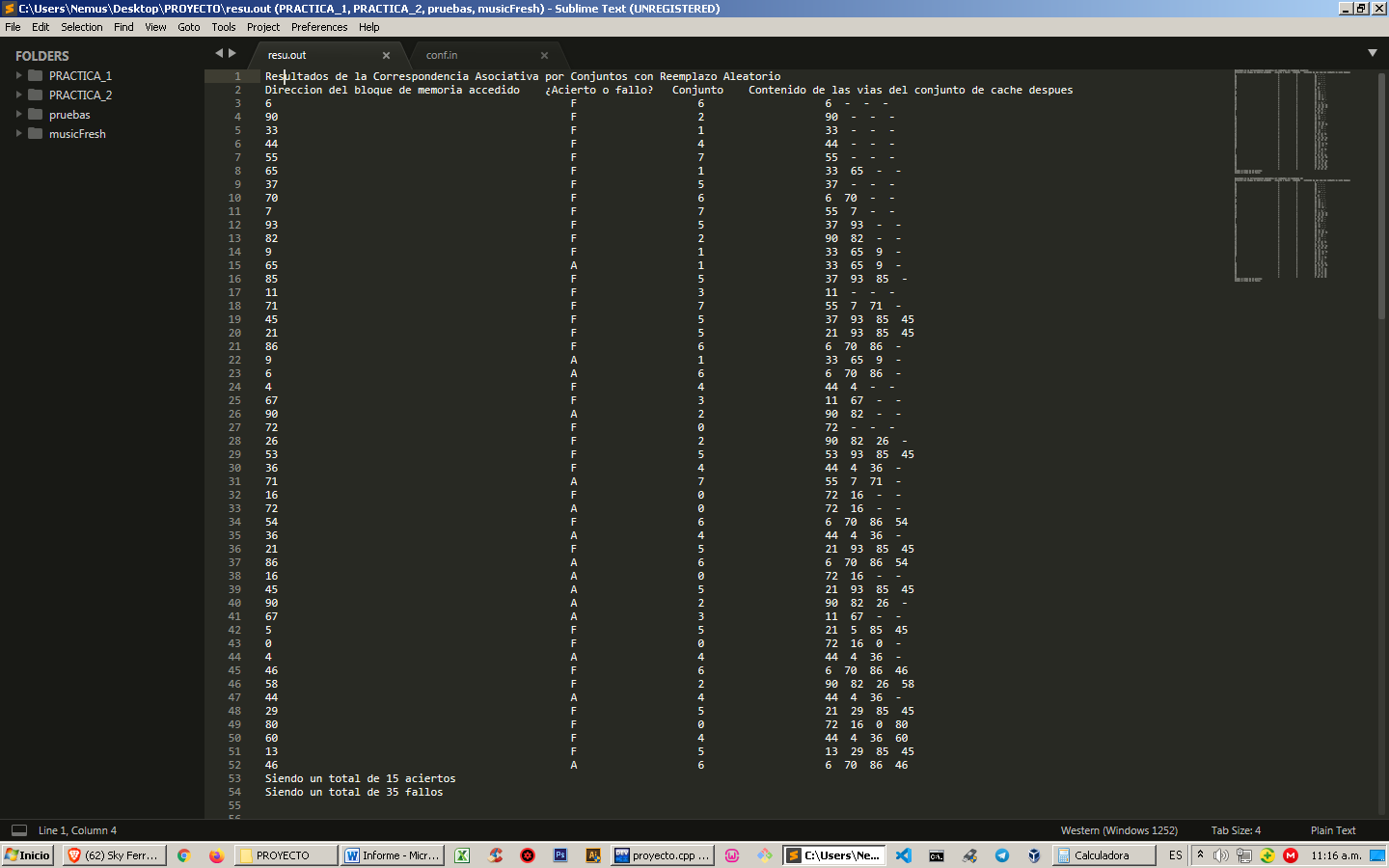


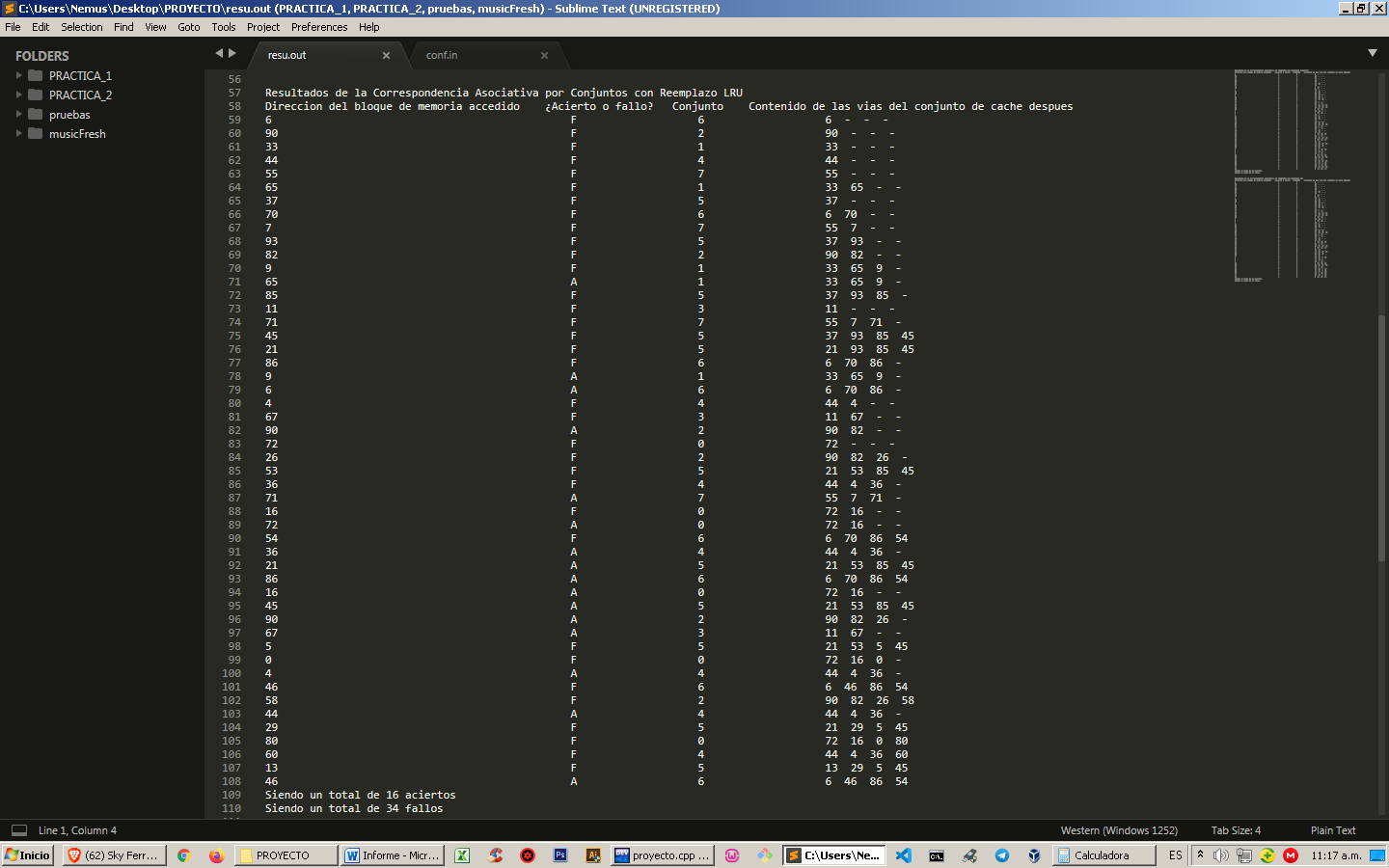


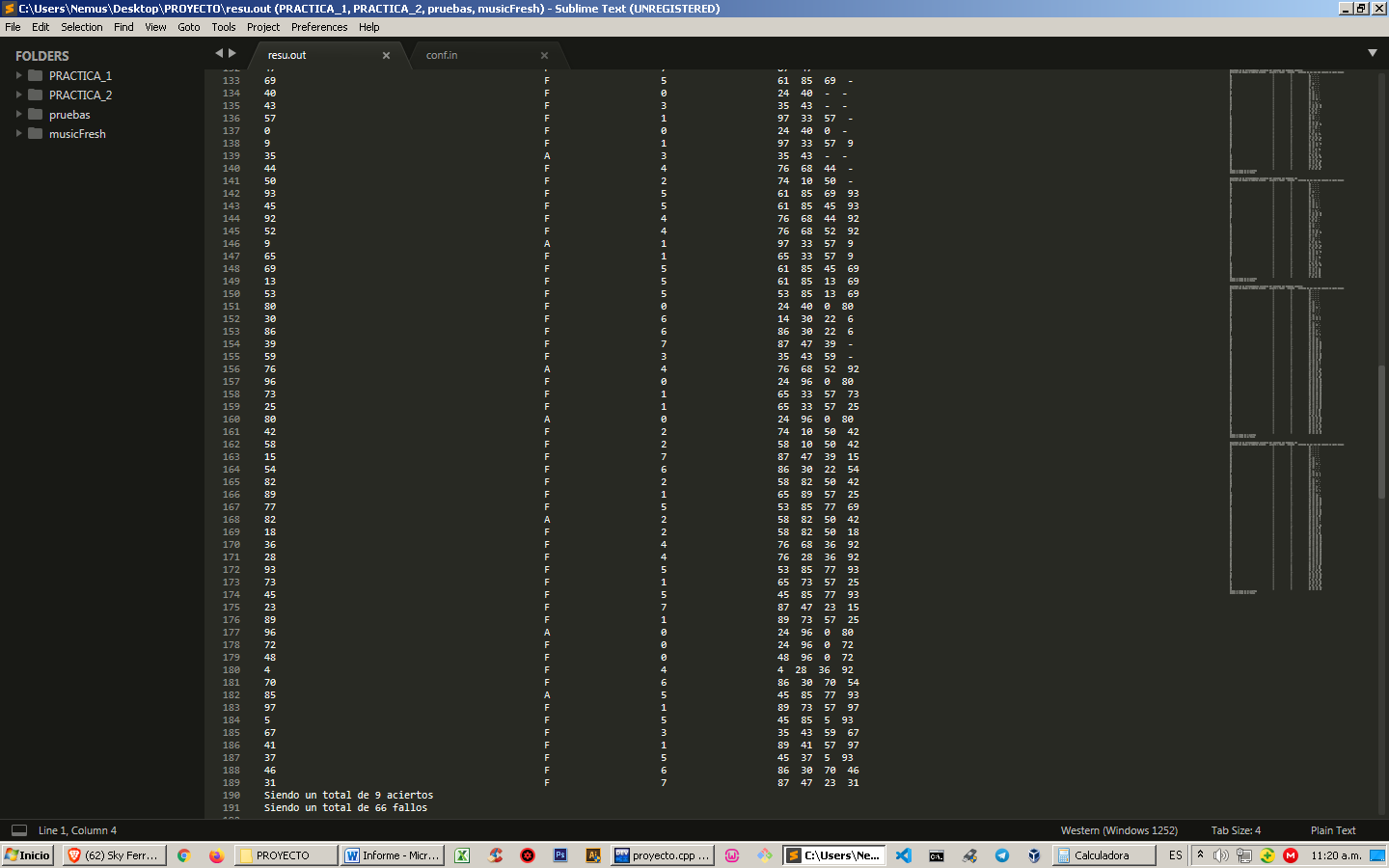


