Tema 12. Colecciones en Java

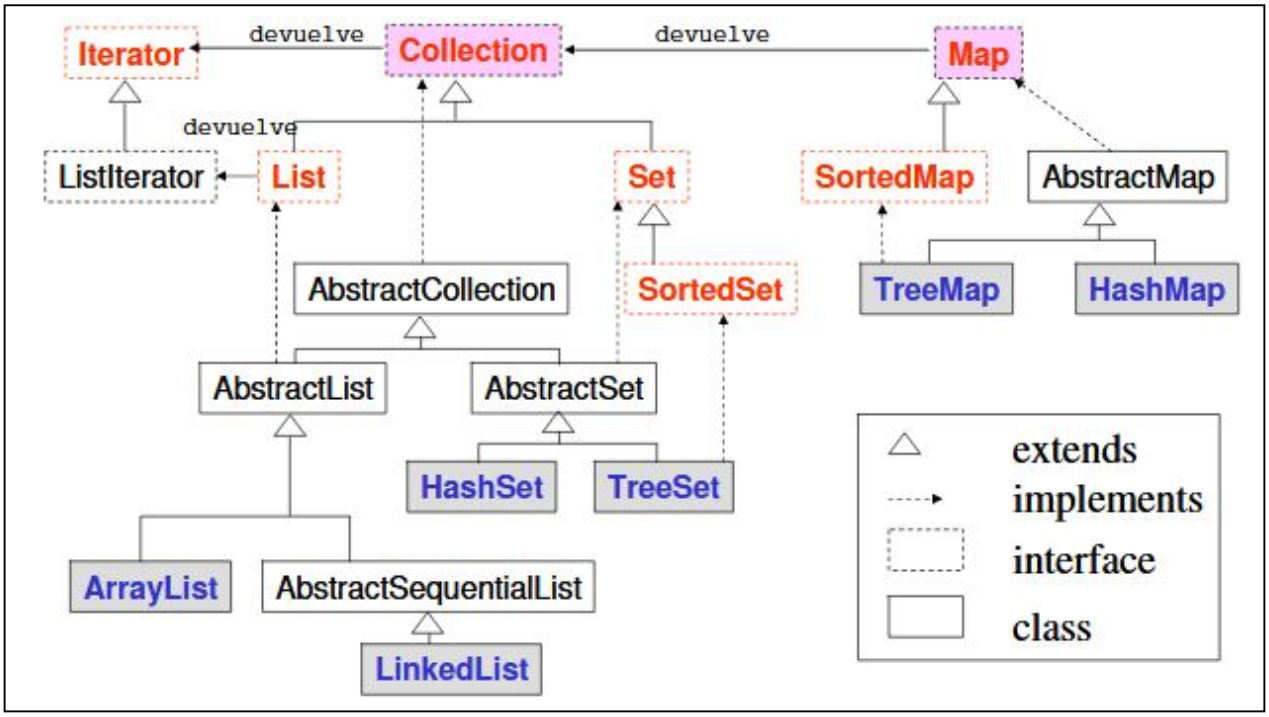
1 Introducción a las colecciones.

Para ver información más detallada consultar:

http://docs.oracle.com/javase/tutorial/collections/TOC.html

Una **colección** es un grupo de objetos almacenados de forma conjunta en una misma estructura. A estos objetos se les llama elementos y para poder trabajar con ellos utilizaremos la **interface** genérica **`Collection`**. Gracias a esta interface, podemos almacenar cualquier tipo de objeto y podemos usar una serie de métodos comunes, como pueden ser: añadir, eliminar, obtener el tamaño de la colección… De la interface genérica **`Collection`** heredan otra serie de interfaces genéricas. Estas subinterfaces imponen más restricciones y aportan distintas funcionalidades sobre la interface anterior.

