



Guion de prácticas

Shopping3

Pair, Index

Abril de 2021



Metodología de la Programación

DGIM-GII-GADE

Curso 2020/2021

Índice

1. Descripción	5
1.1. Los datos	5
1.2. Arquitectura de la práctica (recordatorio)	5
2. Shopping3, práctica a entregar	7
2.1. Descripción de la práctica	9
2.2. Un caso de ejemplo	12
2.3. Tests completos de la práctica	15
2.4. Configuración de la práctica	17
2.5. Entrega de la práctica	18
3. TESTS REPORT OF PROJECT shopping3a	19
4. Ejemplo de uso de las clases Pair e Index	23

1. Descripción

Recordemos que todas las prácticas de este año están orientadas al problema de ventas por internet, analizando el registro de actividad de los clientes de una web de venta de productos, y elaborando informes de comportamiento de las ventas.

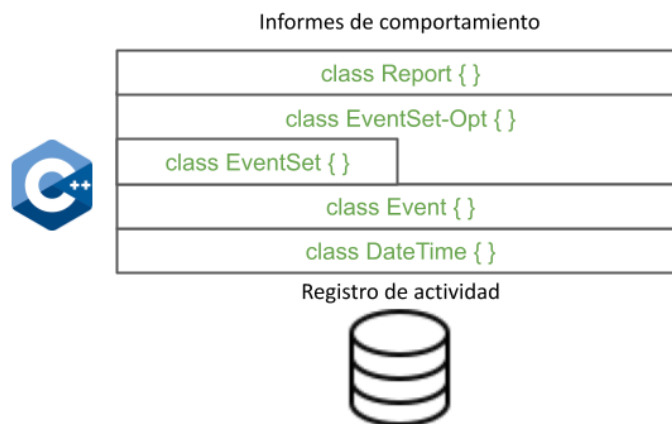


Figura 1: Arquitectura de las prácticas de MP 2021

Para ello, se va a implementar una arquitectura de clases (ver Figura 1) que se irán desarrollando progresivamente a lo largo de la asignatura, desde las más primitivas hasta las más abstractas.

En las prácticas anteriores (Shopping1 y shopping2) ya se implementaron las clases `DateTime`, `Event` y `EventSet`, así como un programa principal para integrar y usar dichas clases en varios conjuntos de datos de prueba. En esta práctica, se van a desarrollar las clases `Pair` e `Index`, que nos permitirán hacer búsquedas más eficientes sobre nuestros conjuntos de datos. Además, se van a visualizar los resultados con la biblioteca `DataTable`, la cual se os proporciona junto con la práctica.

1.1. Los datos

En esta práctica seguiremos trabajando con un pequeño fragmento del conjunto de datos reales (alrededor de 75,000 registros) descargados desde la plataforma Kaggle¹. En la Figura 2 se encuentra un recordatorio del formato de dichos datos en el dataset. Recuerda que un evento puede tener algunos valores vacíos. Los únicos datos que siempre estarán presentes serán `DateTime`, `Product ID`, `UserID`, y `SessionID`.

1.2. Arquitectura de la práctica (recordatorio)

La práctica *Shopping* se ha diseñado como una arquitectura por capas, en la que las capas más internas de la misma representan las estructuras más sencillas, sobre las cuales se asientan las capas más externas,

¹Kaggle ([Abrir en navegador →](#))



Registro de actividad

Date Time	Event Type	Product ID	Category ID	Category Code	Brand	Price	User ID	Session ID
string	string	string	string	string	string	double	string	string

Figura 2: Estructura del registro de actividad

```

5
2019-10-01 00:15:06 UTC, cart, 5869134, 1783999064136745198, , cosmoprofi, 6.35, 554342223, 0b974342-1a53-41c1-a426-23130e770f4b
2019-10-01 00:17:10 UTC, cart, 5787018, 1487580006644188066, , , 6.33, 554342223, 0b974342-1a53-41c1-a426-23130e770f4b
2019-10-01 00:21:02 UTC, cart, 5836843, 1487580009261432856, , pnb, 0.71, 554342223, 0b974342-1a53-41c1-a426-23130e770f4b
2019-10-01 00:22:24 UTC, cart, 5755171, 1487580009387261981, , , 2.48, 554342223, 0b974342-1a53-41c1-a426-23130e770f4b
2019-10-01 00:22:56 UTC, cart, 5691026, 1487580009387261981, , , 2.86, 554342223, 0b974342-1a53-41c1-a426-23130e770f4b

```

Figura 3: Contenido del fichero de datos
./tests/validation/Ecommerce5.keyboard

con las estructuras más complejas. La Figura 1 muestra el diseño de la arquitectura que se va a emplear y que se va a ir desarrollando progresivamente en esta y las siguientes sesiones de prácticas. Durante el desarrollo de la asignatura se implementarán un total de 7 clases que dotarán al programa de la funcionalidad deseada. Concretamente las clases que nos vamos a encontrar y que tendremos que implementar son:

A DateTime.cpp

Implementa la clase DateTime, compuesta por año, mes, día, hora, minutos y segundos. Toda instancia de esta clase ha de ser correcta, esto es, una fecha ha de ser válida y el tiempo dentro de los rangos correctos.

B Event.cpp

Implementa la clase Event, que contiene la información de cada acción registrada en el sitio web.

C EventSet.cpp

Implementa la estructura para almacenar un conjunto de eventos. Como comprobaremos muchos menos que los deseables. Inicialmente usaremos arrays estáticos.

D Pair.cpp

Estructura para almacenar una clave de búsqueda y una posición en el EventSet.

E Index.cpp

Array de claves, que se va a utilizar como índice para la búsqueda y recuperación eficiente en el EventSet.

C, E EventSet-Op.cpp

Manteniendo la interfaz anterior, se cambia la implementación para alojarlas de forma más eficiente en memoria dinámica.

F Report.cpp

Implementa la estructura para almacenar una matriz que nos va a permitir realizar estadísticas y comparativas.

Este trabajo progresivo se ha planificado en hitos sucesivos con entregas en Prado.

2. Shopping3, práctica a entregar

Tras el desarrollo de las prácticas anteriores (Shopping1 y Shopping2) tenemos las tres primeras capas (ver Figura 1) de nuestra arquitectura desarrollada. El programa desarrollado hasta el momento nos permite hacer consultas sencillas sobre nuestro conjunto de datos. Así en la Shopping2 pudimos filtrar nuestro conjunto de datos por tipo de evento para consultar únicamente los eventos asociados a una compra, y obtener el precio total para cada una de las marcas que habían realizado alguna venta. Sin embargo, conforme nuestro conjunto de datos aumenta de tamaño, trabajar con él se vuelve más lento y agota más recursos. Por ello, a la hora de realizar informes avanzados o un análisis más elaborado de los datos, es interesante disponer de estructuras adicionales que nos permitan trabajar con nuestros datos de una forma más eficiente.

Índices. ¿Qué son?

Como habéis podido comprobar, en los archivos de ejemplo con los que hemos trabajado hasta el momento, nuestros eventos están ordenados por fecha; cada vez que se produce un evento nuevo, en el servidor se añade la información de dicho evento al final del archivo de registros.

No obstante, cuando queremos extraer estadísticas de ventas (*purchase*), consultas (*view*) etc., normalmente no estamos interesados en la información temporal (organizada por `DateTime`) si no que, pasan a ser más relevantes otros campos de nuestro histórico de eventos. En particular, en nuestra aplicación tenemos un gran interés en generar informes agrupados por usuarios (`UserID`) o agrupados por marcas (`Brand`) que nos permitan extraer conclusiones y poder hacer promociones de marcas y recomendaciones a usuarios.

Por otro lado, muchas consultas solo hacen referencia a una pequeña parte de los registros de un archivo. Por ejemplo, eventos relativos a un determinado usuario o relativos a una marca en particular. No es eficiente que tengamos que leer todos los eventos para encontrar los eventos del `userID = 446419295` o los eventos asociados a `Brand = runail`. Lo más adecuado sería que pudiéramos localizar directamente estos registros. Para facilitar esta forma de acceso vamos a usar índices.

Un índice, para nuestro conjunto de eventos, es una estructura de datos que funciona como el índice de cualquier libro. Si se quiere buscar información concreta, se busca en el índice ordenado alfabéticamente, lo que facilita la búsqueda, y dicho índice nos indica la(s) página(s) donde encontrar la información buscada. Además, el índice es mucho más pequeño que el libro, con lo que el esfuerzo de búsqueda es aún menor.

El índice que vamos a diseñar juega el mismo papel que el índice de un libro. Por ejemplo, vamos a almacenar una lista de marcas, ordenadas en orden alfabético, junto con cada una de las entradas en el conjunto de eventos donde está implicada cada marca. Así, cuando se requiere una marca determinada, se busca de forma eficiente en el índice ² para localizar las entradas en el conjunto de eventos y poder recuperar rápidamente los registros de dicha marca.

