

## Práctica 2. Estudio experimental de la eficiencia de dos algoritmos de Ordenación Divide y Vencerás, *Merge Sort* y *Quick Sort*

Departamento de Sistemas Informáticos y Computación. Universitat Politècnica de València

### 1. Objetivo de la práctica

Esta práctica tiene dos objetivos. Tras acabarla, por un lado, el alumno deberá ser capaz de comparar experimentalmente los tiempos de ejecución de los algoritmos rápidos de ordenación Divide y Vencerás (DyV), *Merge Sort* y *Quick Sort*. Y, por otro lado, deberá ser capaz de decidir cuál de ellos es el mejor para cada tipo de datos a ordenar, contrastando el efecto que la comparación de los datos individuales tiene en dichos tiempos.

### 2. Descripción del problema

Al elegir un algoritmo para la ordenación *in situ* de los datos de un array se debe de tener en cuenta:

- Su eficiencia expresada en función de la talla del problema. Esto lleva a elegir algoritmos de ordenación directa ( $O(n^2)$ ), como Inserción Directa, si el número de datos a ordenar es pequeño, y algoritmos rápidos ( $O(n * \log(n))$ ), como *Merge Sort* o *Quick Sort*, si el número de datos a ordenar es muy grande.
- El número y coste efectivo de las comparaciones y movimientos de datos (intercambios o copias) que realiza. Así, dado que *Merge Sort* realiza un número de comparaciones bastante menor que *Quick Sort* pero un número de movimientos bastante mayor...
  - será preferible usar *Merge Sort* cuando el coste de comparar los datos sea notorio y, como sucede en Java, el coste de moverlos resulte irrelevante por no afectar a los datos sino únicamente a sus referencias (*Pointer sorting*);
  - será preferible usar *Quick Sort* cuando el coste de mover los datos sea tanto o más relevante que el de compararlos porque, como sucede en C++, dicho movimiento sí afecta directamente a los datos.
- La estabilidad (o preservación del orden inicial de datos idénticos). Esto solo es irrelevante cuando la ordenación se hace según un único criterio (caso de los datos de tipo primitivo). Dado que *Merge Sort* es estable y *Quick Sort* no, *Merge Sort* será preferible para ordenar objetos comparables según diversos criterios.

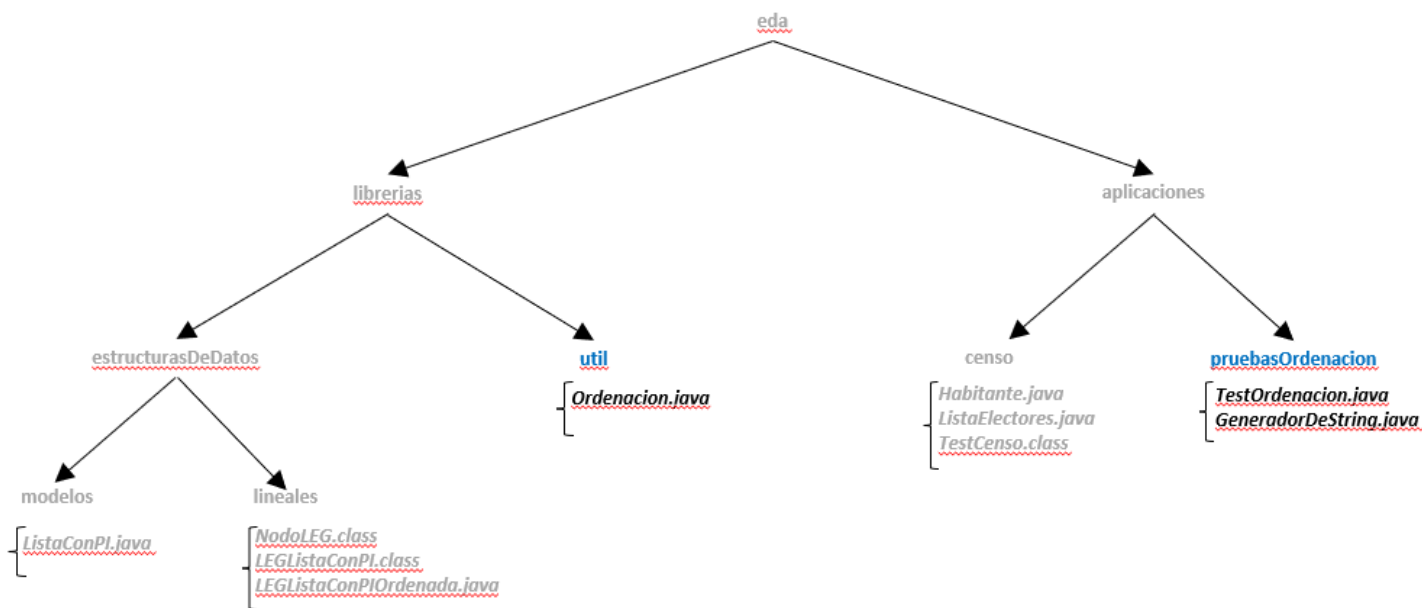
Este análisis permite entender la implementación, en el paquete `java.util`, para los métodos de ordenación de datos de tipo primitivo (por ejemplo, `static void Arrays.sort(int[] a)`) y la de objetos `Comparable` (por ejemplo, `static void Arrays.sort(Object[] a)`). Mientras que los primeros implementan *Quick Sort*, los segundos implementan *Merge Sort*, con el fin de garantizar la estabilidad y favorecer las situaciones en las que el coste efectivo de las comparaciones sea importante.

