**ANÁLISIS DEL RETO**

Iker Barbosa, 202424135, i.barbosac@uniandes.edu.co

Camilo Castro, 202412178, email 2

Melisa Molina, 202312232, email 3

# **Requerimiento 1**

## **Descripción**

El punto del requerimiento es identificar el último registro recopilado según un año de interés. Por este motivo, usamos la librería datetime para poder comparar las fechas de la columna load\_time. Primero, revisamos que la lista no esté vacia, si lo está, retorna None. En caso de que no esté vacía, recorremos la lista para hallar el último registro añadido por medio de comparaciones y guardarlo en una variable. Por último, recorremos la lista otra vez para hallar la fecha mínima encontrada y así agregar los datos al diccionario de retorno. Al encontrar el primero forzamos la salida del bucle con return().

|  |  |
| --- | --- |
| **Entrada** | agro, year:str |
| **Salidas** | retorno\_final = {"numero\_registros": 0, "registro": {}} |
| **Implementado (Sí/No)** | Implementado, Iker Barbosa |

## **Análisis de complejidad**

Análisis de complejidad de cada uno de los pasos del algoritmo

|  |  |
| --- | --- |
| **Pasos** | **Complejidad**  O(N), la función buscar\_anios(), que se guarda en la variable lista, se usa para filtrar solamente los registros por el año que entra por parámetro. Para filtrar se debe recorrer toda la lista.  O(1), se declara el diccionario que se va a retornar el diccionario al final. |
|  |  |
|  | O(1), consultar el tamaño de la lista filtrada. |
|  | O(n), recorrer la lista otra vez para hallar la fecha mas reciente. |
|  | O(n), recorre la lista para encontrar el registro con la fecha máxima y lo añade al diccionario que retorna la función. |
| ***TOTAL*** | ***O(n)*** |

## **Pruebas Realizadas**

Las pruebas realizadas fueron realizadas en una maquina con las siguientes especificaciones. Los datos de entrada fueron el catálogo de datos y el año 2024.

|  |  |
| --- | --- |
| Procesadores | AMD Ryzen 7 5700G |
| Memoria RAM | 22 GB |
| Sistema Operativo | Windows 10 |

|  |  |
| --- | --- |
| **Entrada** | **Tiempo (s)** |
| 100000 datos |  |
| 200000 datos |  |
| 300000 datos |  |
| 400000 datos |  |
| 500000 datos |  |

|  |  |
| --- | --- |
| Procesadores | 12th Gen Intel(R) Core(TM) i5-12450H |
| Memoria RAM | 16 GB |
| Sistema Operativo | Windows 11 |

|  |  |
| --- | --- |
| **Entrada** | **Tiempo (s)** |
| 100000 datos |  |
| 200000 datos |  |
| 300000 datos |  |
| 400000 datos |  |
| 500000 datos |  |

### **Tablas de datos**

Las tablas con la recopilación de datos de las pruebas.

|  |  |  |
| --- | --- | --- |
| **Muestra** | **Salida** | **Tiempo (ms)** |
| 100000 datos |  | 34.431 |
| 200000 datos |  | 58.343 |
| 300000 datos |  | 91.552 |
| 400000 datos |  | 120.980 |
| 500000 datos |  | 136.652 |

### **Graficas**

Las gráficas con la representación de las pruebas realizadas.

|  |  |
| --- | --- |
| Procesadores | AMD Ryzen 7 5700G |
| Memoria RAM | 22 GB |
| Sistema Operativo | Windows 10 |

|  |  |
| --- | --- |
| Procesadores | 12th Gen Intel(R) Core(TM) i5-12450H |
| Memoria RAM | 16 GB |
| Sistema Operativo | Windows 11 |

## **Análisis**

Al tratarse de recorridos de lista completos para realizar comparaciones, por temas de velocidad decidimos usar Single Linked List, la implementación de este requerimiento tiene un orden lineal O(n). Esto se debe a que debemos iterar para recorrer la lista a su totalidad, ya que debemos comparar todos los elementos.

Este comportamiento se puede evidenciar experimentalmente en la gráfica. Por su forma y su pendiente es una función lineal. Por lo que cumple con el comportamiento lineal esperado.