De esta forma, un índice está formado por un conjunto de registros, que va a ser consultado por un campo, la clave de búsqueda, *key*, el string por el que está ordenado y nos va a devolver unas posiciones *Pos*, uno por valor de clave, donde recuperar los registros de datos en el conjunto de eventos. Inicialmente, vamos a considerar un registro en el índice por cada evento en el conjunto de eventos, esto es, si n es el tamaño del conjunto de eventos, n es el número de registros en el índice ³. Cada registro (le llamaremos *Pair*) del índice, está compuesto por un campo string, *key* y un entero *pos* la posición dentro del conjunto de eventos, un valor entre 0 y $n - 1$.

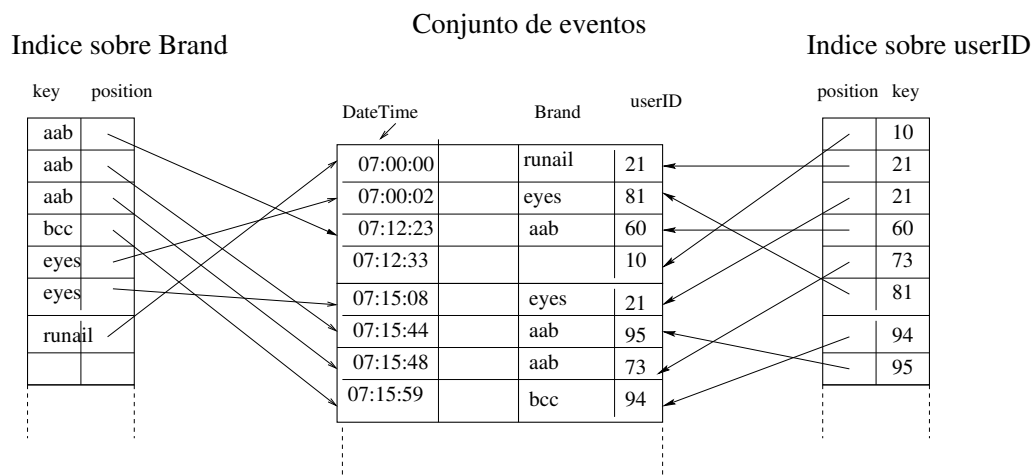


Figura 4: Ejemplo de índices sobre userID y sobre Brand apartir de un conjunto de Eventos.

Todo lo comentado sobre marcas se puede aplicar igualmente sobre la cadena `userID`, pudiendo montarse uno o varios índices sobre nuestro conjunto de eventos como puede verse en la figura 4. Un índice como los que vamos a utilizar en esta práctica, aún ordenado por clave, puede tener valores duplicados. Es decir, pueden existir varios

²Se puede utilizar la búsqueda binaria sobre las cadenas de marca (Brand), ya que existe un orden entre las cadenas, ejemplo: $Abc < Abd$.

³Esto solo es cierto para `userID`, ya que este campo es obligatorio, pero en el caso de Brand, algunos eventos tienen valor desconocido y no se almacena ese valor como *key* en el índice. Esa es la razón por la que la longitud de un índice sobre Brand puede ser menor que n .

Pair con el mismo valor de clave, como se puede comprobar en la mencionada figura. Existirá un valor por cada evento del mismo usuario, o de la misma marca, en instantes diferentes. Cuando se crea el índice, se mantiene el orden por valor de clave, pero a igualdad de clave el nuevo *Pair* se inserta antes de un valor de clave superior.⁴ Para saber cómo se monta un índice sobre un conjunto de eventos, consulta la tabla 3 con los datos del conjunto presentados en la tabla 2.

Por tanto, las dos clases que desarrollamos en esta práctica están enfocadas a la implementación de un índice sobre `UserID` y de un índice sobre `Brand` para facilitar las consultas de la información sobre los campos `UserID` y `Brand` respectivamente.

- **Clase `Pair`:** Esta clase representa un par `<key, position>`. La `key`, un *string*, hace referencia a la clave de búsqueda (`userID` o `Brand`) mientras que la `position`, un *int*, contiene la posición en el conjunto de eventos donde se encuentra el evento con la clave que andamos buscando.
- **Clase `Index`:** Esta clase contiene un array de objetos tipo `Pair`, y dos atributos: el número de entradas activas en el array (`_nEntries`) y un marcador booleano que nos indica si nuestro índice contiene elementos de tipo marca o de tipo usuario (`_onBrand`). La particularidad de nuestro índice es que está ordenado por los valores de la clave y va a contener valores de clave repetidos. La ventaja de la ordenación por la clave es que nos permite hacer búsquedas eficientes sobre el índice, que de otra manera serían muy lentas sobre el `EventSet` completo.

Además, a partir de esta práctica se utilizará la lectura y escritura directamente sobre ficheros, y comenzarán a utilizarse los argumentos de `main` para modificar el funcionamiento del programa. Se puede consultar más información sobre el uso de parámetros del `main` en el Anexo ???. La lectura y escritura de archivos viene detallada en el Anexo ??, y en el Anexo 4 se muestra un caso de ejemplo de las clases `Pair` e `Index` sobre un conjunto de datos sencillo.

2.1. Descripción de la práctica

Para ello, se deberá crear un proyecto NetBeans compuesto por los siguientes ficheros que ya deben estar completos de la práctica anterior:

- **`DateTime.h`**
Nuevo método `sameDay`, para comprobar si dos fechas pertenecen al mismo día
- **`DateTime.cpp`**

⁴O visto de otro modo, el nuevo `Pair` se inserta después del último valor de clave igual.

- **Event.h**

Nuevos métodos `write` and `read`, que permiten leer o escribir un evento desde un flujo de entrada o salida.

- **Event.cpp**

- **EventSet.h** Nuevos métodos `write` and `read`, que permiten leer o escribir un conjunto de eventos desde un flujo de entrada o salida.

- **EventSet.cpp**

- **Pair.h**

Revisar y completar las declaraciones de métodos y parámetros para que pueda funcionar correctamente: métodos `const`, parámetros `const` y parámetros por referencia.

- **Pair.cpp**

Implementar todos los métodos y funciones que faltan. Prestar atención al funcionamiento de los métodos y funciones descritos en los comentarios.

Además se implementará la función externa a la clase `Pair` siguiente:

- `bool equalKey()`; que comprueba si dos objetos de tipo `Pair` son iguales en el campo `Key`.

- **Index.h**

Revisar y completar las declaraciones de métodos y parámetros para que pueda funcionar correctamente: métodos `const`, parámetros `const` y parámetros por referencia.

- **Index.cpp**

Implementar todos los métodos y funciones que faltan. Prestar atención al funcionamiento de los métodos y funciones descritos en los comentarios.

También merecen especial mención las funciones siguientes externas a la clase `Index`:

- `Index rawFilterIndex()`; Extrae un subconjunto de `Index`, para el que el valor del campo especificado coincida con el valor indicado.
- `Event getEvent()`; Extrae el evento que ocupa la posición dada en el `Index` dado sobre el `EventSet`.
- `void write()`; Escribe en un fichero los eventos, en formato `csv`, que haya en el índice indicado.
- `float sumPrice()`; Calcula la suma total de las cantidades gastadas (atributo `Price`) en un conjunto de eventos dados por el índice pasado como parámetro.

■ main.cpp

Las lecturas y escrituras se realizarán directamente desde ficheros, en lugar de utilizar la redirección de entrada utilizada en las prácticas anteriores. Se utilizarán los argumentos del main para modificar el comportamiento del programa. En los comentarios del fichero `main.cpp` se describen con detalle los distintos argumentos que puede recibir nuestro programa. Los argumentos que aparecen entre corchetes son argumentos opcionales:

```
-input <input_file> -events <number_events> -index-by UserID|Brand  
[-report-by UserID|Brand|Type|DateTime -output <output_file>  
-display Framed|Fancy]
```

Para poder ejecutar el programa con una de las órdenes anteriores es necesario utilizar la terminal o modificar la orden de ejecución de las propiedades del proyecto (**Run - Run Command**, igual que se hacía para el redireccionamiento de la entrada). Por ejemplo, Si queremos utilizar el archivo de entrada `Ecommerce926.dataset`, leer 500 registros y hacer un índice por marcas, tendríamos que especificar lo siguiente:

```
"${OUTPUT_PATH}" -input ./data/ECommerce926.dataset  
-events 500 -index-by Brand
```

Además habrá que implementar las funciones siguientes:

- `void errorBreak()` ; Que detiene el programa si ocurre algún error (con los argumentos del programa, al abrir un archivo o al leer los datos del fichero)
- `void extractEventSet()` ; Que dado el conjunto de eventos completo y un índice filtrado, crea un nuevo conjunto de eventos solo con los eventos que están en el índice.

Y completar el código de la función `main` para que el programa tenga el comportamiento que se describe en los comentarios:

1. Procesar los argumentos del main: archivo de entrada, número de eventos a leer, campo sobre el que se realiza el índice (Brand o UserID), campo sobre el que hacer el informe (Brand, UserID, Type o DateTime, usando Brand por defecto ya que es un parámetro opcional), y archivo de salida (también opcional).
2. Utilizar un `EventSet` para almacenar una secuencia de eventos leídos desde el fichero de entrada indicado, leyendo el número de eventos especificados como argumento de entrada (observad que ahora el fichero de eventos no contiene el número de eventos al principio, si no que se especifica desde fuera).
3. Construir el índice sobre las marcas o los usuarios según corresponda.

