Árboles balanceados (AVL)

Estructura de datos

Definición de un árbol balanceado

- Es un árbol binario de búsqueda en el cual se cumple: “Para todo nodo T del árbol, la altura de los subárboles izquierdo y derecho no debe diferir en más de una unidad”.
- Son llamados árboles AVL en honor a sus inventores G.M. Adelson-Velskii y E.M. Landis.
- La idea central de éstos es la de realizar reacomodos o balanceos, después de inserciones o eliminaciones de elementos.
- Los árboles balanceados se parecen mucho, en su mecanismo de formación, a los números de Fibonacci. El árbolde altura 0n es vacío, el árbol de altura 1 tiene un único nodo y en general el número de nodos del árbol con altura h>1 se calcula aplicando la siguiente fórmula recursiva:

$$K_h=K_{h-1}+1+K_{h-2}$$

donde k=número mínimo de nodos de árbol y h=altura

Ejemplo número de nodos mínimo

Para calcular el número de nodos de un árbol AVL con altura 5:

$K_5=K_4+1+K_3$

$K_4=K_3+1+K_2$

$K_3=K_2+1+K_1$

$K_2=K_1+1+K_0$

$K_1=1$

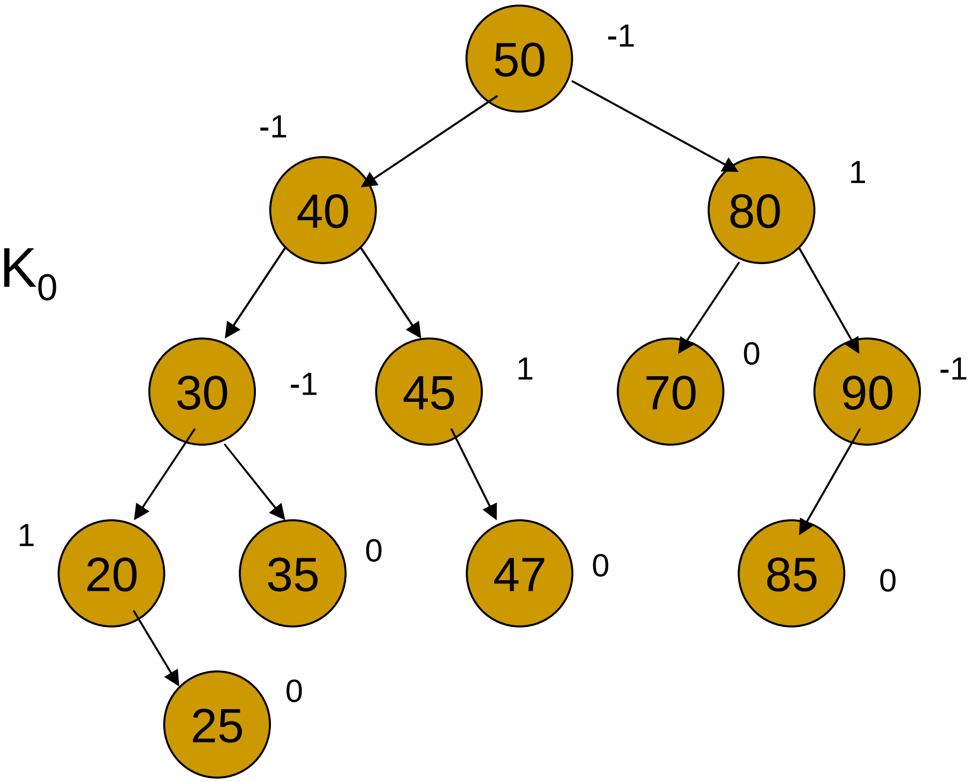
$K_0=0$

$K_2=2$

$K_3=4$

$K_4=7$

$K_5=12$



Algunos estudios demuestran que la altura de un árbol balanceado de n nodos nunca excederá de $1.44*\log_2 n$

Operaciones en un árbol balanceado

- Para poder determinar si un árbol está balanceado o no, debe manejarse información relativa al equilibrio de cada nodo del árbol. Surge así el concepto de factor de equilibrio del un nodo (FE) que se define como: la altura del subárbol derecho menos la altura del subárbol izquierdo.

$$FE=H_{RD}-H_{RI}$$

- Se debe considerar que cada operación de inserción y eliminación puede alterar el equilibrio del árbol.
- Por lo cual es necesario revisar por cada operación el factor de equilibrio de cada nodo involucrado en el camino de búsqueda y si el equilibrio es afectado, será necesario realizar una reestructuración del árbol.
- El proceso termina al llegar a la raíz del árbol, o cuando se realiza la reestructuración del mismo; en cuyo caso no es necesario determinar el FE de los restantes nodos.