La jerarquía de interfaces y clases que permiten manipular este tipo de estructuras es la siguiente:



Para **obtener los elementos** almacenados **en una colección** hay que **usar iteradores**, que permiten obtenerlos uno a uno de forma secuencial (*no hay otra forma de acceder a los elementos de una colección*). Los iteradores, que veremos más adelante, se pueden usar de forma transparente, a través de una estructura for especial, denominada bucle "`for-each`" o bucle "para cada". En el siguiente código se usa un bucle `for-each`, en él la variable ***i*** va tomando todos los valores almacenados en el conjunto hasta que llega al último:

```
for (Integer i: conjunto) {
      System.out.println("Elemento almacenado:"+i);
}
```

La estructura for-each es muy sencilla: la palabra for seguida de "**`(tipo variable:colección)`**" y el cuerpo del bucle; ***`tipo`*** es el tipo del objeto sobre el que se ha creado la colección, ***`variable`*** es la variable donde se almacenará cada elemento de la colección y ***`colección`*** es la colección en sí. Los bucles `for-each` se pueden usar para todas las colecciones.

EjemploHashSet

2 Tipos de colecciones

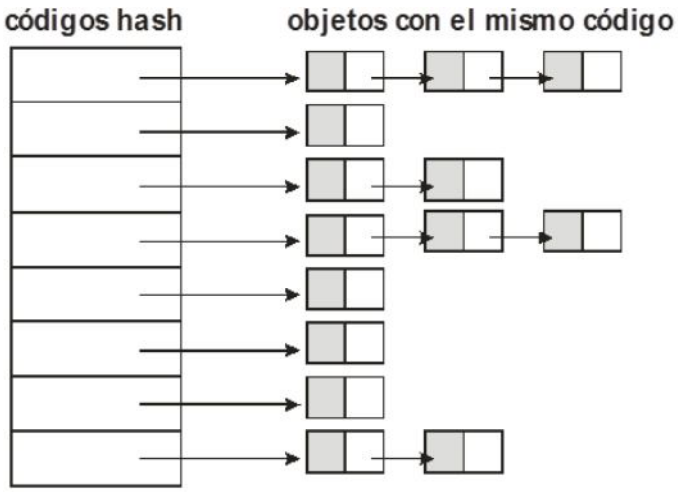
Hay varias interfaces que **heredan** de la interface **`Collection`**, y por ahora vamos a ver las interfaces **`Set`** y **`List`**:

2.1 Interface Set

La **interface `Set`** define una colección que **no puede contener** elementos **duplicados**. Esta interface contiene, únicamente, los métodos heredados de **`Collection`** añadiendo la restricción de que los elementos duplicados están prohibidos. Para comprobar si los elementos están duplicados o no, es necesario que las clases de los objetos que se almacenan en la colección implementen de forma correcta los métodos **`equals`** y **`hashCode`**. Para comprobar si dos **`Set`** son iguales, se comprobará si todos los elementos que los componen son iguales sin importar el orden que ocupen dichos elementos.

La **interface `Set`** es implementada, entre otras por las **clases**:

- **`HashSet (java.util.HashSet)`**: Es la clase más utilizada para implementar una colección sin duplicados. Almacena los elementos en una tabla *hash* (tabla que permite asociar clave con valores). La igualdad de los objetos se comprueba comparando los códigos hash y, si son iguales, se compara con `equals`. Es la implementación con mejor rendimiento de todas pero **no almacena** los objetos **de forma ordenada**.



- De la clase anterior hereda **`LinkedHashSet (java.util.LinkedHashSet)`**, que almacena los objetos combinando tablas hash, para un acceso rápido a los datos, y listas enlazadas para recuperarlos **en el orden en que fueron insertados**.

- **`TreeSet (java.util.TreeSet)`**: esta clase implementa la interface **`SortedSet`**, que almacena **los objetos usando** unas estructuras conocidas como **árboles** rojo-negro, **que los ordenan en función de sus valores**, de forma que aunque se inserten los elementos de forma desordenada, internamente se ordenan dependiendo del valor de cada uno. Es bastante más lento que **`HashSet`**. Las clases de los elementos almacenados deben implementar la interface **`Comparable`** de Java (está en **`java.lang`**). Esta interface define el método **`compareTo`** que utiliza como argumento un objeto a comparar y que devuelve 0 si los objetos son iguales, un número positivo si el primero es mayor que el segundo y negativo en caso contrario.