4. Generar los informes sobre el EventSet siguientes (podéis utilizar la librería DataTable adjunta a la práctica de igual forma que en la práctica 2):
 - Actividad por cada día de la semana (al igual que en las prácticas anteriores, utilizando la función `computeActivity`).
 - Tráfico agregado por el campo indicado con el argumento `-report-by` (suponiendo que agregamos por marca, recuento del número de Eventos para cada marca en el EventSet). Además, almacenar el valor del campo que genera más tráfico, el número de eventos en los que está involucrado, y el índice que contiene dichos eventos, ya que lo necesitaremos más adelante.
 - Precio agregado por el campo indicado con el argumento `-report-by` (suponiendo que agregamos por marca, suma total de los Eventos para cada marca en el EventSet). Para imprimir la información de los distintos DataVector, hay tres funciones disponibles, que se utilizarán en función del parámetro de main `-display`:
 - `dataVector.showPlainReportV(0)` muestra el tipo de informe más simple, y es el que se utiliza por defecto
 - `dataVector.showFramedReportV(0)` muestra un informe de tipo tabla de texto, y es el que se utiliza cuando el parámetro display toma el valor `Framed`
 - `dataVector.showFancyReportV(0)` muestra un informe de tipo gráfico, y es el que se utiliza cuando el parámetro display toma el valor `Fancy`
5. Si se ha especificado un archivo de salida, guardar el eventSet que corresponde al valor del campo que más tráfico ha generado en dicho fichero.
6. Utilizando la función `REPORT_DATA2`, se reportan los siguientes informes para los tests de integración:
 - EventSet original, con todos los eventos leídos del archivo de datos
 - EventSet filtrado, que solo contiene los eventos indexados en el Index principal construido
 - DataVector con el informe de tráfico generado
 - DataVector con el informe de precios generado

2.2. Un caso de ejemplo

En esta práctica los ficheros en `/tests/validation/` contendrán los parámetros que serán necesarios para que el main pueda utilizar los distintos ficheros de datos que estarán en `/data/`. Así por ejemplo para poder leer del fichero `ECommerce926.dataset`, que contiene 926 eventos, 500 eventos y hacer un índice por marcas se deberá especificar en la redirección de entrada lo siguiente:



```
"${OUTPUT_PATH}" -input ./data/ECommerce926.dataset  
-events 500 -index-by Brand
```

Por lo que se podría comprobar que la salida obtenida sería:

```
Read 500 records from file ./data/ECommerce926.dataset  
Main Index by Brand  
299 pairs  
201 records could not be indexed  
amount :4926.5
```

Donde se indica que se han leído 500 eventos de `ECommerce926.dataset` y se ha construido un índice por marcas (`Brand`). Sin embargo no todos los eventos contienen una marca por lo que el índice solo contiene 299 pares (se podría utilizar el método `size` de `Index`) y 201 eventos no han sido indexados. El último número nos indica la suma de cantidades (`Price`) de los eventos indexados.

Si calculamos la actividad por día de la semana para los eventos indexados (utilizando `computeActivity()`), después visualizaremos cuántos registros hay en cada día de la semana, lo que produce el siguiente informe:

```
Activity found: SUNDAY(11) MONDAY(20) TUESDAY(73) WEDNESDAY(154)  
THURSDAY(17) FRIDAY(23) SATURDAY(1)
```

Si utilizamos un `DataVector` al que llamaremos `weeklyData`, al igual que hicimos en la práctica `shopping2` podremos usar las funciones `showFancyReportH`, `showFramedReportH` o `showPlainReportH` para obtener un informe como el siguiente:

Events/day							
11	20	73	154	17	23	1	299
SUNDAY	MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY	

donde además se indica el total de eventos (299) procesados.

A continuación, se usa la función `findUnique()` para extraer los valores únicos de la columna `Brand`. Se puede ver que son todos distintos ya que los que no tenían valor no se habían indexado:

```
art-visage  artex  beautix  ...  uskusi  yoko  zinger
```

A continuación se extrae, por cada marca de las anteriores encontradas, el número de veces que aparece cada marca (tráfico por marca) y se suman los precios de los productos que aparecen en cada marca, lo que produce los siguientes dos informes, que llamaremos `ammountData` y `hitsData`, de nuevo usando el tipo de datos `DataVector`. En total, se han vendido 4926.50 \$.

```
Brand (traffic)  
art-visage      2  
    artex       1  
    beautix     2
```



```
.  
. .  
uskusi      5  
  yoko      1  
  zinger     3  
  
299
```

```
Brand (amount)  
art-visage   9.36  
  artex     15.08  
  beautix   31.42  
.  
.  
.  
uskusi      19.05  
  yoko       7.83  
  zinger     48.15  
  
4926.50
```

Además calcularemos qué marca es la que ha conseguido un mayor tráfico (número de visitas) en la web:

```
Highest traffic choice (runail) with 45 hits
```

Finalmente, el programa llama a `REPORT_DATA2` con cada variable interesada, produciendo la salida siguiente, la cual muestra el nombre de la variable, el número de elementos que contiene y el código encriptado de su contenido para que las herramientas de testeo puedan comprobar si se han hecho bien las operaciones internas.

```
Highest traffic choice (runail) with 45 hits  
  
[originalES(Original read from dataset)] 500 18248242769591794958  
  
[filteredES(Events indexed)] 299 3323149042970569627  
  
[hitsData( Report traffic)] 56 art-visage( 2.00) artex( 1.00) beautix( 2.00)  
beauty-free( 4.00) beautyblender( 4.00) bioaqua( 7.00) bpw.style( 1.00)  
browxenna( 2.00) cnd( 4.00) consly( 5.00) depilflax( 1.00) dermal( 4.00)  
domix( 1.00) emil( 1.00) enas( 1.00) enigma( 1.00) estel( 5.00) estelare( 1.00)  
eunyul( 5.00) farmona( 3.00) farmstay( 1.00) grattol( 6.00) haruyama(12.00)  
i-laq( 4.00) ingarden(27.00) irisk(16.00) italwax( 2.00) jas( 2.00)  
jessnail( 1.00) kaaral( 6.00) kapous( 6.00) keen( 2.00) kinetics( 2.00)  
kosmekka( 4.00) lebelage( 1.00) lianail( 1.00) lovely( 5.00) lowence( 4.00)  
marathon( 7.00) masura(14.00) max(11.00) milv( 1.00) oniq(16.00) osmo( 1.00)  
pnb( 2.00) provoc( 3.00) runail(45.00) shik( 2.00) solomeya( 1.00) strong( 6.00)  
sunuv( 1.00) thuya( 8.00) uno(15.00) uskusi( 5.00) yoko( 1.00) zinger( 3.00)  
  
[ammountData( Report ammount)] 56 art-visage( 9.36) artex(15.08) beautix(31.42)  
beauty-free(18.80) beautyblender(99.04) bioaqua(54.46) bpw.style( 0.79)  
browxenna(29.68) cnd(34.28) consly(23.50) depilflax(10.54) dermal( 4.12)  
domix( 4.24) emil(55.56) enas( 3.54) enigma( 6.90) estel(27.45) estelare( 3.02)  
eunyul(35.55) farmona(80.73) farmstay( 0.63) grattol(25.14) haruyama(47.64)  
i-laq(15.88) ingarden(174.99) irisk(129.85) italwax(26.04) jas(51.42)  
jessnail( 8.73) kaaral(59.22) kapous(21.90) keen(19.56) kinetics(18.74)  
kosmekka(128.88) lebelage( 3.75) lianail( 6.33) lovely(34.93) lowence(69.84)  
marathon(964.46) masura(57.88) max(579.86) milv( 3.17) oniq(139.68) osmo( 9.71)  
pnb( 7.38) provoc(19.71) runail(63.61) shik(20.64) solomeya( 1.75)  
strong(1172.22) sunuv(33.33) thuya(252.06) uno(134.48) uskusi(19.05) yoko( 7.83)  
zinger(48.15)
```

2.3. Tests completos de la práctica

Se recomienda repasar primero los videotutoriales completos sobre este la metodología TDD en Google Drive

([Abrir en navegador →](#))

y la preparación del entorno de trabajo con MPTTest

([Abrir en navegador →](#))

Esta tercera práctica mantiene los tres niveles de testeo, que incluyen tanto tests unitarios como de integración y que incluyen los tests de la práctica anterior más los nuevos tests de esta práctica (Ver Sección ??).

- Nivel Básico: 27 tests
- Nivel Intermedio: 10 tests
- Nivel Avanzado: 22 tests. Varios de ellos hacen un chequeo de memoria adicional, para comprobar que ningún vector de datos se sale más allá de sus límites.