# **Requerimiento 2**

## **Descripción**

El punto del requerimiento es identificar el último registro cargado dado un departamento de interés. Similar al requerimiento anterior, usamos la librería datetime para poder comparar las fechas de la columna load\_time y así identificar el último registro cargado. En este caso, simplemente recorremos la lista y con un condicional verificamos los datos correspondan al departamento que recibimos por parámetro y guardamos las fechas en variables para compararlas. Por último, recorremos la lista otra vez para hallar la fecha mínima encontrada y así agregar los datos al diccionario de retorno. Al encontrar el primero forzamos la salida del bucle con return().

|  |  |
| --- | --- |
| **Entrada** |  |
| **Salidas** | retorno\_final = {"numero\_registros": 0, "registro": {}} |
| **Implementado (Sí/No)** | Implementado, Iker Barbosa |

## **Análisis de complejidad**

Análisis de complejidad de cada uno de los pasos del algoritmo

|  |  |
| --- | --- |
| **Pasos** | **Complejidad** |
|  | O(n), recorrer la lista para hallar el último registro añadido dependiendo del departamento pasado por parámetro. |
|  | O(n), recorrer la lista para hallar el último registro añadido en el departamento pasado por parámetro. |
| ***TOTAL*** | ***O(n)*** |

## **Pruebas Realizadas**

Las pruebas realizadas fueron realizadas en una maquina con las siguientes especificaciones. Los datos de entrada fueron el catálogo de datos y el estado de CALIFORNIA.

|  |  |
| --- | --- |
| Procesadores | AMD Ryzen 7 5700G |
| Memoria RAM | 22 GB |
| Sistema Operativo | Windows 10 |

|  |  |
| --- | --- |
| **Entrada** | **Tiempo (s)** |
| 100000 datos |  |
| 200000 datos |  |
| 300000 datos |  |
| 400000 datos |  |
| 500000 datos |  |

|  |  |
| --- | --- |
| Procesadores | 12th Gen Intel(R) Core(TM) i5-12450H |
| Memoria RAM | 16 GB |
| Sistema Operativo | Windows 11 |

|  |  |
| --- | --- |
| **Entrada** | **Tiempo (s)** |
| 100000 datos |  |
| 200000 datos |  |
| 300000 datos |  |
| 400000 datos |  |
| 500000 datos |  |

### **Tablas de datos**

Las tablas con la recopilación de datos de las pruebas.

|  |  |  |
| --- | --- | --- |
| **Muestra** | **Salida** | **Tiempo (ms)** |
| 100000 datos |  | 29.489 |
| 200000 datos |  | 86.088 |
| 300000 datos |  | 120.451 |
| 400000 datos |  | 145.154 |
| 500000 datos |  | 166.005 |

### **Graficas**

Las gráficas con la representación de las pruebas realizadas.

|  |  |
| --- | --- |
| Procesadores | AMD Ryzen 7 5700G |
| Memoria RAM | 22 GB |
| Sistema Operativo | Windows 10 |

|  |  |
| --- | --- |
| Procesadores | 12th Gen Intel(R) Core(TM) i5-12450H |
| Memoria RAM | 16 GB |
| Sistema Operativo | Windows 11 |

## **Análisis**

Al tratarse de recorridos de lista completos para realizar comparaciones, por temas de velocidad decidimos usar Single Linked List, la implementación de este requerimiento tiene un orden lineal O(n). Esto se debe a que debemos iterar para recorrer la lista a su totalidad, ya que debemos comparar todos los elementos.

Este comportamiento se puede evidenciar experimentalmente en la gráfica. Al inicio tiene una pendiente más inclinada, pero a medida que tenemos más datos, adopta su forma lineal y una pendiente más fija. Por lo que cumple con el comportamiento lineal esperado.