2.2 Interface List

La **interface `List`** define una sucesión de elementos que están dispuestos en un cierto orden. A diferencia de la interface **`Set`**, la interface **`List`** sí **admite** elementos **duplicados**. Aparte de los métodos heredados de **`Collection`**, añade métodos que permiten mejorar los siguientes puntos:

- Acceso posicional a elementos: manipula elementos en función de su posición en la lista. Las posiciones se empiezan a numerar desde cero.
- Búsqueda de elementos: busca un elemento concreto de la lista y devuelve su posición.
- Rango de operación: permite realizar ciertas operaciones sobre rangos de elementos dentro de la propia lista, por ejemplo insertar los elementos de otra colección a partir de una posición dada.

La **interface `List`** es implementada por dos **clases**:

- **`ArrayList (java.util.ArrayList)`**: esta es la implementación típica. Se basa en un *array* redimensionable que aumenta su tamaño según crece la colección de elementos. La redimensión es transparente al usuario, nosotros, no nos enteramos cuando se produce, pero eso redunda en una diferencia de rendimiento notable dependiendo del uso. Los ArrayList son muy rápidos para acceder a un elemento según su posición. En cambio, eliminar o insertar un elemento implica muchas operaciones ya que hay que mover todos los elementos que van después del nodo que queremos borrar o insertar.
- **`LinkedList (java.util.LinkedList)`**: esta implementación permite que mejore el rendimiento en ciertas ocasiones. Se basa en una lista doblemente enlazada de los elementos, teniendo cada uno de los elementos un puntero al anterior y al siguiente elemento.

  Esta clase también puede implementar las interfaces **`java.util.Queue`** y **`java.util.Deque`**. Dichas interfaces permiten hacer uso de las listas como si fueran una cola de prioridad o una pila, respectivamente.

Si se van a realizar muchas operaciones de eliminación de elementos sobre la lista, conviene usar una lista enlazada (**`LinkedList`**), pero si no se van a realizar muchas eliminaciones, sino que solamente se van a insertar y consultar elementos por posición, conviene usar una lista basada en arrays redimensionados (**`ArrayList`**).

2.3 Interface Map

Los mapas permiten definir colecciones de elementos que poseen pares de datos clave-valor, no pueden contener claves duplicadas y cada una de dichas claves, sólo puede tener asociado un valor como máximo. Esto se utiliza para localizar valores en función de la clave que poseen.

Es la raíz de todas las clases capaces de implementar mapas. Hasta la versión 1.5, los mapas eran colecciones de pares clave, valor donde tanto la clave como el valor eran de tipo **Object**. Desde la versión 1.5 esta interfaz tiene dos genéricos: **K** para el tipo de datos de la clave y **V** para el tipo de los valores (**`Map<k, V>`**).

Esta interfaz no deriva de **`Collection`** y no usa iteradores porque la obtención, búsqueda y borrado de elementos se hace de manera muy distinta, pero veremos que existe un truco interesante. Los mapas **no permiten insertar objetos nulos** (provocan excepciones de tipo **`NullPointerException`**).

Dentro de la interface **`Map`** existen varios tipos de implementaciones realizadas dentro de la plataforma Java. Vamos a analizar cada una de ellas:

- **`HashMap (java.util.HashMap)`**: esta implementación almacena las claves en una tabla *hash*. Es la implementación con mejor rendimiento de todas pero **no garantiza ningún orden** a la hora de realizar iteraciones. Esta implementación proporciona tiempos constantes en las operaciones básicas siempre y cuando la función *hash* disperse de forma correcta los elementos dentro de la tabla *hash*. Es importante definir el tamaño inicial de la tabla ya que este tamaño marcará el rendimiento de esta implementación.
- **`TreeMap (java.util.TreeMap)`**: esta implementación **almacena las claves ordenándolas** en función de sus valores. Es bastante más lento que *HashMap*. La clase de las claves debe implementar la interface **Comparable** o bien durante la creación de un objeto **TreeMap** pasar como parámetro al constructor un objeto **Comparator.**
- **`LinkedHashMap (java.util.LinkedHashMap)`**: esta implementación almacena las claves en función del **orden de inserción**. Es, simplemente, un poco más costosa que **`HashMap`**.