Y este debería ser el resultado del test de la aplicación satisfaciendo todos los tests que se han diseñado (en azul aparece la llamada a los tests desde una terminal del proyecto).

```
make test

[=====] Running 59 tests from 3 test suites.
[-----] Global test environment set-up.
[-----] 27 tests from _01_Basics
[ RUN      ] _01_Basics.DateTime_Constructors
[ OK       ] _01_Basics.DateTime_Constructors (5 ms)
[ RUN      ] _01_Basics.DateTime_getters
[ OK       ] _01_Basics.DateTime_getters (6 ms)
[ RUN      ] _01_Basics.DateTime_set
[ OK       ] _01_Basics.DateTime_set (2 ms)
[ RUN      ] _01_Basics.DateTime_sameDay
[ OK       ] _01_Basics.DateTime_sameDay (5 ms)
[ RUN      ] _01_Basics.Event_ConstructorBase
[ OK       ] _01_Basics.Event_ConstructorBase (4 ms)
[ RUN      ] _01_Basics.Event_Setters_getters
[ OK       ] _01_Basics.Event_Setters_getters (9 ms)
[ RUN      ] _01_Basics.EventSet_Constructor
[ OK       ] _01_Basics.EventSet_Constructor (3 ms)
[ RUN      ] _01_Basics.EventSet_add_event
[ OK       ] _01_Basics.EventSet_add_event (5 ms)
[ RUN      ] _01_Basics.EventSet_add_line
[ OK       ] _01_Basics.EventSet_add_line (4 ms)
[ RUN      ] _01_Basics.EventSet_at_basic
[ OK       ] _01_Basics.EventSet_at_basic (4 ms)
[ RUN      ] _01_Basics.Pair_Constructors
[ OK       ] _01_Basics.Pair_Constructors (4 ms)
[ RUN      ] _01_Basics.Pair_isEmpty
[ OK       ] _01_Basics.Pair_isEmpty (2 ms)
[ RUN      ] _01_Basics.Pair_setters
[ OK       ] _01_Basics.Pair_setters (1 ms)
[ RUN      ] _01_Basics.Pair_getters
[ OK       ] _01_Basics.Pair_getters (0 ms)
[ RUN      ] _01_Basics.Index_Constructors
[ OK       ] _01_Basics.Index_Constructors (1 ms)
[ RUN      ] _01_Basics.Index_getIONWhich
[ OK       ] _01_Basics.Index_getIONWhich (1 ms)
[ RUN      ] _01_Basics.Index_clear
[ OK       ] _01_Basics.Index_clear (2 ms)
[ RUN      ] _01_Basics.Integrated_5_records
[ MEMCHECK ] ECommerce5-valgrind
[ OK       ] ECommerce5-valgrind
[ OK       ] _01_Basics.Integrated_5_records (28 ms)
[ RUN      ] _01_Basics.Integrated_30_records
[ MEMCHECK ] ECommerce30-valgrind
[ OK       ] ECommerce30-valgrind
[ OK       ] _01_Basics.Integrated_30_records (24 ms)
[ RUN      ] _01_Basics.Integrated_41_records
[ MEMCHECK ] ECommerce41-valgrind
[ OK       ] ECommerce41-valgrind
[ OK       ] _01_Basics.Integrated_41_records (31 ms)
[ RUN      ] _01_Basics.Integrated_162_records
[ MEMCHECK ] ECommerce162-valgrind
[ OK       ] ECommerce162-valgrind
[ OK       ] _01_Basics.Integrated_162_records (37 ms)
[ RUN      ] _01_Basics.Integrated_926_records
[ MEMCHECK ] ECommerce926-valgrind
```



```
[ OK ] ECommerce926-valgrind
[ OK ] _01_Basics.Integrated_926_records (139 ms)
[ RUN ] _01_Basics.Integrated_Args_5_records
[ MEMCHECK ] ECommerce5-valgrind
[ OK ] ECommerce5-valgrind
[ OK ] _01_Basics.Integrated_Args_5_records (27 ms)
[ RUN ] _01_Basics.Integrated_Args_30_records
[ MEMCHECK ] ECommerce30-valgrind
[ OK ] ECommerce30-valgrind
[ OK ] _01_Basics.Integrated_Args_30_records (28 ms)
[ RUN ] _01_Basics.Integrated_Args_41_records
[ MEMCHECK ] ECommerce41-valgrind
[ OK ] ECommerce41-valgrind
[ OK ] _01_Basics.Integrated_Args_41_records (28 ms)
[ RUN ] _01_Basics.Integrated_Args_162_records
[ MEMCHECK ] ECommerce162-valgrind
[ OK ] ECommerce162-valgrind
[ OK ] _01_Basics.Integrated_Args_162_records (29 ms)
[ RUN ] _01_Basics.Integrated_Args_926_records
[ MEMCHECK ] ECommerce926-valgrind
[ OK ] ECommerce926-valgrind
[ OK ] _01_Basics.Integrated_Args_926_records (134 ms)
[-----] 27 tests from _01_Basics (563 ms total)

[-----] 10 tests from _02_Intermediate
[ RUN ] _02_Intermediate.DateTime_isBefore
[ OK ] _02_Intermediate.DateTime_isBefore (4 ms)
[ RUN ] _02_Intermediate.DateTime_weekDay
[ OK ] _02_Intermediate.DateTime_weekDay (4 ms)
[ RUN ] _02_Intermediate.Event_getField
[ OK ] _02_Intermediate.Event_getField (4 ms)
[ RUN ] _02_Intermediate.EventSet_add_event_partial
[ OK ] _02_Intermediate.EventSet_add_event_partial (4 ms)
[ RUN ] _02_Intermediate.EventSet_at_intermediate
[ OK ] _02_Intermediate.EventSet_at_intermediate (5 ms)
[ RUN ] _02_Intermediate.Index_3x3_just_build
[ OK ] _02_Intermediate.Index_3x3_just_build (4 ms)
[ RUN ] _02_Intermediate.Index_B_BxU_build_at
[ OK ] _02_Intermediate.Index_B_BxU_build_at (7 ms)
[ RUN ] _02_Intermediate.Index_U_BxU_build_at
[ OK ] _02_Intermediate.Index_U_BxU_build_at (4 ms)
[ RUN ] _02_Intermediate.Integrated_EMPTY
[ MEMCHECK ] EMPTY-valgrind
[ OK ] EMPTY-valgrind
[ OK ] _02_Intermediate.Integrated_EMPTY (44 ms)
[ RUN ] _02_Intermediate.Integrated_ErrorLoading
[ MEMCHECK ] ErrorLoading-valgrind
[ OK ] ErrorLoading-valgrind
[ OK ] _02_Intermediate.Integrated_ErrorLoading (40 ms)
[-----] 10 tests from _02_Intermediate (121 ms total)

[-----] 22 tests from _03_Advanced
[ RUN ] _03_Advanced.DateTime_BadValues
[ OK ] _03_Advanced.DateTime_BadValues (8 ms)
[ RUN ] _03_Advanced.Event_setType_Bad_Values
[ OK ] _03_Advanced.Event_setType_Bad_Values (3 ms)
[ RUN ] _03_Advanced.Event_Others_Bad_Values
[ OK ] _03_Advanced.Event_Others_Bad_Values (4 ms)
[ RUN ] _03_Advanced.EventSet_add_event_full
[ OK ] _03_Advanced.EventSet_add_event_full (6 ms)
[ RUN ] _03_Advanced.EventSet_at_advanced
[ OK ] _03_Advanced.EventSet_at_advanced (4 ms)
[ RUN ] _03_Advanced.EventSet_externalfunctions
[ OK ] _03_Advanced.EventSet_externalfunctions (5 ms)
[ RUN ] _03_Advanced.EventSet_write
[ OK ] _03_Advanced.EventSet_write (3 ms)
[ RUN ] _03_Advanced.EventSet_read
[ OK ] _03_Advanced.EventSet_read (2 ms)
[ RUN ] _03_Advanced.Index_U_BxU_bounds
[ OK ] _03_Advanced.Index_U_BxU_bounds (1 ms)
[ RUN ] _03_Advanced.Index_B_BxU_bounds
[ OK ] _03_Advanced.Index_B_BxU_bounds (1 ms)
[ RUN ] _03_Advanced.Index_add
[ OK ] _03_Advanced.Index_add (2 ms)
[ RUN ] _03_Advanced.Index_B_BxU_rawFilterIndex
[ OK ] _03_Advanced.Index_B_BxU_rawFilterIndex (1 ms)
[ RUN ] _03_Advanced.Index_U_BxU_rawFilterIndex
[ OK ] _03_Advanced.Index_U_BxU_rawFilterIndex (2 ms)
[ RUN ] _03_Advanced.Index_Type_BxU_rawFilterIndex
[ OK ] _03_Advanced.Index_Type_BxU_rawFilterIndex (2 ms)
[ RUN ] _03_Advanced.Index_DateTime_BxU_rawFilterIndex
[ OK ] _03_Advanced.Index_DateTime_BxU_rawFilterIndex (2 ms)
[ RUN ] _03_Advanced.Index_BxU_sumPrice
[ OK ] _03_Advanced.Index_BxU_sumPrice (5 ms)
[ RUN ] _03_Advanced.Integrated_ErrorData
[ MEMCHECK ] ErrorData-valgrind
[ OK ] ErrorData-valgrind
[ OK ] _03_Advanced.Integrated_ErrorData (34 ms)
[ RUN ] _03_Advanced.Integrated_ErrorSaving
[ MEMCHECK ] ErrorSaving-valgrind
[ OK ] ErrorSaving-valgrind
[ OK ] _03_Advanced.Integrated_ErrorSaving (36 ms)
[ RUN ] _03_Advanced.Integrated_Args_no_open
[ MEMCHECK ] ECommerce926-valgrind
[ OK ] ECommerce926-valgrind
[ OK ] _03_Advanced.Integrated_Args_no_open (34 ms)
[ RUN ] _03_Advanced.Integrated_Args_error_data
[ MEMCHECK ] ECommerce162-valgrind
[ OK ] ECommerce162-valgrind
[ OK ] _03_Advanced.Integrated_Args_error_data (38 ms)
```




```
[ RUN      ] _03_Advanced.Integrated_Args_error_missing_arg
[ MEMCHECK ] ECommerce162-valgrind
[ OK       ] ECommerce162-valgrind
[ MEMCHECK ] ECommerce162-valgrind
[ OK       ] ECommerce162-valgrind
[ OK       ] _03_Advanced.Integrated_Args_error_missing_arg (61 ms)
[ RUN      ] _03_Advanced.Integrated_Args_error_bad_arg
[ MEMCHECK ] ECommerce162-valgrind
[ OK       ] ECommerce162-valgrind
[ OK       ] _03_Advanced.Integrated_Args_error_bad_arg (30 ms)
[-----] 22 tests from _03_Advanced (288 ms total)

[-----] Global test environment tear-down
[=====] 59 tests from 3 test suites ran. (972 ms total)
[ PASSED ] 59 tests.
```