# **Requerimiento 4**

## **Descripción**

El punto del requerimiento es consultar los registros recopilados para un tipo de producto de interés dado un rango de años de recopilación dado. Por este motivo, use una función auxiliar para que me dé un Array List filtrado entre años de carga. Desde acá es más fácil recorrer de nuevo un Array List más pequeño y filtrarlo por tipo de producto. En caso de que el número de registros sea menor o igual a 20 devuelvo una Pila, en caso contrario, devuelvo un Array List con acceso a elementos.

|  |  |
| --- | --- |
| **Entrada** |  |
| **Salidas** | 1. Tupla con pila, boolean, size, numero de registros tipo celsus, numero de registros tipo survey. 2. Tupla con Array List, boolean, size, numero de registros tipo celsus, numero de registros tipo survey. |
|  | Implementado, Iker Barbosa |

## **Análisis de complejidad**

Análisis de complejidad de cada uno de los pasos del algoritmo

|  |  |
| --- | --- |
| **Pasos** | **Complejidad** |
|  | O(n), la función auxiliar buscar\_entre\_anios\_al() es una función que recorre todo el Array List para filtrar entre los años recibidos por parametro. |
|  | O(n), recorrer la lista filtrada entre años para hallar los registros que coincidan con el tipo de producto que se recibió por parámetro. |
|  | O(n), si el tamaño de lista es menor o igual a 20, recorremos la lista para añadir sus elementos a una pila y retornarla. |
|  | O(n), en caso de que el tamaño de la lista sea mayor a 20. Creamos una nueva Array List con new\_list(), e iteramos la lista para añadir los primeros 5 elementos y los últimos 5 con get.element() y retornamos el Array List. |
| ***TOTAL*** | ***O(n)*** |

## **Pruebas Realizadas**

Las pruebas realizadas fueron realizadas en una maquina con las siguientes especificaciones. Los datos de entrada fueron el catálogo de datos y el tipo de producto EGGS, entre 1980 y 2024.

|  |  |
| --- | --- |
| Procesadores | AMD Ryzen 7 5700G |
| Memoria RAM | 22 GB |
| Sistema Operativo | Windows 10 |

|  |  |
| --- | --- |
| **Entrada** | **Tiempo (s)** |
| 100000 datos |  |
| 200000 datos |  |
| 300000 datos |  |
| 400000 datos |  |
| 500000 datos |  |

|  |  |
| --- | --- |
| Procesadores | 12th Gen Intel(R) Core(TM) i5-12450H |
| Memoria RAM | 16 GB |
| Sistema Operativo | Windows 11 |

|  |  |
| --- | --- |
| **Entrada** | **Tiempo (s)** |
| 100000 datos |  |
| 200000 datos |  |
| 300000 datos |  |
| 400000 datos |  |
| 500000 datos |  |

### **Tablas de datos**

Las tablas con la recopilación de datos de las pruebas.

|  |  |  |
| --- | --- | --- |
| **Muestra** | **Salida** | **Tiempo (ms)** |
| 100000 datos |  | 72.495 |
| 200000 datos |  | 142.293 |
| 300000 datos |  | 249.642 |
| 400000 datos |  | 293.690 |
| 500000 datos |  | 359.728 |

### **Graficas**

Las gráficas con la representación de las pruebas realizadas.

|  |  |
| --- | --- |
| Procesadores | AMD Ryzen 7 5700G |
| Memoria RAM | 22 GB |
| Sistema Operativo | Windows 10 |

|  |  |
| --- | --- |
| Procesadores | 12th Gen Intel(R) Core(TM) i5-12450H |
| Memoria RAM | 16 GB |
| Sistema Operativo | Windows 11 |

## **Análisis**

Por temas de retorno decidimos usar Array List, ya que nos facilita el acceso a elementos con get\_element() y es de complejidad O(1), la implementación de este requerimiento tiene un orden lineal O(n). Esto se debe a que debemos iterar los Array List varias veces poder realizar filtros, añadir elementos y obtener elementos.

Este comportamiento se puede evidenciar experimentalmente en la gráfica. Por poco es una recta completa, indicando su crecimiento lineal.