Inserción en árboles balanceados

1. Las ramas izquierdas (RI) y derecha (RD) del árbol tienen la misma altura ($H_{RI}=H_{RD}$), por lo tanto:
   a) Si se inserta un elemento en RI entonces H_{RI} será mayor a H_{RD}
   b) Si se inserta un elemento en RD entonces H_{RD} será mayor a H_{RI}

   No se altera el equilibrio del árbol
2. Las ramas izquierda (RI) y derecha (RD) del árbol tienen altura diferente ($H_{RI}!=H_{RD}$):
   a) Supóngase que $H_{RI}<H_{RD}$:
      - Si se inserta un elemento en RI entonces H_{RI} será igual H_{RD}. Se mejora el equilibrio del árbol.
      - Si se inserta un elemento en RD entonces se rompe el criterio de equilibrio del árbol y es necesario reestructurarlo.
   b) Supóngase que $H_{RI}>H_{RD}$:
      - Si se inserta un elemento en RI, entonces se rompe el criterio de equilibrio del árbol y es necesario reestructurarlo.
      - Si se inserta un elemento en RD entonces H_{RD} será igual H_{RI}. Se mejora el equilibrio del árbol.

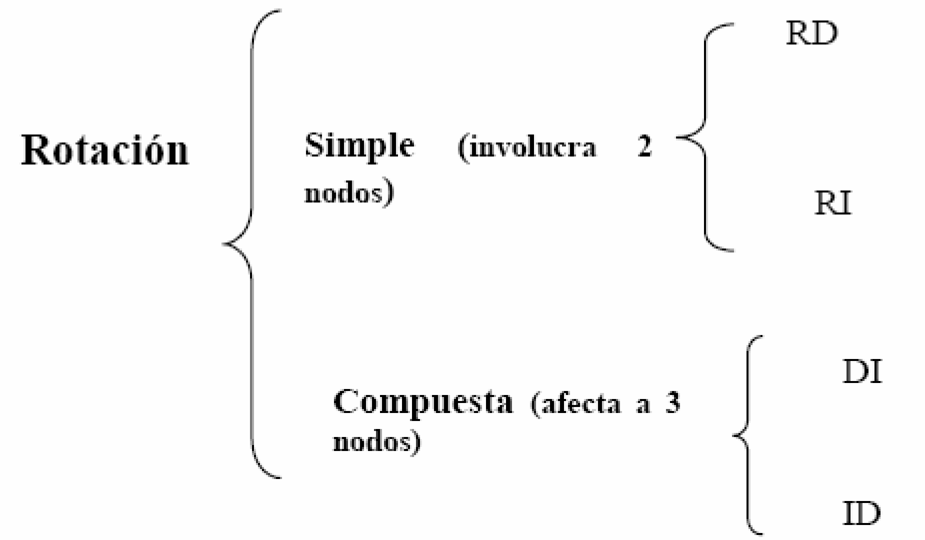
Borrado en árboles balanceados

1. Localizar la posición del nodo a eliminar en el árbol.
2. Se elimina siguiendo los siguientes criterios:
   a) Si el elemento a borrar es terminal u hoja, simplemente se suprime.
   b) Si el elemento a borrar tiene un solo descendiente entonces, tiene que sustituirse por ese descendiente.
   c) Si el elemento a borrar tiene los dos descendientes, entonces tiene que sustituirse por el nodo que se encuentra más a la izquierda en el subárbol derecho o por el nodo que se encuentra más a la derecha en el subárbol izquierdo.
3. Se regresa por el camino de búsqueda calculando el FE de los nodos visitados. Si en alguno de los nodos se viola el criterio de equilibrio, entonces debe reestructurarse el árbol. El proceso termina cuando se llega hasta la raíz del árbol.