2.4. Configuración de la práctica

Para esta práctica se puede seguir con la misma configuración de la práctica anterior pero con las siguientes novedades:

- Se incluirán en la carpeta **tests** los nuevos tests correspondientes a las clases **Pair** e **Index**, así como los nuevos tests de integración.
- Se incluirán los ficheros correspondientes a las clases **Pair** e **Index** a las carpetas **include** y **src**.
- Se deberá añadir la carpeta **include** de la librería **DataTable** mediante **Project Properties - C++ Compiler - Include Directories**. Opcionalmente se podrá añadir también la librería **AnsiTerminal**, incluyendo también su carpeta **include**. Revisar que ya están también incluidos los directorios **include** siguientes:
 - del propio proyecto **./include**
 - las de testeo **../MPTest/include**
 - las de googletest **../googletest-master/googletest/include**
- Se modificará el fichero **main.cpp** de acuerdo a las instrucciones en la documentación del fichero de la práctica.
- Recordar añadir desde la vista lógica del proyecto.
 1. En **Header Files, Add existing item** añadir los ficheros **Pair.h** e **Index.h**.
 2. En **Source Files, Add existing item** añadir los ficheros **Pair.cpp** e **Index.cpp**.
 3. En **Test Files, Add existing item** añadir los nuevos ficheros de los tests.

La práctica deberá ser entregada en Prado, en la fecha que se indica en cada entrega, y consistirá en un fichero ZIP del proyecto en su estado actual (es decir con todas las clases implementadas) incluyendo las clases implementadas para la práctica anterior.



2.5. Entrega de la práctica

Una vez terminada la práctica que, al menos, haya superado los tests básicos, se debe hacer un zip (se sugiere utilizar la script `runZipProject.sh`) excluyendo las carpetas `./dist/`, `./build/`, `./nbproject/private/`, `./doc/html/` y `./dos/latex/` y subirla a Prado antes de la fecha de cierre de la entrega.



3. TESTS REPORT OF PROJECT shopping3a

This is the list of tests of the project

ID	NAME	DESCRIPTION
1::1	.01_Basics.DateTime_Constructors	A newly create instance of DateTime gives the default date
1::2	.01_Basics.DateTime_Constructors	A newly created instance of DateTime by using a string gives the same string
2::1	.01_Basics.DateTime_getters	The year of the default datetime must be 2021
2::2	.01_Basics.DateTime_getters	The month of the default datetime must be 1
2::3	.01_Basics.DateTime_getters	The day of the default datetime must be 1
2::4	.01_Basics.DateTime_getters	The hour of the default datetime must be 0
2::5	.01_Basics.DateTime_getters	The minutes of the default datetime must be 0
2::6	.01_Basics.DateTime_getters	The seconds of the default datetime must be 0
3::1	.01_Basics.DateTime_set	Setting an instance of DateTime with a valid string gives the same date
4::1	.01_Basics.DateTime_sameDay	Two DateTimes with different initializations give different sameDay
4::2	.01_Basics.DateTime_sameDay	When a DateTime is used to initialize a second DateTime, both must be sameDay
5::1	.01_Basics.Event_ConstructorBase	A newly created instance of Event gives the default event
5::2	.01_Basics.Event_ConstructorBase	A newly create instance of Event, initialized with a string, gives the same string
6::1	.01_Basics.Event_Setters_getters	Setting the datetime of an event to TODAY gives TODAY
6::2	.01_Basics.Event_Setters_getters	Setting the type of an event to a good value X gives X
6::3	.01_Basics.Event_Setters_getters	Setting the productid of an event to a good value X gives X
6::4	.01_Basics.Event_Setters_getters	Setting the categoryID of an event to a good value X gives X
6::5	.01_Basics.Event_Setters_getters	Setting the category code of an event to a good value X gives X
6::6	.01_Basics.Event_Setters_getters	Setting the brand of an event to a good value X gives X
6::7	.01_Basics.Event_Setters_getters	Setting the price of an event to a good value X gives X
7::1	.01_Basics.EventSet_Constructor	A newly created instance of EventSet must have size = 0
7::2	.01_Basics.EventSet_Constructor	A newly created instance of EventSet must have a to_string empty '0'
8::1	.01_Basics.EventSet_add_event	Adding an event to a newly created instance of EventSet must have size = 1
8::2	.01_Basics.EventSet_add_event	Adding the default event to an empty event set must have a to_string equal to the default event
8::3	.01_Basics.EventSet_add_event	Adding 1 event to a filled EventSet increases its size in 1
9::1	.01_Basics.EventSet_add_line	Adding a line to a newly created instance of EventSet must have size = 1
9::2	.01_Basics.EventSet_add_line	Adding the default event, as a line, to an empty event set must have a to_string equal to the default event
9::3	.01_Basics.EventSet_add_line	Adding 1 event, as a line, to a filled EventSet increases its size in 1
10::1	.01_Basics.EventSet_at_basic	Querying the event at the 0 position should match with the first event added to the EventSet
10::2	.01_Basics.EventSet_at_basic	Querying the event at the middle position should match with the event added which was added at that point
11::1	.01_Basics.Pair_Constructors	After the creation of a new instance of Pair, it shows its default value
11::2	.01_Basics.Pair_Constructors	After the creation of a new instance of Pair by using a string and an int, it shows these two same values
11::3	.01_Basics.Pair_Constructors	After the creation of a new instance of Pair by using a string and NO int, it shows the string and -1 default value
12::1	.01_Basics.Pair_isEmpty	Any newly created instance of Pair, without specifying any arguments, is considered to be empty
12::2	.01_Basics.Pair_isEmpty	Any instance of Pair with valid key and pos, could not be considered empty
13::1	.01_Basics.Pair_setters	After the creation of a new instance of Pair and setting it by using a string and an int, it shows these two same values
13::2	.01_Basics.Pair_setters	After the creation of a new instance of Pair and setting the key to a given string, it shows this same string
13::3	.01_Basics.Pair_setters	After the creation of a new instance of Pair and setting the value to a given number, it shows this same number