En esta práctica se implementará una versión eficiente de *Merge Sort* y se efectuará la comparación con la de *Quick Sort* vista en clase de teoría, analizando la influencia que sobre su coste temporal puede tener el incremento del coste de la comparación de elementos. Para ello, se usarán arrays de dos clases del estándar de Java que implementan la interfaz `Comparable`: `Integer` y `String`. Como se recordará, los valores de tipo `Integer` se comparan entre sí en tiempo constante, mientras que para los de tipo `String` el tiempo de la comparación individual depende del número de caracteres iniciales que tengan iguales los objetos a comparar. Así, inicializando adecuadamente los `String` a ordenar, se podrá comprobar cómo cambia el tiempo de ejecución de los dos algoritmos cuando crece el tiempo de ejecución de cada comparación individual.

### 3. Actividades a realizar

Antes de llevar a cabo las actividades propuestas, es necesario que el alumno actualice la estructura de paquetes y ficheros de su proyecto *BlueJ eda* siguiendo estos pasos:

- Ejecutar *BlueJ*.
  - Abrir el proyecto *eda*.
  - Abrir el paquete *aplicaciones* y crear en él un nuevo paquete llamado *pruebasOrdenacion*, que contendrá las clases de la aplicación a desarrollar en esta práctica.
  - Abrir el paquete *librerias* y crear en él un nuevo paquete de nombre *util*.
- Salir de *BlueJ* seleccionando la opción **Salir de la pestaña Proyecto**.
- Descargar los ficheros (.java) disponibles en *PoliformaT* en sus correspondientes directorios, tal como muestra la siguiente figura:



#### NOTA:

- La clase `Ordenacion` contiene métodos estáticos para ordenar arrays de datos de tipo genérico `Comparable`, y un método auxiliar, `sonIguales`, para comprobar si los arrays pasados como parámetros coinciden elemento a elemento.
- La clase `TestOrdenacion` contiene los métodos necesarios para llevar a cabo el análisis empírico de la eficiencia de los métodos de ordenación de la clase `Ordenacion` (entre otros, los de creación de los arrays de tallas distintas sobre los que se ejecutan los métodos a analizar y el de medición/tabulación de sus tiempos promedio de ejecución sobre dichos arrays).
- La clase `GeneradorDeString` es un generador aleatorio de objetos `String` con los primeros `n` caracteres iguales.

- Entrar en el proyecto *BlueJ eda* y situarse en su paquete *librerias.util*.

#### 3.1. Diseñar una implementación eficiente del algoritmo *Merge Sort*

La versión del algoritmo *Merge Sort* proporcionado en la clase `Ordenacion`, similar a la vista en teoría, es mejorable si en la operación de fusión o mezcla (método `merge1`) se evita copiar dos veces cada elemento del array a fusionar. Nótese que los elementos fusionados en este método se escriben en un array auxiliar `aux` (primeros bucles de `merge1`) y luego, se copian de nuevo al array a ordenar (en el último bucle de `merge1`).

En esta actividad el alumno debe añadir, en la clase `Ordenacion`, una nueva versión de *Merge Sort* que evite esa doble copia y, por tanto, sea más eficiente. Para ello, los pasos a seguir son:

1. Modificar el método de mezcla `merge1` para que devuelva el array auxiliar `aux`, con lo que tan solo se copiará cada elemento una vez. El perfil del nuevo método será:

```
private static <T extends Comparable<T>> T[] merge2(T[] v1, T[] v2)
```

2. Modificar el método `mergeSort1(T[], int, int)` para que use `merge2(T[], T[])` y, por tanto, devuelva el array resultante de la ordenación en lugar de `void`. El perfil del nuevo método será:

```
private static <T extends Comparable<T>> T[] mergeSort2(T[] v, int i, int f)
```

Además, para reducir el coste que supone la generación excesiva de arrays de un solo elemento, en su diseño se deben definir dos casos base de la recursión: uno para arrays de talla 1 y otro para arrays de talla 2.