# **Requerimiento 6**

## **Descripción**

El punto del requerimiento es conocer la estadística del departamento de mi interés para un rango de fechas de carga de los registros dado. Por este motivo, use una función auxiliar para que me dé un Array List filtrado entre fechas de carga que funciona con la librería datetime. Desde acá es más fácil recorrer de nuevo un Array List más pequeño y filtrarlo por departamento. En caso de que el número de registros sea menor o igual a 20 devuelvo una Pila, en caso contrario, devuelvo un Array List con acceso a elementos.

|  |  |
| --- | --- |
| **Entrada** |  |
| **Salidas** | 1. Tupla con pila, boolean, size, numero de registros tipo celsus, numero de registros tipo survey. 2. Tupla con Array List, boolean, size, numero de registros tipo celsus, numero de registros tipo survey. |
|  | Implementado, Iker Barbosa |

## **Análisis de complejidad**

Análisis de complejidad de cada uno de los pasos del algoritmo

|  |  |
| --- | --- |
| **Pasos** | **Complejidad** |
|  | O(n), la función auxiliar buscar\_entre\_fechas\_al() es una función que recorre todo el Array List para filtrar entre fechas de carga recibidas por parámetro. |
|  | O(n), recorrer la lista filtrada entre fechas de carga para hallar los registros que coincidan con el tipo de departamento que se recibió por parámetro. |
|  | O(n), si el tamaño de lista es menor o igual a 20, recorremos la lista para añadir sus elementos a una pila y retornarla. |
|  | O(n), en caso de que el tamaño de la lista sea mayor a 20. Creamos una nueva Array List con new\_list(), e iteramos la lista para añadir los primeros 5 elementos y los últimos 5 con get.element() y retornamos el Array List. |
| ***TOTAL*** | ***O(n)*** |

## **Pruebas Realizadas**

Las pruebas realizadas fueron realizadas en una maquina con las siguientes especificaciones. Los datos de entrada fueron el catálogo de datos, el estado de CALIFORNIA, la fecha de carga 2012-01-01 00:00:00 y la fecha de carga 2024-11-07 12:00:00 .

|  |  |
| --- | --- |
| Procesadores | AMD Ryzen 7 5700G |
| Memoria RAM | 22 GB |
| Sistema Operativo | Windows 10 |

|  |  |
| --- | --- |
| **Entrada** | **Tiempo (s)** |
| 100000 datos |  |
| 200000 datos |  |
| 300000 datos |  |
| 400000 datos |  |
| 500000 datos |  |

|  |  |
| --- | --- |
| Procesadores | 12th Gen Intel(R) Core(TM) i5-12450H |
| Memoria RAM | 16 GB |
| Sistema Operativo | Windows 11 |

|  |  |
| --- | --- |
| **Entrada** | **Tiempo (s)** |
| 100000 datos |  |
| 200000 datos |  |
| 300000 datos |  |
| 400000 datos |  |
| 500000 datos |  |

### **Tablas de datos**

Las tablas con la recopilación de datos de las pruebas.

|  |  |  |
| --- | --- | --- |
| **Muestra** | **Salida** | **Tiempo (ms)** |
| 100000 datos |  | 571.472 |
| 200000 datos |  | 1167.862 |
| 300000 datos |  | 1743.198 |
| 400000 datos |  | 2334.033 |
| 500000 datos |  | 2891.148 |

### **Graficas**

Las gráficas con la representación de las pruebas realizadas.

|  |  |
| --- | --- |
| Procesadores | AMD Ryzen 7 5700G |
| Memoria RAM | 22 GB |
| Sistema Operativo | Windows 10 |

|  |  |
| --- | --- |
| Procesadores | 12th Gen Intel(R) Core(TM) i5-12450H |
| Memoria RAM | 16 GB |
| Sistema Operativo | Windows 11 |

## **Análisis**

Similar al requerimiento 4, decidimos usar Array List, ya que nos facilita el acceso a elementos con get\_element() y es de complejidad O(1), la implementación de este requerimiento tiene un orden lineal O(n). Esto se debe a que debemos iterar los Array List varias veces poder realizar filtros, añadir elementos y obtener elementos.

Este comportamiento se puede evidenciar perfectamente en la gráfica, especialmente en la del Ryzen 7 5700G. Una recta completa, lo que indica su crecimiento lineal.