ID	NAME	DESCRIPTION
14::1	.01_Basics.Pair_getters	After the creation of a new instance of Pair and setting it by using a string and an int, getKey() returns the same string
14::2	.01_Basics.Pair_getters	After the creation of a new instance of Pair and setting it by using a string and an int, getPost() returns the same integer
14::3	.01_Basics.Pair_getters	After the creation of a new instance of Pair and setting the key to a given string, getKey() returns this same string
14::4	.01_Basics.Pair_getters	After the creation of a new instance of Pair and setting the value to a given number, getPos() returns the same number
15::1	.01_Basics.Index_Constructors	A newly created instance of Index must have size = 0
15::2	.01_Basics.Index_Constructors	A newly created instance of Index must write as ": On brand? 0"
16::1	.01_Basics.Index_getOnWhich	Building an index by Users reports 0
16::2	.01_Basics.Index_getOnWhich	Building an index by brand reports 1
17::1	.01_Basics.Index_clear	Clearing a brand new instance of Index leaves it empty
17::2	.01_Basics.Index_clear	Clearing an instance of Index with existing records leaves it empty
18::1	.01_Basics.Integrated_5_records	Execution of a simple test with five records only
18::2	.01_Basics.Integrated_5_records	The output file of this test with five records only is not as expected, please have a look at ./tests/validation/ECommerce5-hit.dataset
19::1	.01_Basics.Integrated_30_records	Execution of a test with 30 records only
20::1	.01_Basics.Integrated_41_records	Execution of a simple test with 41 records
20::2	.01_Basics.Integrated_41_records	The output file of this test with 41 records is not as expected, please have a look at ./tests/validation/ECommerce41-hit.dataset
21::1	.01_Basics.Integrated_162_records	Execution of a simple test with 162 records
21::2	.01_Basics.Integrated_162_records	The output file of this test with 162 records is not as expected, please have a look at ./tests/validation/ECommerce162-hit.dataset
22::1	.01_Basics.Integrated_926_records	Execution of a simple test with 926 records
22::2	.01_Basics.Integrated_926_records	The output file of this test with 926 records is not as expected, please have a look at ./tests/validation/ECommerce926-hit.dataset
23::1	.01_Basics.Integrated_Args_5_records	Execution of a simple test with five records only + mainargs
23::2	.01_Basics.Integrated_Args_5_records	The output file of this test with five records only is not as expected, please have a look at ./tests/validation/ECommerce5-hit.dataset
24::1	.01_Basics.Integrated_Args_30_records	Execution of a test with 30 records only + mainargs
25::1	.01_Basics.Integrated_Args_41_records	Execution of a simple test with 41 records only + mainargs
25::2	.01_Basics.Integrated_Args_41_records	The output file of this test with 41 records only is not as expected, please have a look at ./tests/validation/ECommerce41-hit.dataset
26::1	.01_Basics.Integrated_Args_162_records	Execution of a simple test with 162 records only + mainargs
26::2	.01_Basics.Integrated_Args_162_records	The output file of this test with 162 records only is not as expected, please have a look at ./tests/validation/ECommerce162-hit.dataset
27::1	.01_Basics.Integrated_Args_926_records	Execution of a simple test with 500 records only + mainargs
27::2	.01_Basics.Integrated_Args_926_records	The output file of this test with 500 records only is not as expected, please have a look at ./tests/validation/ECommerce926-hit.dataset
28::1	.02_Intermediate.DateTime_isBefore	A DateTime cannot be before itself
28::2	.02_Intermediate.DateTime_isBefore	Default date is before today
28::3	.02_Intermediate.DateTime_isBefore	Today is not before the Default date
29::1	.02_Intermediate.DateTime_weekDay	Today must be Thursday
29::2	.02_Intermediate.DateTime_weekDay	Tomorrow is Friday
29::3	.02_Intermediate.DateTime_weekDay	Yesterday was Wednesday
30::1	.02_Intermediate.Event_getField	getField("DateTime") on any Event, must be equal to getDateTime()
30::2	.02_Intermediate.Event_getField	In any row, querying the Event for field "Price" will produce the same string than getDateTime
30::3	.02_Intermediate.Event_getField	In any case, querying the Event for field "UserID" will produce the same string than the former getUserID()
30::4	.02_Intermediate.Event_getField	In any row, querying the EventS a mandatory field cannot give an empty string
30::5	.02_Intermediate.Event_getField	In any row, querying an optional field will produce the same string than the former event added to the EvenSet



ID	NAME	DESCRIPTION
31::1	.02.Intermediate.EventSet.add_event_partial	Adding MAXEVENT events to a newly created EventSet increases its size in MAXEVENT
32::1	.02.Intermediate.EventSet.at.intermediate	Querying the event at the last position should match with the last event added to the EventSet
32::2	.02.Intermediate.EventSet.at.intermediate	Accessing EventSet at() a certain position and changing the user ID of the event, this change will remain in the EventSet itself
32::3	.02.Intermediate.EventSet.at.intermediate	Accessing EventSet at() a certain position and changing the brand of the event, this change will remain in the EventSet itself
33::1	.02.Intermediate.Index_3x3.just_build	In any case, building the same index twice gives the same result
33::2	.02.Intermediate.Index_3x3.just_build	The index by user built within experiment 3x3 differs from the expected one
33::3	.02.Intermediate.Index_3x3.just_build	The index by brand built within experiment 3x3 differs from the expected one
34::1	.02.Intermediate.Index_B.BxU.build.at	0 The index by brand built within experiment BxU is the same key during U consecutive positions
34::2	.02.Intermediate.Index_B.BxU.build.at	0 The index by brand built within experiment BxU two consecutive indexes of the same brand differ in B units
34::3	.02.Intermediate.Index_B.BxU.build.at	1 The index by brand built within experiment BxU is the same key during U consecutive positions
34::4	.02.Intermediate.Index_B.BxU.build.at	1 The index by brand built within experiment BxU two consecutive indexes of the same brand differ in B units
34::5	.02.Intermediate.Index_B.BxU.build.at	2 The index by brand built within experiment BxU is the same key during U consecutive positions
34::6	.02.Intermediate.Index_B.BxU.build.at	2 The index by brand built within experiment BxU two consecutive indexes of the same brand differ in B units
34::7	.02.Intermediate.Index_B.BxU.build.at	3 The index by user built within experiment BxU changes the key at point U
34::8	.02.Intermediate.Index_B.BxU.build.at	4 The index by brand built within experiment BxU is the same key during U consecutive positions
34::9	.02.Intermediate.Index_B.BxU.build.at	4 The index by brand built within experiment BxU two consecutive indexes of the same brand differ in B units
34::10	.02.Intermediate.Index_B.BxU.build.at	5 The index by brand built within experiment BxU is the same key during U consecutive positions
34::11	.02.Intermediate.Index_B.BxU.build.at	5 The index by brand built within experiment BxU two consecutive indexes of the same brand differ in B units
34::12	.02.Intermediate.Index_B.BxU.build.at	6 The index by brand built within experiment BxU is the same key during U consecutive positions
34::13	.02.Intermediate.Index_B.BxU.build.at	6 The index by brand built within experiment BxU two consecutive indexes of the same brand differ in B units
34::14	.02.Intermediate.Index_B.BxU.build.at	7 The index by user built within experiment BxU changes the key at point U
34::15	.02.Intermediate.Index_B.BxU.build.at	8 The index by brand built within experiment BxU is the same key during U consecutive positions
34::16	.02.Intermediate.Index_B.BxU.build.at	8 The index by brand built within experiment BxU two consecutive indexes of the same brand differ in B units
34::17	.02.Intermediate.Index_B.BxU.build.at	9 The index by brand built within experiment BxU is the same key during U consecutive positions
34::18	.02.Intermediate.Index_B.BxU.build.at	9 The index by brand built within experiment BxU two consecutive indexes of the same brand differ in B units
34::19	.02.Intermediate.Index_B.BxU.build.at	10 The index by brand built within experiment BxU is the same key during U consecutive positions
34::20	.02.Intermediate.Index_B.BxU.build.at	10 The index by brand built within experiment BxU two consecutive indexes of the same brand differ in B units
34::21	.02.Intermediate.Index_B.BxU.build.at	11 The index by user built within experiment BxU changes the key at point U
35::1	.02.Intermediate.Index_U.BxU.build.at	0 The index by user built within experiment BxU is the same key during B consecutive positions
35::2	.02.Intermediate.Index_U.BxU.build.at	0 The index by user built within experiment BxU indexes B consecutive positions for the same key
35::3	.02.Intermediate.Index_U.BxU.build.at	1 The index by user built within experiment BxU is the same key during B consecutive positions
35::4	.02.Intermediate.Index_U.BxU.build.at	1 The index by user built within experiment BxU indexes B consecutive positions for the same key
35::5	.02.Intermediate.Index_U.BxU.build.at	2 The index by user built within experiment BxU is the same key during B consecutive positions
35::6	.02.Intermediate.Index_U.BxU.build.at	2 The index by user built within experiment BxU indexes B consecutive positions for the same key