Nota: Cabe aclarar que mientras que en el algoritmo de inserción una vez que era efectuada una rotación podía detenerse el proceso, en este algoritmo debe continuarse puesto que se puede producir más de una rotación en el camino hacia atrás.

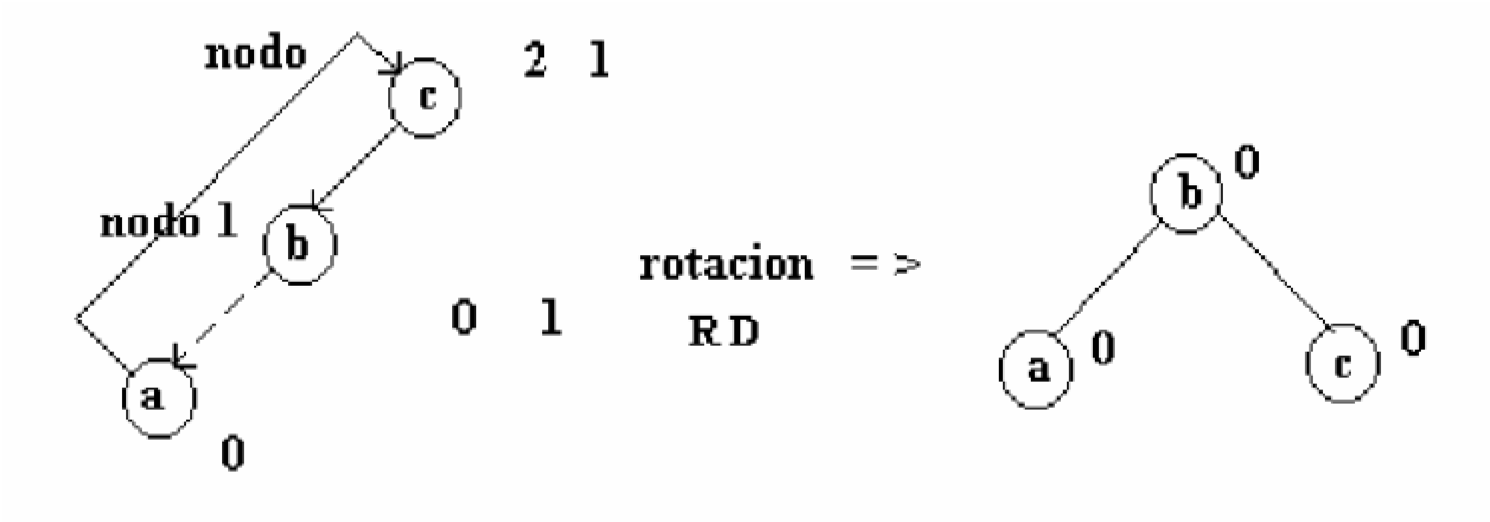
Reestructuración de un árbol balanceado

- Reestructurar el árbol significa rotar los nodos del mismo. La rotación puede ser simple o compuesta:
- La rotación simple involucra dos nodos: puede realizarse por las ramas derechas (RD) o por las ramas izquierdas (RI).
- La rotación compuesta involucra tres nodos: puede realizarse por las ramas derecha e izquierda (DI) o por las ramas izquierda y derecha (ID).