# **Requerimiento 7**

## **Descripción**

El punto del requerimiento es conocer el año con mayor y menor ingresos para un departamento de interés en un rango de fechas de recolección de los registros dado. Para esta consulta, solo se deben tener en cutna los registros que contengan un “$”.

|  |  |
| --- | --- |
| **Entrada** |  |
| **Salidas** |  |
| **Implementado (Sí/No)** | Implementado, Iker Barbosa. Camilo Castro |

|  |  |
| --- | --- |
| **Pasos** | **Complejidad** |
|  | O(n), buscar\_entre\_anios() es una función que se usa para recorrer la lista completa y filtrar entre dos años que recibe por parámetro. El retorno de esta función se guarda en la variable lista. |
|  | O(n), la variable lista\_filtrada guarda el valor de retorno de la función remover\_value\_lista(). Esta función recibe un departamento por parametro y se encarga de filtrar por departamento, guarda solamente las funciones que contengan $ en su unidad de medida y también ignora los que contienen $ y value (D) |
|  | O(1), guarda en la variable total\_registros el tamaño de la lista\_filtrada calculada anteriormente. |
|  | O(n), se crea un diccionario para poder almacenar los ingresos por año. Se recorre la lista filtrada y se extraen los años para que sean llaves de diccionario. Diccionarios de listas enlazadas. También cuenta con las llaves lista, que almacena las listas enlazadas del año en cuestión, total\_ingresos, que almacena la sumatoria de ingresos, census, que almacena el conteo de registros tipo CENSUS, survey, que almacena el conteo de registros tipo SURVEY, validos, almacena el conteo de registros validos, no\_validos, que almacena el conteo de registros no\_validos.  La inserción en la última posición es O(1) |
|  | Una vez creados los diccionarios, se recorren y se accede a la llave lista. El recorrido de esta llave es O(n), ya que se recorren listas enlazadas por años dentro del diccionario y se hacen los respectivos incrementos al propio diccionario. Como el número de años es fijo, la complejidad completa es O(n), ya que se recorren solamente una vez.  La función es\_numero() O(1), solamente verifica si el valor del registro se puede convertir a flotante, si retorna True, cuenta como registro valido. De lo contrario retorna false y se cuenta como registro no valido. |
|  | *O(n), ya que en el peor de los casos estas comparaciones se harían n veces si tenemos muchos años.* |
| ***TOTAL*** | ***O(n)*** |

## **Pruebas Realizadas**

Las pruebas realizadas fueron realizadas en una maquina con las siguientes especificaciones. Los datos de entrada fueron el catálogo de datos, CALIFORNIA, 1970, 2024.

|  |  |
| --- | --- |
| Procesadores | 12th Gen Intel(R) Core(TM) i5-12450H |
| Memoria RAM | 16 GB |
| Sistema Operativo | Windows 11 |

|  |  |
| --- | --- |
| **Entrada** | **Tiempo (s)** |
| 100000 datos |  |
| 200000 datos |  |
| 300000 datos |  |
| 400000 datos |  |
| 500000 datos |  |

# **Requerimiento 8**

## **Descripción**

El punto del requerimiento es conocer el cuál es el departamento con mayor diferencia promedio de tiempo entre la recopilación de los registros y su carga a la plataforma de entre todos los registros. Para este requerimiento decidimos crear un diccionario de listas enlazadas en donde la llave del diccionario representa un departamento y la lista enlazada contiene la información necesaria para saber la mayor diferencia. Para el cálculo de diferencias hicimos uso de la librería Datetime.

|  |  |
| --- | --- |
| **Entrada** |  |
| **Salidas** | ret = (estado, promedio, año\_max, año\_min,diferencia\_max,  diferencia\_min, census, survey) |
| **Implementado (Sí/No)** | Implementado, Iker Barbosa |