ID	NAME	DESCRIPTION
35::7	.02.Intermediate.Index.U.BxU.build.at	3 The index by user built within experiment BxU changes the key at point B
35::8	.02.Intermediate.Index.U.BxU.build.at	4 The index by user built within experiment BxU is the same key during B consecutive positions
35::9	.02.Intermediate.Index.U.BxU.build.at	4 The index by user built within experiment BxU indexes B consecutive positions for the same key
35::10	.02.Intermediate.Index.U.BxU.build.at	5 The index by user built within experiment BxU is the same key during B consecutive positions
35::11	.02.Intermediate.Index.U.BxU.build.at	5 The index by user built within experiment BxU indexes B consecutive positions for the same key
35::12	.02.Intermediate.Index.U.BxU.build.at	6 The index by user built within experiment BxU is the same key during B consecutive positions
35::13	.02.Intermediate.Index.U.BxU.build.at	6 The index by user built within experiment BxU indexes B consecutive positions for the same key
35::14	.02.Intermediate.Index.U.BxU.build.at	7 The index by user built within experiment BxU changes the key at point B
35::15	.02.Intermediate.Index.U.BxU.build.at	8 The index by user built within experiment BxU is the same key during B consecutive positions
35::16	.02.Intermediate.Index.U.BxU.build.at	8 The index by user built within experiment BxU indexes B consecutive positions for the same key
35::17	.02.Intermediate.Index.U.BxU.build.at	9 The index by user built within experiment BxU is the same key during B consecutive positions
35::18	.02.Intermediate.Index.U.BxU.build.at	9 The index by user built within experiment BxU indexes B consecutive positions for the same key
35::19	.02.Intermediate.Index.U.BxU.build.at	10 The index by user built within experiment BxU is the same key during B consecutive positions
35::20	.02.Intermediate.Index.U.BxU.build.at	10 The index by user built within experiment BxU indexes B consecutive positions for the same key
35::21	.02.Intermediate.Index.U.BxU.build.at	11 The index by user built within experiment BxU changes the key at point B
36::1	.02.Intermediate.Integrated.EMPTY	Execution of a simple test with an empty set of records
37::1	.02.Intermediate.Integrated.ErrorLoading	Execution of a simple test with an error when opening the data file
38::1	.03.Advanced.DateTime.BadValues	Setting a bad date or time gives the default datetime
38::2	.03.Advanced.DateTime.BadValues	Setting a bad date or time gives the default datetime
38::3	.03.Advanced.DateTime.BadValues	Setting a bad date or time gives the default datetime
38::4	.03.Advanced.DateTime.BadValues	Setting a bad date or time gives the default datetime
38::5	.03.Advanced.DateTime.BadValues	Setting a bad date or time gives the default datetime
38::6	.03.Advanced.DateTime.BadValues	Setting a bad date or time gives the default datetime
38::7	.03.Advanced.DateTime.BadValues	Setting a bad date or time gives the default datetime
38::8	.03.Advanced.DateTime.BadValues	Setting a bad date or time gives the default datetime
38::9	.03.Advanced.DateTime.BadValues	Setting a bad date or time gives the default datetime
38::10	.03.Advanced.DateTime.BadValues	Setting a date with the incorrect format throws an exception
39::1	.03.Advanced.Event.setType.Bad.Values	Setting the type of an event to a good value X gives X
39::2	.03.Advanced.Event.setType.Bad.Values	Setting the type of an event to a good value X gives X
40::1	.03.Advanced.Event.Others.Bad.Values	Setting the productid of an event to a good value X gives X
40::2	.03.Advanced.Event.Others.Bad.Values	Setting the productid of an event to a good value X gives X
40::3	.03.Advanced.Event.Others.Bad.Values	Setting the productid of an event to a good value X gives X
41::1	.03.Advanced.EventSet.add_event_full	Adding one single Event events to a partly filled EventSet must return 1
41::2	.03.Advanced.EventSet.add_event_full	Adding one single Event events to a completely filled EventSet does not produce any change in the EventSet
41::3	.03.Advanced.EventSet.add_event_full	Adding one single Event events to a completely filled EventSet must return 0
42::1	.03.Advanced.EventSet.at.advanced	Querying an EventSet beyond its legal limits gives an EMPTY event
42::2	.03.Advanced.EventSet.at.advanced	Querying an EventSet beyond its legal limits gives an EMPTY event
42::3	.03.Advanced.EventSet.at.advanced	Querying an EventSet within its legal limits always gives a NON-EMPTY event
43::1	.03.Advanced.EventSet.externalfunctions	sumPrice must give the sum of all prices of a given EventSet, and it doesnt
43::2	.03.Advanced.EventSet.externalfunctions	UniqueBrands() must give the number of different brands, but it doesnt



ID	NAME	DESCRIPTION
43::3	.03.Advanced.EventSet.externalfunctions	UniqueUsers() must give the number of different users, but it doesnt
44::1	.03.Advanced.EventSet.write	Any EventSet writes into a disk file exactly its to.string() without the number of events
45::1	.03.Advanced.EventSet.read	Any EventSet read from a disk file must have a to.string(), without the number of events, comparable to the content of the file
46::1	.03.Advanced.Index.U.BxU.bounds	0 The index by user built within experiment BxU lower bounds User u with $u \cdot B$
46::2	.03.Advanced.Index.U.BxU.bounds	0 The index by user built within experiment BxU upper bounds User u with $(u+1) \cdot B$
47::1	.03.Advanced.Index.B.BxU.bounds	0 The index by brand built within experiment BxU lower bounds Brand b with $b \cdot U$
47::2	.03.Advanced.Index.B.BxU.bounds	0 The index by brand built within experiment BxU upper bounds Brand b with $(b+1) \cdot U$
48::1	.03.Advanced.Index.add	Inserting any pair in any index increases its size in one
48::2	.03.Advanced.Index.add	Adding a key at the end of the index places the new key in its previous upperbound
48::3	.03.Advanced.Index.add	Adding a key at the beginning of the index gives a lower bound equals to 0 after the insertion
49::1	.03.Advanced.Index.B.BxU.rawFilterIndex	In an Index BxU , filtering the index by brand-i produces an index with exactly size U
50::1	.03.Advanced.Index.U.BxU.rawFilterIndex	In an Index BxU , filtering the index by user-i produces an index with exactly size B
51::1	.03.Advanced.Index.Type.BxU.rawFilterIndex	In an Index BxU , filtering the index by Type=cart produces an index with exactly size B
52::1	.03.Advanced.Index.DateTime.BxU.rawFilterIndex	In an Index BxU , filtering the index by DateTime equals to the first available date produces an index with exactly size B
53::1	.03.Advanced.Index.BxU.sumPrice	Given a BxU Index, a filter by user 0 the sum of prices amounts to $(UB)/4(1+B)/2$
53::2	.03.Advanced.Index.BxU.sumPrice	Given a BxU Index, a filter by brand 0 the sum of prices amounts to $U \cdot (b+1)$
53::3	.03.Advanced.Index.BxU.sumPrice	Given a BxU Index, a filter by Type=view the sum of prices amounts to $(BU)/4(1+B)/2$
53::4	.03.Advanced.Index.BxU.sumPrice	Given a BxU Index, a filter by date=2021-03-1 10:32:00 UTC the sum of prices amounts to $U \cdot \text{day}$
54::1	.03.Advanced.Integrated.ErrorData	Execution of a simple test with an error when reading the data file
55::1	.03.Advanced.Integrated.ErrorSaving	Execution of a simple test with an error when saving the data file
56::1	.03.Advanced.Integrated.Args.no.open	Execution of a simple test with 500 records only + mainargs
57::1	.03.Advanced.Integrated.Args.error_data	Execution of a simple test with 500 records only + mainargs
58::1	.03.Advanced.Integrated.Args.error.missing_arg	Execution of a simple test with 500 records only + mainargs
58::2	.03.Advanced.Integrated.Args.error.missing_arg	Execution of a simple test with 500 records only + mainargs
59::1	.03.Advanced.Integrated.Args.error.bad_arg	Execution of a simple test with 500 records only + mainargs

4. Ejemplo de uso de las clases Pair e Index

En esta sección se ejemplifica el uso de las clases Pair e Index. Supongamos que estamos interesados en extraer información sobre el usuario con código 446419295. En particular, estamos interesados en saber qué cantidad de eventos ha generado dicho usuario, así como el precio total de dichos eventos, y trabajamos con el conjunto de datos que se muestra a continuación:



DateTime	EventType	ProductID	CategoryID	Cat Code	Brand	Price	UserID	Session ID
2019-10-01 00:15:06 UTC	cart	5869134	1783999064136745198		cosmoprofi	6.35	554342223	0b974342-1a53-41c1-a426-23130e770f4b
2019-10-01 00:17:10 UTC	cart	5787018	1487580006644188066			6.33	554342223	0b974342-1a53-41c1-a426-23130e770f4b
2019-10-01 00:21:02 UTC	cart	5836843	1487580009261432856		pnb	0.71	554342223	0b974342-1a53-41c1-a426-23130e770f4b
2019-10-01 00:29:21 UTC	cart	5857007	1487580009496313889		runail	3.17	554342223	0b974342-1a53-41c1-a426-23130e770f4b
2019-10-01 00:36:22 UTC	view	7652	1487580008657454075			2.86	446419295	d5e0977b-b6ac-43a1-b173-63e9742e95b2
2019-10-01 00:37:27 UTC	view	39918	1487580008657454075			2.86	446419295	d5e0977b-b6ac-43a1-b173-63e9742e95b2
2019-10-01 00:38:33 UTC	view	5582544	1958278551207674674			2.7	446419295	d5e0977b-b6ac-43a1-b173-63e9742e95b2
2019-10-01 00:39:45 UTC	view	7652	1487580008657454075			2.86	446419295	d5e0977b-b6ac-43a1-b173-63e9742e95b2
2019-10-01 00:43:06 UTC	view	5802490	1487580009471148064		shary	5.08	554342223	0b974342-1a53-41c1-a426-23130e770f4b
2019-10-01 00:44:00 UTC	cart	5643943	1487580010100293687			1.05	554342223	0b974342-1a53-41c1-a426-23130e770f4b
2019-10-01 00:44:31 UTC	cart	5809921	1487580010100293687		grattol	2.06	554342223	0b974342-1a53-41c1-a426-23130e770f4b
2019-10-01 00:45:03 UTC	cart	5875322	1487580010100293687			0.92	554342223	0b974342-1a53-41c1-a426-23130e770f4b
2019-10-01 00:56:43 UTC	view	5756762	1487580008909111303			3.02	446419295	d5e0977b-b6ac-43a1-b173-63e9742e95b2
2019-10-01 00:57:25 UTC	view	5697103	1487580008657454075			2.22	544418344	d5e0977b-b6ac-43a1-b173-63e9742e95b2
2019-10-01 03:18:29 UTC	remove_from_cart	5830552	1487580005671109489		masura	1.73	544418344	513b523a-be03-4bcb-85a1-ee0c340170a8
2019-10-01 03:18:35 UTC	remove_from_cart	5796770	1487580004916134735			4.29	544418344	513b523a-be03-4bcb-85a1-ee0c340170a8
2019-10-01 03:32:00 UTC	view	5653177	1487580012994363565			9.84	429179215	125c71ea-435a-4ef3-9df8-bdcfad1c46a3
2019-10-01 03:37:12 UTC	purchase	5864600	1487580013011140782		italwax	1.24	429179215	125c71ea-435a-4ef3-9df8-bdcfad1c46a3
2019-10-01 03:37:12 UTC	purchase	6731	1487580007256556476			14.27	429179215	125c71ea-435a-4ef3-9df8-bdcfad1c46a3
2019-10-01 03:37:12 UTC	purchase	5770380	1487580012994363565		italwax	10.32	429179215	125c71ea-435a-4ef3-9df8-bdcfad1c46a3

Cuadro 2: Conjunto de eventos de referencia, sobre el que se va montar un índice sobre userID, como clave de búsqueda.