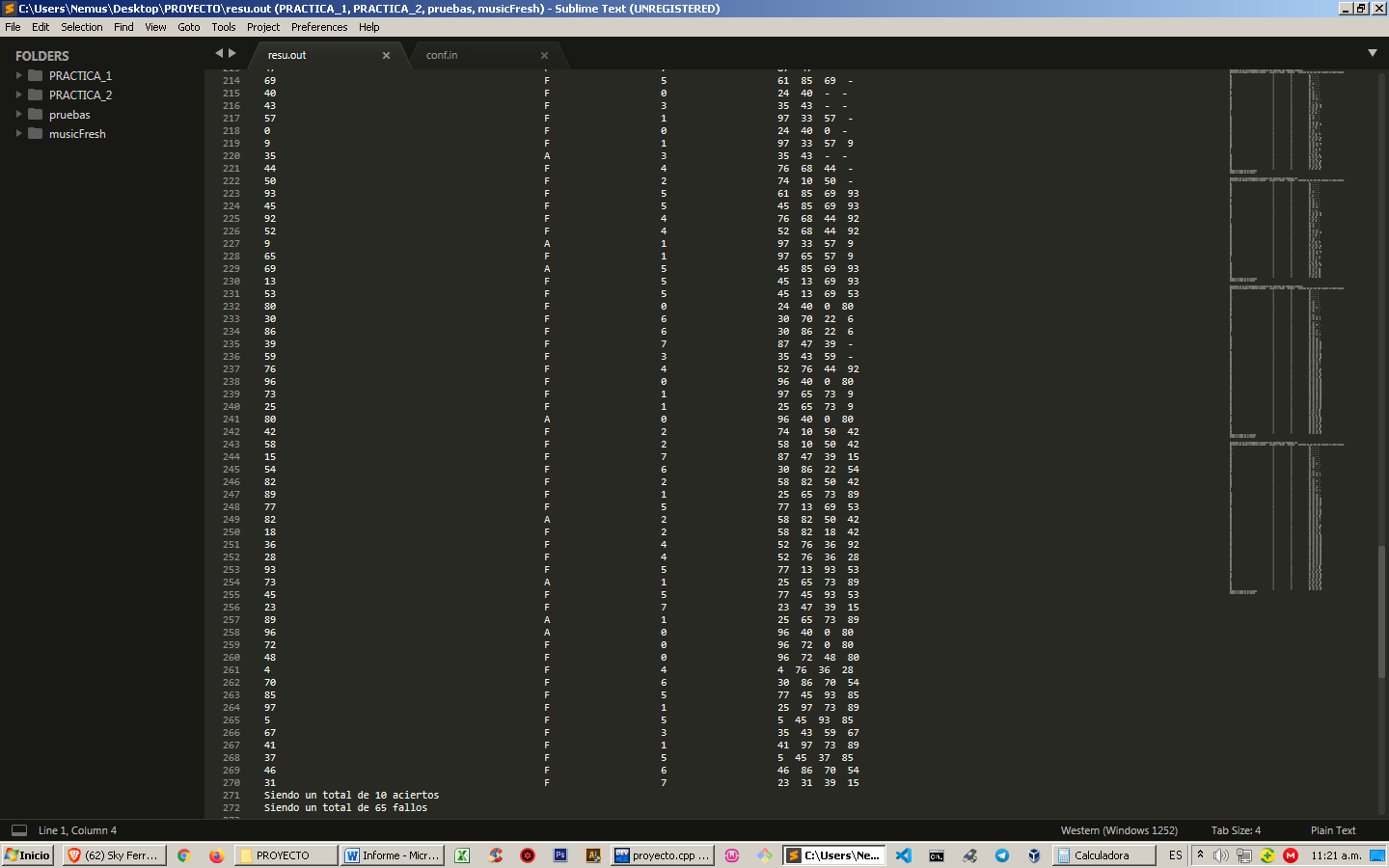
Para 50 direcciones de memoria hubo un fallo del 86% y un acierto del 14%, para 75 direcciones de memoria hubo un fallo del 78.6% y un acierto del 21.4%, para 100 direcciones de memoria hubo un fallo del 82% y un acierto del 18%. A un mayor número de líneas la correspondencia directa incrementa mucho el fallo a las solicitudes.