## **Análisis de complejidad**

Análisis de complejidad de cada uno de los pasos del algoritmo

|  |  |
| --- | --- |
| **Pasos** | **Complejidad** |
| **Texto  El contenido generado por IA puede ser incorrecto.** | O(n), filtrar los registros para evitar añadir los valores que contengan “$” y (“D”). Se recorre la lista completa y se retorna la lista filtrada. |
| Texto  El contenido generado por IA puede ser incorrecto.  Texto  El contenido generado por IA puede ser incorrecto. | O(n), recorrer la lista para así crear un diccionario donde las llaves son cada departamento. Si no existe, crea la llave y crea una nueva Single Linked List con sl.new\_list(). En caso de que existe, crea un nodo con información,  O(1) para añadirlo al final en la Single Linked List con sl.add\_last(). |
| Interfaz de usuario gráfica, Texto, Chat o mensaje de texto  El contenido generado por IA puede ser incorrecto. | O(N), Se accede a las llaves del diccionario y se recorre la lista enlazada solamente una vez por estado hasta el final para realizar las comparaciones y sumatorias necesarias. |
|  |  |
| ***TOTAL*** | ***O(n)*** |

## **Análisis de complejidad**

Análisis de complejidad de cada uno de los pasos del algoritmo

|  |  |
| --- | --- |
| **Pasos** | **Complejidad** |
|  | O(n), filtrar los registros para evitar añadir los valores que contengan “$” y (“D”) |
|  | O(n), recorrer la lista para así crear un diccionario donde las llaves son cada departamento. Si no existe, crea la llave y crea una nueva Single Linked List con sl.new\_list(). En caso de que existe, crea un nodo con información, para añadirlo en la Single Linked List con sl.add\_last(). |
|  | O(k), el diccionario tiene un numero de estados fijo. Por lo tanto, se recorre k veces, siendo k una constante. |
|  | O(N), recorre la lista enlazada hasta el final para realizar las comparaciones y sumatorias necesarias. |
| ***TOTAL*** | ***O(n)*** |

## **Pruebas Realizadas**

Las pruebas realizadas fueron realizadas en una maquina con las siguientes especificaciones. Los datos de entrada fueron el catálogo de datos.

|  |  |
| --- | --- |
| Procesadores | AMD Ryzen 7 5700G |
| Memoria RAM | 22 GB |
| Sistema Operativo | Windows 10 |

|  |  |
| --- | --- |
| **Entrada** | **Tiempo (s)** |
| 100000 datos |  |
| 200000 datos |  |
| 300000 datos |  |
| 400000 datos |  |
| 500000 datos |  |

|  |  |
| --- | --- |
| Procesadores | 12th Gen Intel(R) Core(TM) i5-12450H |
| Memoria RAM | 16 GB |
| Sistema Operativo | Windows 11 |

|  |  |
| --- | --- |
| **Entrada** | **Tiempo (s)** |
| 100000 datos |  |
| 200000 datos |  |
| 300000 datos |  |
| 400000 datos |  |
| 500000 datos |  |

### **Tablas de datos**

Las tablas con la recopilación de datos de las pruebas.

|  |  |  |  |
| --- | --- | --- | --- |
| **Muestra** | **Salida** | **Tiempo (ms)** |  |
| 100000 datos |  | 1150.047 |  |
| 200000 datos |  | 2524.231 |  |
| 300000 datos |  | 3260.040 |  |
| 400000 datos |  | 4533.304 |  |
| 500000 datos |  | 5719.113 |  |

### **Graficas**

Las gráficas con la representación de las pruebas realizadas.

|  |  |
| --- | --- |
| Procesadores | AMD Ryzen 7 5700G |
| Memoria RAM | 22 GB |
| Sistema Operativo | Windows 10 |

|  |  |
| --- | --- |
| Procesadores | 12th Gen Intel(R) Core(TM) i5-12450H |
| Memoria RAM | 16 GB |
| Sistema Operativo | Windows 11 |

## **Análisis**

El requerimiento 8 presenta un crecimiento lineal. Simplemente se recorre por completo una primera vez para filtrar los datos, una segunda para crear diccionarios de listas enlazadas y luego se recorre el diccionario completo para poder hacer los cálculos necesarios.

Este comportamiento se puede evidenciar experimentalmente en la gráfica. Por su forma y su pendiente es una función lineal. Por lo que cumple con el comportamiento lineal esperado.

Por la cantidad de variables, comparaciones y sumatorias que se deben realizar aumenta su tiempo de ejecución, comparado a otros requerimientos.