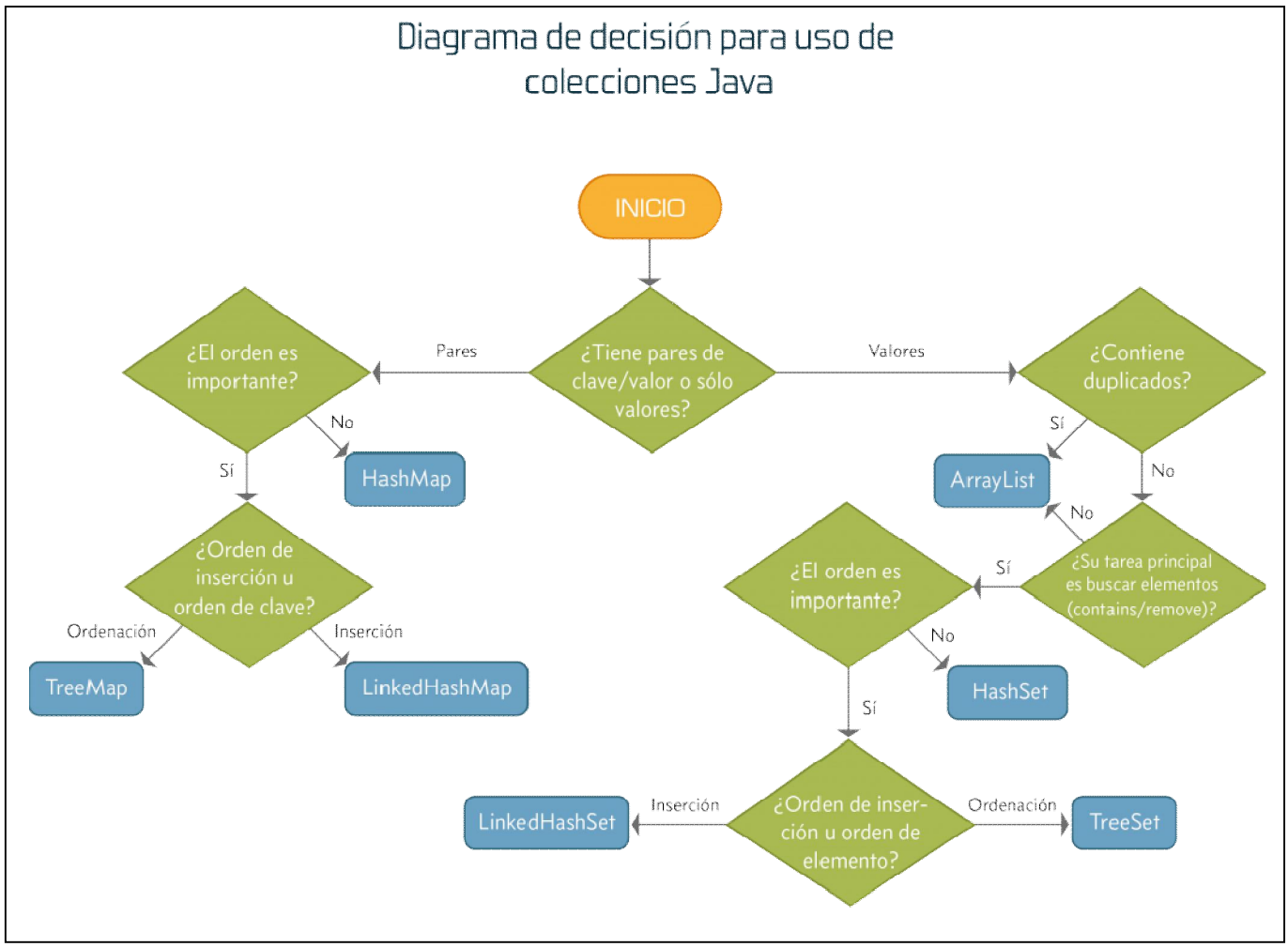
Los **mapas utilizan clases genéricas** y permiten definir un tipo base para la clave, y otro tipo diferente para el valor. Veamos un ejemplo de cómo crear un mapa, que es extensible a los otros dos tipos de mapas:

```
HashMap<String,Integer> t=new HashMap<String,Integer>();
```

El mapa anterior permite usar cadenas como claves y almacenar de forma asociada a cada clave, un número entero.