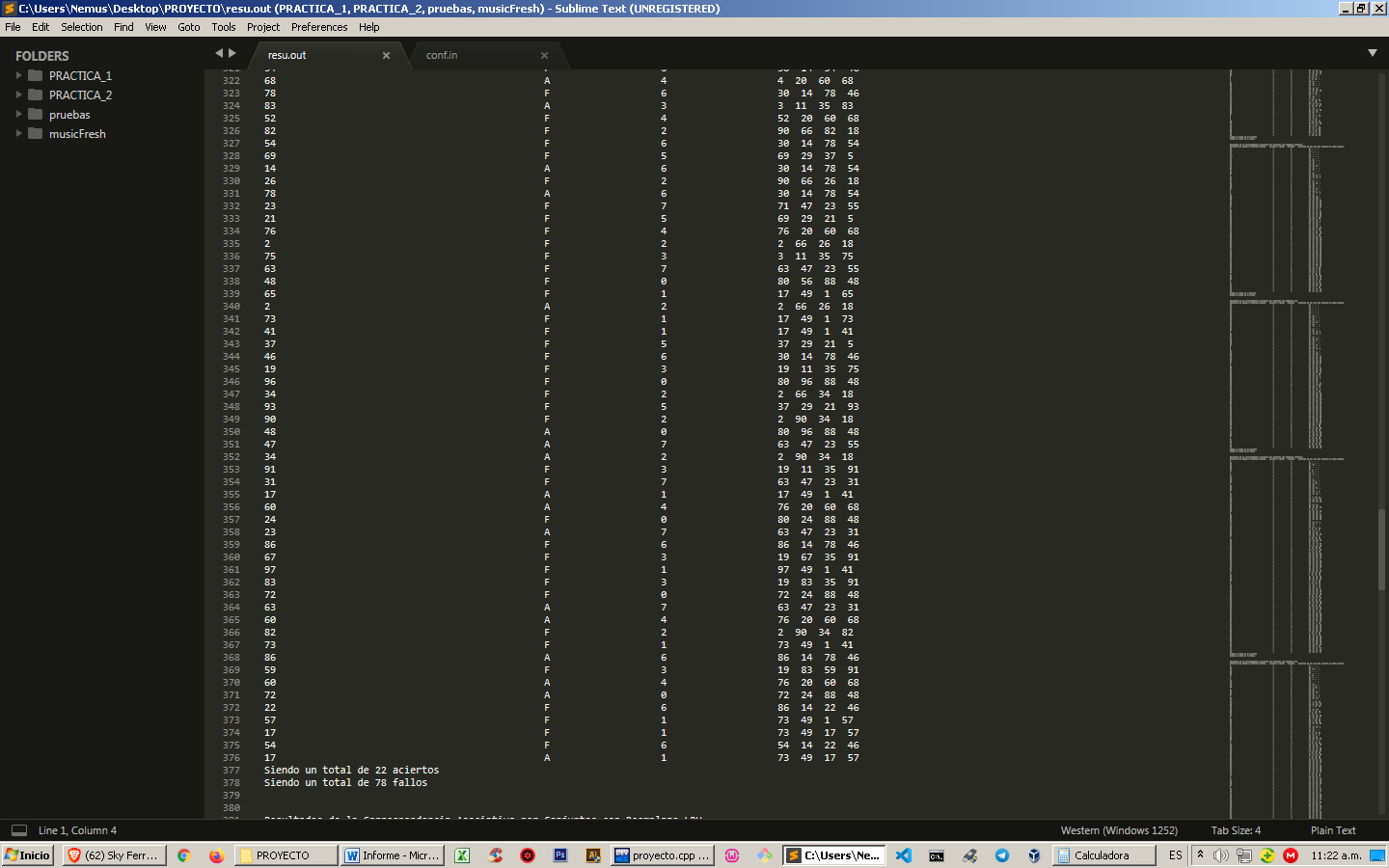
Continuando con el segundo caso de prueba, los resultados para la correspondencia asociativa por conjuntos.