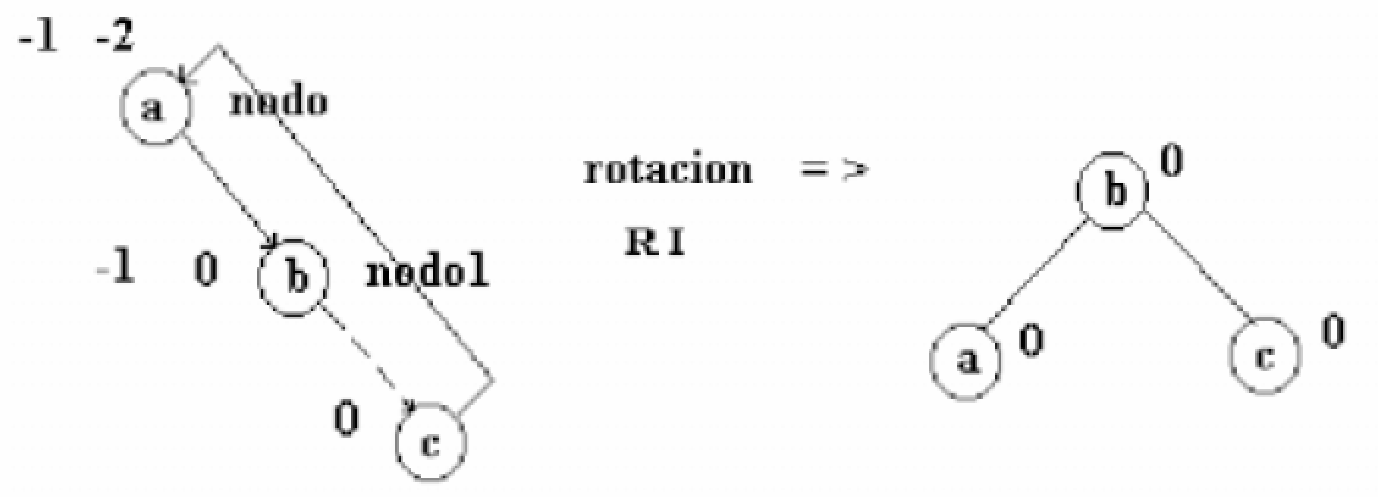
Rotación simple a la derecha (RD)

- El subárbol izquierdo de un nodo sea 2 unidades más alto que el derecho, es decir, cuando su FE sea de 2. Y además, la raíz del subárbol izquierdo tenga una FE de 1, es decir, que esté cargado a la izquierda.



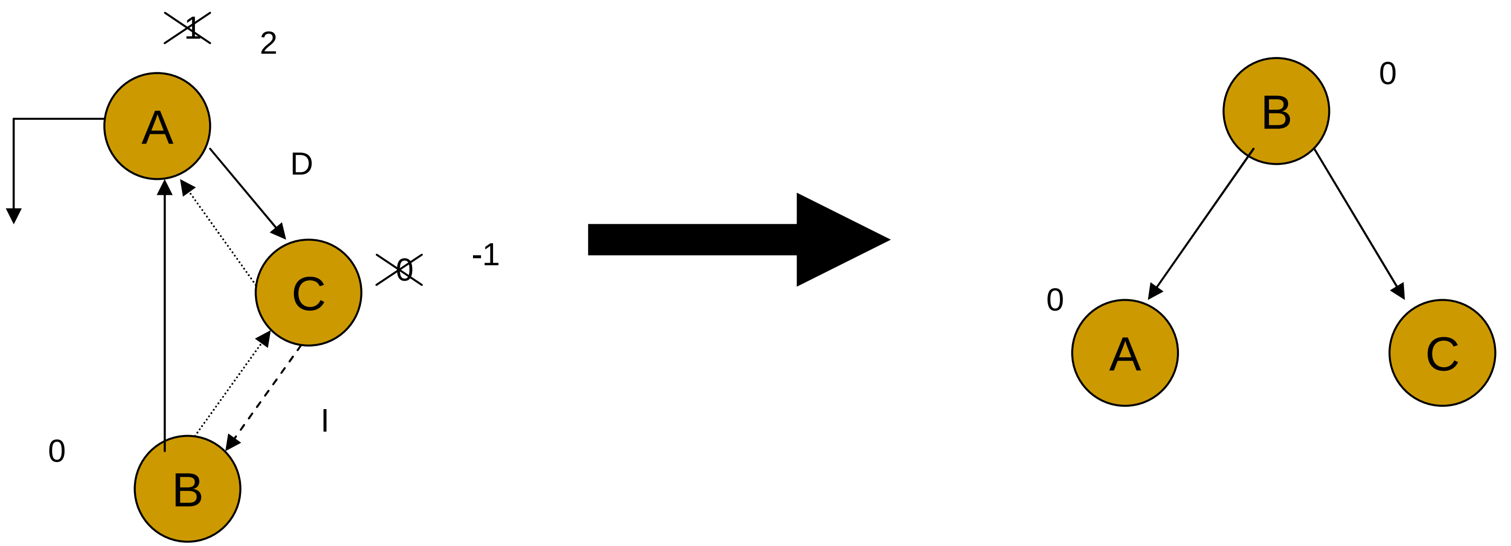
Rotación simple a la izquierda (RI)

- El subárbol derecho de un nodo sea 2 unidades más alto que el izquierdo, es decir, cuando su FE sea de -2. Y además, la raíz del subárbol izquierdo tenga una FE de -1, es decir, que esté cargado a la derecha.



