

Estructuras de datos

Diferentes tipos de árboles binarios

Investigación

TID42M

Walter González

Aldo Barrera

Docente Milton Batres



20 nov 2024

Contenido

Objetivo:.....	3
Material:	3
Desarrollo:.....	4
Arbol Binario	4
Concepto general	4
Altura	5
Profundidad	5
Definición concreta.....	5
Arbol Balanceado.....	6
Concepto general	6
Definición concreta.....	7
Arbol R	8
Concepto general	8
Definición concreta.....	9
Árbol AVL.....	10
Conclusiones:	12
Referencias.....	13

Objetivo:

El presente documento tiene como objetivo analizar y documentar las características fundamentales de cuatro tipos de estructuras de datos arbóreas: árbol binario, árbol balanceado, árbol R y árbol AVL. Se presentará la definición y concepto de cada estructura, estableciendo sus propiedades básicas y aplicaciones prácticas en el contexto de la programación y manejo de datos.

Material:

Internet

Word

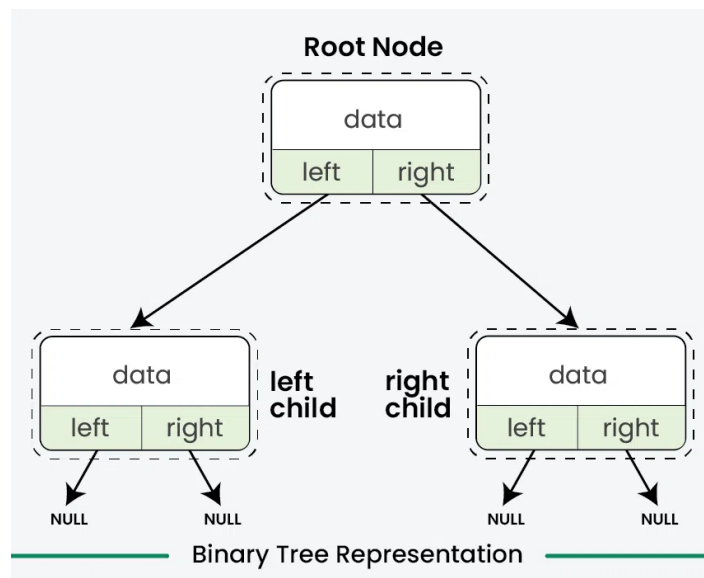
Desarrollo:

Arbol Binario

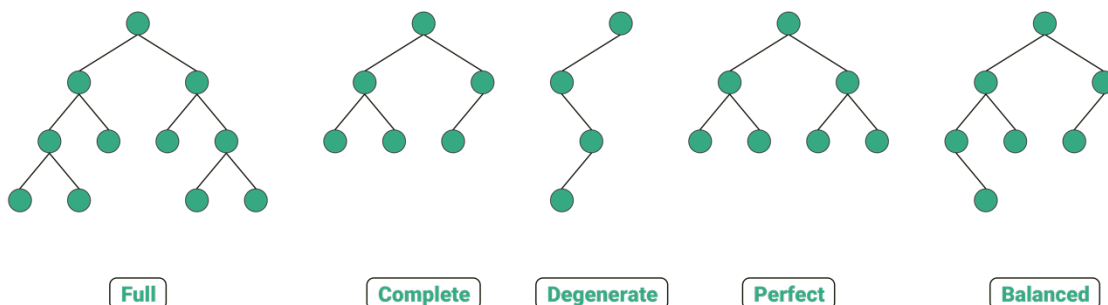
Concepto general

Esta estructura de datos no lineal puede tener como máximo dos nodos hijo, permite el uso de diferentes algoritmos para la inserción, borrado y operaciones transversales. Los atributos clave de un arbol binario son los siguientes:

Los **nodos** guardan un elemento, un puntero hacia un hijo izquierdo y uno hacia hijo derecho.