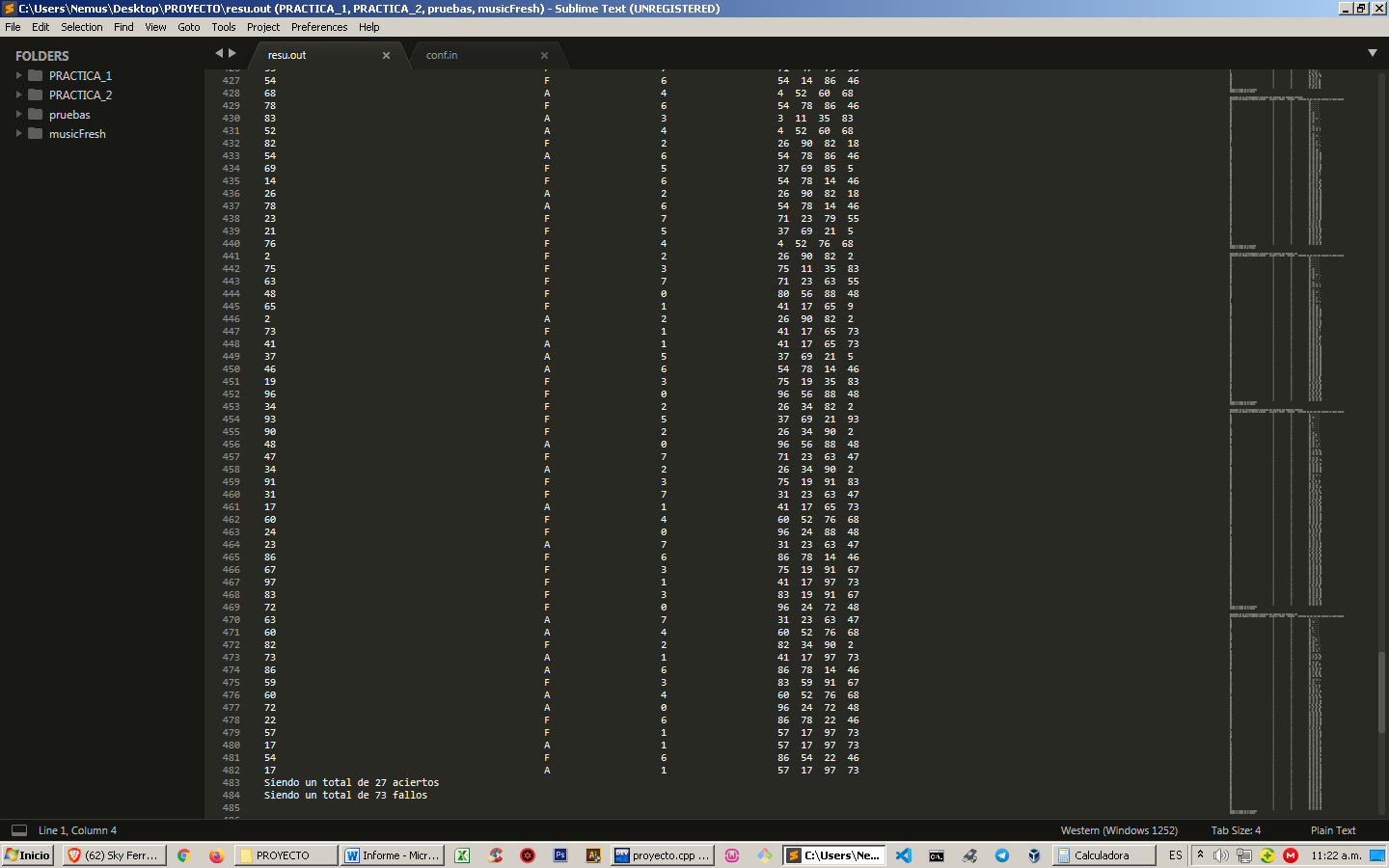












Para 50 direcciones de memoria hubo un fallo del 70% y acierto del 30% para Reemplazo Aleatorio mientras que para LRU un fallo del 68% y acierto del 32%, para 75 direcciones de memoria hubo un fallo del 88% y un acierto del 12% para Reemplazo Aleatorio mientras que para LRU un fallo del 86.6% y un acierto del 13.4%, para 100 direcciones de memoria hubo un fallo del 78% y un acierto del 22% para Reemplazo Aleatorio mientras que para LRU un fallo del 73% y un acierto del 27%. Según estos resultados el algoritmo de Reemplazo LRU a mayor número de direcciones solicitadas es mejor opción que realizar Reemplazos aleatorios, la deficiencia en cuanto a la predicción de direcciones que serán solicitadas se debe a que Reemplazo Aleatorio no toma en cuenta ningún criterio para sustituir un bloque.