3. Modificar el método `mergeSort1(T[])` para que use `mergeSort2(T[], int, int)` y, al mismo tiempo, mantenga el perfil típico de un método de ordenación *in situ*. Nótese que `mergeSort2(T[], int, int)` devuelve un array ordenado y el tipo del resultado de un método de ordenación *in situ* es `void`, no `T[]`. El perfil del nuevo método será:

```
public static <T extends Comparable<T>> void mergeSort2(T[] v)
```

### 3.2. Validar la nueva versión de *Merge Sort*

Es necesario verificar que la nueva versión de *Merge Sort* implementada, `mergeSort2(T[])`, ordena correctamente. Para ello, comprobaremos si su resultado, ante diferentes entradas, es el mismo que el conseguido por cualquier otro método que se sepa que es correcto.

En esta actividad el alumno debe completar el cuerpo del método `comprobar` de la clase `TestOrdenacion` para que, usando los métodos `quickSort` y `sonIguales` de la clase `Ordenacion`, compruebe si, en efecto, `mergeSort2(T[])` es un método correcto.

### 3.3. Comparar los tres métodos DyV de la clase `Ordenacion`

Para comparar los tres métodos DyV de la clase `Ordenacion`, hay que realizar el análisis experimental de su eficiencia. El método `temporizar` de la clase `TestOrdenacion` ya realiza dicho análisis para los métodos `quickSort` y `mergeSort1(T[])`. En concreto, las condiciones en las que `temporizar` realiza las medidas del tiempo de ejecución de ambos métodos son las siguientes:

- Los arrays a ordenar contienen valores de tipo `Integer` generados aleatoriamente, como resultado de invocar al método `crearAleatorioInteger`.
- Los tiempos de ejecución se miden para arrays de tallas (crecientes), desde 10.000 hasta 100.000 elementos con pasos de 10.000 en 10.000.
- Para garantizar resultados significativos, para cada talla se promedian los resultados de 200 ejecuciones, ordenando en cada una de ellas un array diferente.
- El reloj para medir los tiempos de ejecución es `nanoTime()`, método estático de la clase `java.lang.System` que devuelve el valor `long` actual del temporizador más preciso empleado en el sistema en nanosegundos (aunque la resolución real suele ser menor).

Como `temporizar` también tabula los tiempos medios de ejecución de `quickSort` y `mergeSort1(T[])`, para realizar esta actividad el alumno solo debe...

- Añadir al método `temporizar` las líneas de código correspondientes a la temporización de `mergeSort2(T[])`.
- Compilar `TestOrdenacion`, y ejecutar su método `temporizar`.
- Extraer las conclusiones correspondientes sobre cuál de los tres métodos analizados es el más rápido ordenando números enteros.

### 3.4. Comparar los tres métodos DyV de la clase Ordenación cuando crece el coste de la comparación individual

En esta actividad el alumno ha de usar la clase `GeneradorDeString` que permite generar objetos de tipo `String` con sus primeros `n` caracteres iguales. Por ejemplo, para generar dos `String` con sus 50 primeros caracteres iguales, las instrucciones son:

```
GeneradorDeString g = new GeneradorDeString(50);
String ejem1 = g.generar();
String ejem2 = g.generar();
// ejem1 y ejem2 tienen iguales sus 50 primeros caracteres
```

Teniendo en cuenta este ejemplo, el alumno debe modificar el método `crearAleatorioString`, de la clase `TestOrdenacion`, para que devuelva un array de `talla` objetos `String` generados aleatoriamente con sus primeros `n` caracteres iguales.

Hecho esto, se debe repetir la Actividad 3.3, aunque esta vez para arrays de tipo `String` con los 50 primeros caracteres iguales y usando el método `temporizarString` de la clase `TestOrdenacion`...

¿Cuál de los tres métodos de ordenación DyV de la clase `Ordenacion` será en este caso el más rápido? ¿Por qué?

## 4. Otras actividades

Si lo desea, el alumno puede resolver alguna de las siguientes actividades para continuar analizando los problemas planteados en esta práctica.

- Repetir la última temporización realizada pero usando ahora arrays de tipo `String` que tengan, respectivamente, sus primeros 10 y 80 caracteres iguales. Interpretar luego los resultados obtenidos.
- Repetir los estudios de coste realizados añadiendo en una última columna los resultados obtenidos usando el método `sort(Object[] a)` de la clase `java.util.Arrays`.
- Ajustar y representar gráficamente los resultados del análisis experimental de costes (hecho en las anteriores actividades). El alumno ha de utilizar `gnuplot` para ajustar (mediante el comando `fit`) y representar gráficamente (mediante el comando `plot`) tanto los resultados de la ejecución del método `temporizar` como sus ajustes.<sup>1</sup>

---

<sup>1</sup>Si fuera necesario, puede consultar el documento “Resumen de gnuplot” disponible en *PoliformaT*, que contiene los comandos más habituales de `gnuplot` así como un ejemplo de una sesión con su uso.