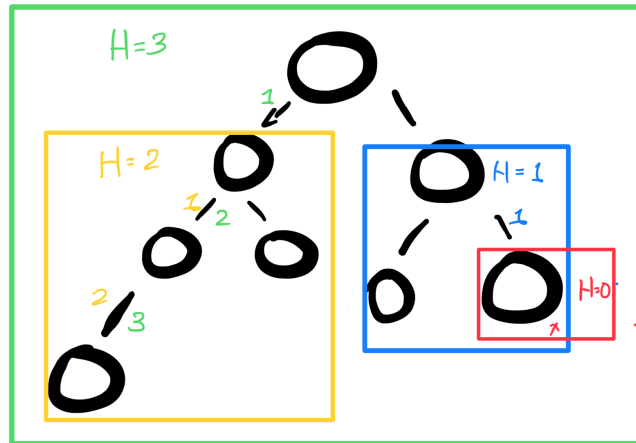
Se tiene un **nodo raíz**, que es el nodo más alto de la estructura. Los nodos sin hijos son llamados **hoja** y nodos con uno o más hijos son **nodos internos**.



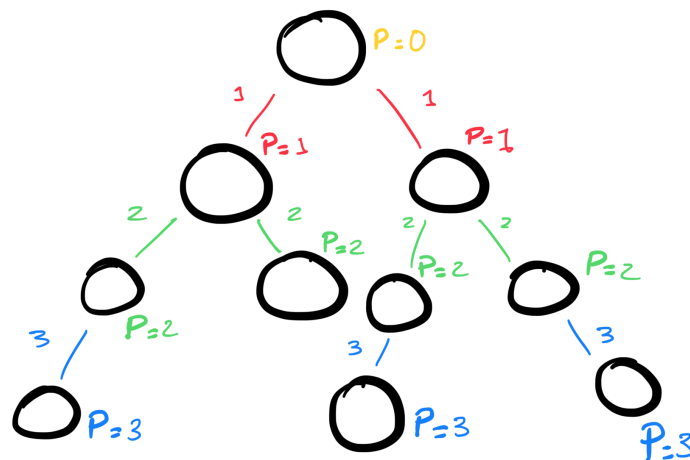
REPORTE DE PRÁCTICA

La **altura** y la **profundidad** son metricas para medir un árbol binario. La altura se define como el número de aristas en el camino más largo desde la raíz hasta una hoja. La profundidad representa el número de aristas que hay entre la raíz y un nodo específico x.

Altura



Profundidad



Definición concreta

Un árbol **binario** se define como una estructura de datos jerárquica en la que cada nodo puede tener un máximo de dos hijos, conocidos como hijo izquierdo y hijo derecho. (GeeksforGeeks, 2024)

Arbol Balanceado

Concepto general

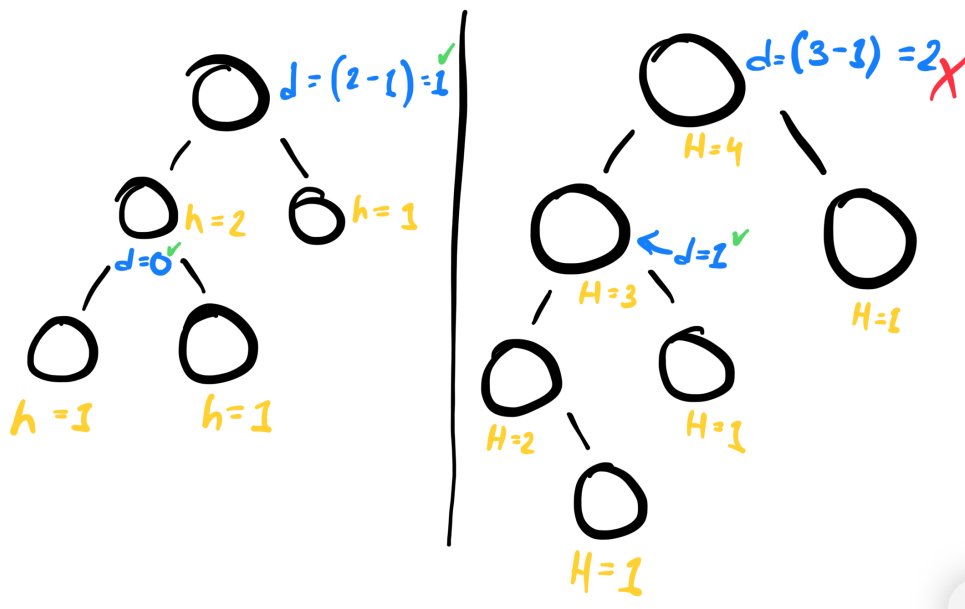
Tipo de árbol binario, diseñados para mantener una estructura específica que garantiza operaciones eficientes.