El cuándo usar una implementación u otra de **Map** variará en función de la situación en la que nos encontremos. Generalmente, **HashMap** será la implementación que usemos en la mayoría de situaciones. **HashMap** es la implementación con mejor rendimiento, pero en algunas ocasiones podemos decidir renunciar a este rendimiento a favor de cierta funcionalidad como la ordenación de sus elementos.

Para conocer qué tipo de colección usar, podemos emplear el siguiente diagrama:



2.4 Interface Iterator (java.util.Iterator)

La interface `Collection` tiene el método **`iterator()`** que devuelve un iterador, que es un objeto que implementa, o bien la interface `Iterator`, o bien la interface `ListIterator`, y con los métodos de este objeto podemos recorrer la colección.

Las colecciones se pueden recorrer de dos formas: con **bucles for-each** (existentes en Java a partir de la versión 1.5) **y a través de un bucle normal creando un iterador**. Los bucles for-each ya los hemos visto. Nos falta ver cómo se crea un iterador: sólo hay que invocar al método "`iterator()`" de cualquier colección.

```
Iterator<Integer> it = t.iterator();
```

Hemos especificado un parámetro para el tipo de dato genérico en el iterador (poniendo "`<Integer>`" después de `Iterator`) porque es una interface genérica y es necesario especificar el tipo base que contendrá el iterador. Si no se especifica el tipo base del iterador, igual se puede recorrer la colección, pero devolverá objetos tipo `Object` (clase de la que derivan todas las clases), con lo que nos veremos obligados a forzar la conversión de tipo.

Para recorrer y gestionar la colección, el iterador ofrece tres métodos básicos:

- `boolean hasNext()`. Devuelve *true* si le quedan más elementos por visitar en la colección y *False* en caso contrario.
- `E next()`. Devuelve el siguiente elemento de la colección, si no existe siguiente elemento lanzará la excepción `NoSuchElementException`, por eso conviene chequear primero si el siguiente elemento existe.

- `remove()`. Elimina de la colección el último elemento devuelto en la última invocación de *next* (no es necesario pasárselo como parámetro). Si *next* no ha sido invocado todavía saltará una excepción.

EjemploIterator01

2.5 Interface ListIterator

Hereda de la interface `Iterator`, por tanto puede usar sus métodos y además añade otros nuevos para modificar elementos o recorrer la lista en cualquier sentido.

3 Recorrido de mapas

Un iterador está pensado para no sobrepasar los límites de la colección, ocultando operaciones más complicadas que pueden producir errores. Puede haber problemas cuando es necesario hacer la operación de conversión de tipos. Si la colección no contiene los objetos esperados, al intentar hacer la conversión, saltará una excepción. Usar genéricos aporta grandes ventajas, pero hay que usarlos adecuadamente.

Para recorrer los mapas con iteradores, hay que hacer un pequeño truco. Usamos el método **`entrySet`** que ofrecen los mapas para generar un conjunto con las entradas (pares de clave-valor), o bien, el método **`keySet`** para generar un conjunto con las claves existentes en el mapa. Veamos cómo sería para el segundo caso, con **`keySet`**, que es el más sencillo:

```
Map<Integer,Integer> mapa=new HashMap<Integer,Integer>();

for (int i=1;i<=10;i++) mapa.put(i, i*2);//Insertamos datos de prueba en el mapa.

for (Integer clave:mapa.keySet()) // Recorremos el conjunto generado por keySet,
                                  // contendrá las claves.
{
   Integer valor=mapa.get(clave); //Para cada clave, accedemos a su valor si
                                  // es necesario.
   System.out.println(valor);
}
```

Lo único que hay que tener en cuenta es que el conjunto generado por `keySet` no tendrá el método `add` para añadir elementos al mismo, dado que eso habrá que hacerlo a través del mapa.

Si usas iteradores, y piensas eliminar elementos de la colección (e incluso de un mapa), debes usar el método remove del iterador y no el de la colección. Si eliminas los elementos utilizando el método remove de la colección mientras estás dentro de un bucle de iteración los fallos que pueden producirse en tu programa son impredecibles.

Los problemas son debidos a que el método remove del iterador elimina el elemento de dos sitios: de la colección y del iterador en sí (que mantiene interiormente información del orden de los elementos). Si usamos el método remove de la colección la información solo se elimina de un lugar, de la colección.