Si extrajásemos información del conjunto de datos completo, sin ningún tipo de estructura de índice, tendríamos que visitar las 20 filas para contar el número de eventos en los que interviene nuestro usuario. En este caso, dicho usuario participa en 5 eventos, y la cantidad total de dinero es $2,86 + 2,86 + 2,7 + 2,86 + 3,02 = 14,30$. Si construyésemos el índice a priori, el funcionamiento sería el siguiente. Comenzamos construyendo dicho índice por usuarios, que es el campo en el que estamos interesados:

- Comenzamos con el primer elemento de la tabla. Como el índice está vacío metemos el elemento al principio, y nos queda el siguiente índice:

Key	Position
55434223	0

- Continuamos recorriendo la tabla, como los primeros eventos pertenecen todos al mismo usuario, todos van entrando consecutivamente, hasta llegar al siguiente índice:

Key	Position
554342223	0
554342223	1
554342223	2
554342223	3

- Llegamos a un usuario nuevo, que al entrar en orden en nuestro índice, se coloca antes que el usuario que traíamos. De nuevo tenemos varios registros seguidos para el mismo usuario, que en este ejemplo metemos simultáneamente, hasta llegar al siguiente índice:



Key	Position
446419295	4
446419295	5
446419295	6
446419295	7
554342223	0
554342223	1
554342223	2
554342223	3

- Ahora, nos volvemos a encontrar con el usuario 554342223. Lo introducimos a continuación de los registros existentes para ese usuario, quedando así el índice:

Key	Position
446419295	4
446419295	5
446419295	6
446419295	7
554342223	0
554342223	1
554342223	2
554342223	3
554342223	8
554342223	9
554342223	10
554342223	11

- De nuevo encontramos registros del usuario 446419295. Los introducimos a continuación de los registros ya existentes para ese usuario, quedando así el índice:

Key	Position
446419295	4
446419295	5
446419295	6
446419295	7
446419295	12
554342223	0
554342223	1
554342223	2
554342223	3
554342223	8
554342223	9
554342223	10
554342223	11

- Nos encontramos a un usuario nuevo, cuyo ID es mayor que el



446419295 pero menor que el 554342223, por lo que se colocará entre los dos, quedando así el índice:

Key	Position
446419295	4
446419295	5
446419295	6
446419295	7
446419295	12
544418344	13
544418344	14
544418344	15
554342223	0
554342223	1
554342223	2
554342223	3
554342223	8
554342223	9
554342223	10
554342223	11

- Y finalmente, el último usuario, cuyo ID es menor que todos los demás, por lo que se coloca al principio:

Key	Position
429179215	16
429179215	17
429179215	18
429179215	19
446419295	4
446419295	5
446419295	6
446419295	7
446419295	12
544418344	13
544418344	14
544418344	15
554342223	0
554342223	1
554342223	2
554342223	3
554342223	8
554342223	9
554342223	10
554342223	11

Cuadro 3: Índice sobre userID, final obtenido a partir del conjunto de Eventos

Una vez construido el índice, podemos hacer búsquedas sobre nuestros usuarios de forma mucho más eficiente. El orden del índice nos aporta la ventaja de poder utilizar el algoritmo de búsqueda binaria. Este algoritmo funciona de la siguiente manera:

1. Tenemos una colección `array` en el que queremos localizar un elemento `elem`
2. Accedemos a la posición intermedia de `array` (`index=size/2`), y la comparamos con `elem`:
 - a) Si `array[index] == elem`, hemos encontrado el elemento y lo devolvemos
 - b) Si `array[index] > elem`, el elemento se tiene que encontrar en la mitad derecha del `array` (si estuviese en la colección)
 - c) Si `array[index] < elem`, el elemento se tiene que encontrar en la mitad izquierda del `array` (si estuviese en la colección)
3. En función del resultado anterior, devolvemos `index`, o repetimos la búsqueda con la mitad correspondiente (utilizando el vector desde el 0 hasta `index`, o desde `index` hasta el final)

Basadas en este algoritmo, dispondremos de dos funciones, que se llamarán `lower_bound` y `upper_bound`. El método `lower_bound` nos devolverá la primera posición en la que ocurre el elemento que estamos buscando, y `upper_bound` nos devolverá la primera posición en la que ocurre el elemento siguiente, para poder obtener todas las entradas del índice que corresponden al valor con el que estamos trabajando.

Aplicamos nuestra función `lower bound` para encontrar la primera ocurrencia del individuo 446419295. Comenzamos con la búsqueda binaria sobre el índice que hemos construido:

- Accedemos a la posición intermedia (posición 9 del índice). Nos encontramos con la pareja [544418344, 13]. Como el ID que vamos buscando es menor que el que hemos encontrado, continuamos la búsqueda en la mitad inferior del vector.
- Accedemos a la posición intermedia de la mitad inferior (posición 4). Encontramos la pareja [446419295, 4] que contiene al usuario que buscamos, pero aún no sabemos si es la primera ocurrencia. Comenzamos la búsqueda hacia atrás hasta encontrar un usuario distinto.
- Accedemos a la posición 3, en la que encontramos a la pareja [429179215, 19]. Como el id ha cambiado, sabemos que en la posición 4 del índice está la primera ocurrencia del usuario que nos interesa.

En total, hemos necesitado consultar 3 posiciones del índice para localizar el `lower bound`.

Aplicamos ahora la función `upper_bound`. Comenzamos la búsqueda binaria:

- Accedemos a la posición intermedia (posición 9 del índice). Nos encontramos con la pareja [544418344, 13]. Como el ID que vamos buscando es menor que el que hemos encontrado, continuamos la búsqueda en la mitad inferior del vector.
- Accedemos a la posición intermedia de la mitad inferior (posición 4). Encontramos la pareja [446419295, 4] que contiene al usuario que buscamos, pero aún no sabemos si es la última ocurrencia. Comenzamos la búsqueda hacia delante hasta encontrar un usuario distinto.
- Accedemos a la posición 5, en la que encontramos a la pareja [446419295, 5]. Pertenece al mismo usuario, continuamos.
- Accedemos a la posición 6, en la que encontramos a la pareja [446419295, 6]. Pertenece al mismo usuario, continuamos.
- Accedemos a la posición 7, en la que encontramos a la pareja [446419295, 7]. Pertenece al mismo usuario, continuamos.
- Accedemos a la posición 8, en la que encontramos a la pareja [446419295, 12]. Pertenece al mismo usuario, continuamos.
- Accedemos a la posición 9, en la que encontramos a la pareja [544419344, 13]. Pertenece a un usuario distinto, por lo que hemos encontrado el upper bound.

De esta forma, hemos localizado la primera y la última ocurrencia del usuario 446419295 en nuestro conjunto de eventos (que sabemos que están entre nuestras posiciones 4 y 9 del índice). Ahora, el total de eventos en los que está involucrado el usuario coincide con la diferencia entre el upper bound y el lower bound ($9 - 4 = 5$), y para calcular el precio total sólo tenemos que recorrer los elementos del índice entre el 4 y el 9 acceder a la posición del conjunto de eventos que se nos indica (posiciones 4, 5, 6, 7 y 12) del vector de eventos, extraer el precio de cada uno de ellos, y hacer la suma.

La eficiencia de un método frente al otro es fácilmente observable. Mientras que trabajando con el conjunto de eventos completo hemos tenido que recorrer sus 20 filas para asegurarnos de que encontrábamos todos los eventos en los que participaba nuestro usuario, utilizando el índice sólo hemos tenido que acceder a 10 posiciones del índice (3 cuando buscamos el lower bound y 7 cuando buscamos el upper bound), y a 5 posiciones del conjunto de eventos (las cinco posiciones que hemos obtenido en el índice para nuestro usuario). Estas diferencias serán mucho más notables cuando tengamos conjuntos de cientos o miles de elementos, ya que solo construiremos el índice una vez, pero haremos varias consultas sobre él.