La característica principal de los árboles balanceados es que la diferencia de altura de los subárboles izquierdo y derecho de cualquier nodo esté entre -1 y 1.

Además, la altura es en función logarítmica al número de nodos, en $O(\log n)$, permitiendo que las operaciones de búsqueda, inserción y borrado sean sumamente efectivas, ya que la altura del árbol no se incrementa abruptamente mientras que el número de nodos crece.

(GeeksforGeeks, 2023)

En estos ejemplos se muestra un árbol balanceado (izquierda) contra uno no balanceado (derecha). Esto sucede a que la diferencia de alturas de los subárboles nunca es mayor a 1 en el primer ejemplo, sin embargo, al calcular la diferencia de alturas de los nodos hijos del nodo raíz del subárbol derecho obtenemos que la altura 3 (hijo izquierdo) menos la altura 1 (hijo derecho) es igual a 2, rompiendo el axioma.



Este tipo de árboles deben implementar mecanismos para balancearse automáticamente después de inserciones o eliminaciones.

Definición concreta

Un árbol binario en el que la altura de los subárboles izquierdo y derecho de cualquier nodo difiere en no más de uno. Matemáticamente, esto puede expresarse como: $\text{Altura subárbol izquierdo} - \text{Altura subárbol derecho} \leq 1$.

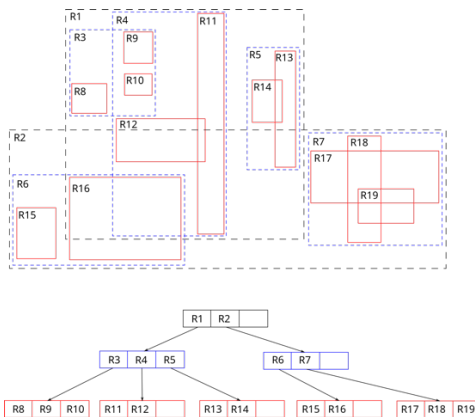
Además de esta condición, tanto el subárbol izquierdo como el derecho deben ser por sí mismos árboles binarios balanceados. Esta definición asegura que las operaciones en el árbol permanezcan eficientes al mantener su altura logarítmica en relación con el número de nodos que contiene.

(Balanced Binary Tree, n.d.)

Arbol R

Concepto general

El árbol R está diseñado para indexar datos espaciales multidimensionales como coordenadas geográficas, rectángulos o polígonos. La R hace referencia a Rectángulo. Una de sus aplicaciones más comunes es el almacenamiento de objetos espaciales para implementar mapas virtuales o mostrar puntos de interés, permitiendo consultas rápidas como "Encontrar todas las gasolineras en un radio de dos kilómetros".



Su estructura se basa en un algoritmo de balanceo por altura, similar al árbol B, pero sin limitarse a dos nodos hijo. Los nodos hoja contienen índices con punteros a objetos de datos.