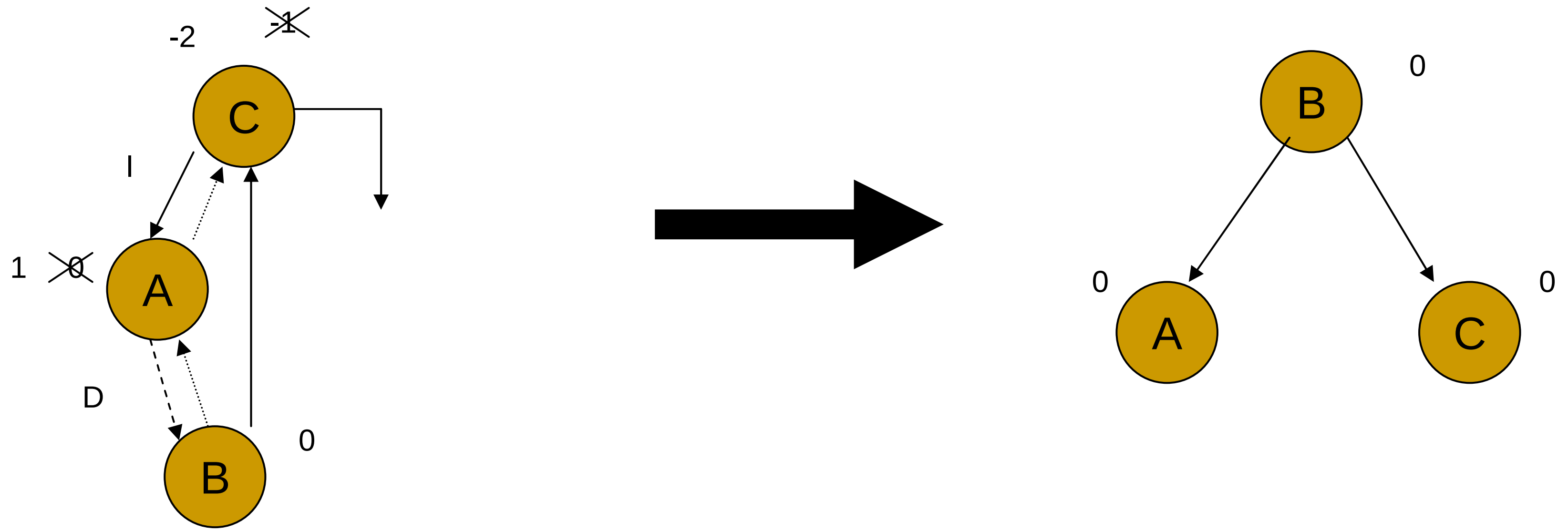
Rotación compuesta derecha Izquierda (DI)

- Cuando el subárbol izquierdo de un nodo sea 2 unidades más alto que el derecho, es decir, cuando su FE sea -2. Y además, la raíz del subárbol izquierdo tenga una FE de 1, es decir, que esté cargado a la derecha.



Rotación compuesta Izquierda Derecha (ID)

- Cuando el subárbol derecho de un nodo sea 2 unidades más alto que el izquierdo, es decir, cuando su FE sea 2. Y además, la raíz del subárbol izquierdo tenga una FE de -1, es decir, que esté cargado a la izquierda.



Análisis de eficiencia

- El análisis matemático de los algoritmos de inserción y eliminación demuestran que es posible buscar, insertar y eliminar un elemento en un árbol balanceado de n nodos en O(log n) unidades de tiempo.
- Diversos análisis demuestran que son más frecuentes las rotaciones en las operaciones de inserción que en las de eliminación.
- Mientras se produce aproximadamente una rotación por cada dos inserciones, se produce una rotación por cada cinco eliminaciones.