Antonin Guttman, su creador, diseñó esta estructura para aplicaciones que manejan colecciones de tuplas representando objetos espaciales. Cada tupla tiene un identificador único para su recuperación. Los nodos hoja del árbol R almacenan registros de índice con el formato:

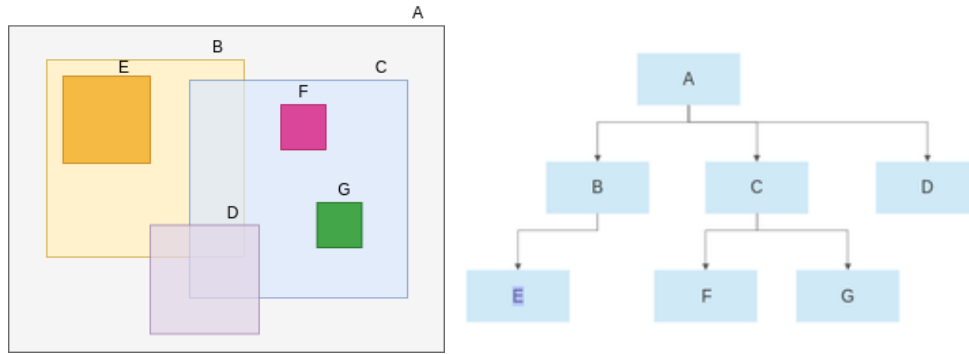
(I, identificador_de_tupla)

El identificador_de_tupla referencia una tupla en la base de datos, mientras que I representa un rectángulo n-dimensional que funciona como caja delimitadora del objeto espacial indexado. Los nodos internos, por su parte, almacenan registros con el formato:

(I, apuntador_a_hijo)

En el siguiente diagrama, se muestran contenedores de puntos espaciales en un plano, tomando al contenedor A como el super contenedor o el nodo raíz.

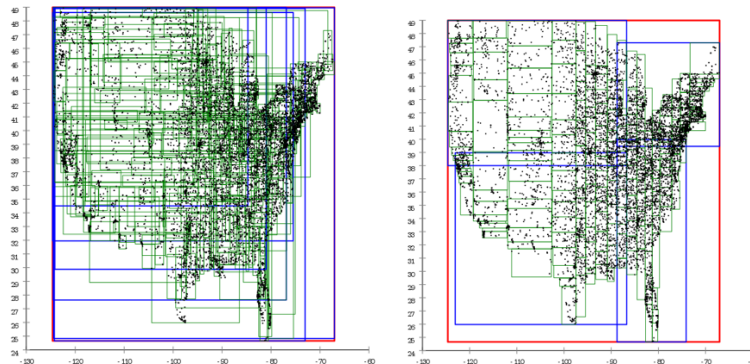
REPORTE DE PRÁCTICA



Este nodo raíz A contiene un puntero hacia la región mas grande del espacio.

Los nodos internos contienen apuntes hacia nodos hijos donde la region de los nodos hijos se superpone por completo dentro de la región de nodos padre.

Esta estructura tambien propone el concepto de MBR (Región de interés minima) refiriendose al parametro contenedor más pequeño de algun objeto de interés en concreto.



Aquí se muestra una aplicación al indexar la base de datos de los distritos postales de USA con un árbol R.

Definición concreta

Un árbol-R es una estructura de datos en árbol diseñada para indexar eficientemente datos espaciales multidimensionales, como coordenadas geográficas o rectángulos. Introducido por Antonin Guttman en 1984, el árbol-R organiza los datos en una jerarquía de nodos, donde cada nodo contiene Rectángulos Delimitadores Mínimos (MBR) que engloban sus nodos hijos u objetos de datos.

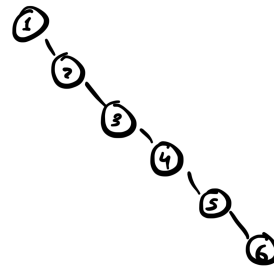
(The R-Tree: A Dynamic Index Structure for Spatial Searching, n.d.)

Árbol AVL

Concepto general

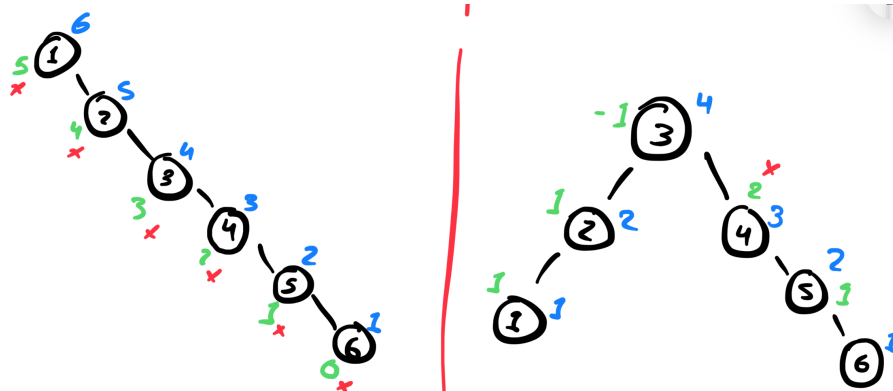
Aunque un árbol binario busca conseguir una complejidad de $O(\log N)$ existen escenarios donde la estructura se aleja mucho del objetivo. En el siguiente ejemplo se plantea una inserción de 6 números consecuentes, que al ser ingresados se apendizan como hijo derecho de la hoja.

[1, 2, 3, 4, 5, 6]

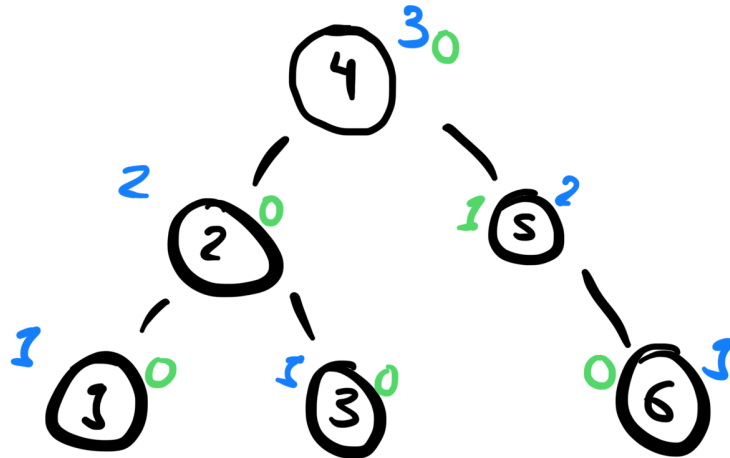


Los árboles AVL resuelven este problema. Son árboles binarios que se balancean automáticamente después de cada inserción o eliminación, según la altura de sus nodos. Esta estructura garantiza complejidades y tiempos específicos de ejecución.

También, este cumple con la condición de que, para cada nodo, la altura de sus subárboles difiere por máximo 1, a esta diferencia se le conoce como **factor de balanceo**.



Ambos árboles tienen uno o mas nodos con un factor de balanceo fuera de menor a -1 o mayor a 1



Este árbol se encuentra perfectamente balanceado y conserva los mismos datos del arreglo original. Su importancia radica en la implementación: cómo logra mantener su balance después de realizar operaciones, lo que le permite seguir siendo eficiente y óptimo.

Definición concreta

Un árbol AVL (nombrado por sus inventores Adelson-Velsky y Landis) es un árbol binario de búsqueda auto balanceado donde la diferencia de altura entre los subárboles izquierdo y derecho de cualquier nodo no puede ser mayor que uno. Esta estructura garantiza operaciones de búsqueda, inserción y eliminación en tiempo logarítmico.

Conclusiones:

A lo largo de la investigación de este reporte, se encontraron conceptos que verbalizaban muchos de los temas vistos en clase y han ayudado a la comprensión espacial del tema, y eso particular llama mucho nuestra atención, que son temas que requieren de una comprensión espacial más que una lógica matemática, donde a través del entendimiento de patrones podemos discernir a simple vista la naturaleza de un árbol en concreto.

El más interesante que se encontró fue el árbol de tipo AVL, debido a su implementación más robusta y compleja, lo interesante es cuando se analiza material de apoyo, se ven los cambios constantes en la estructura y parece ser una estructura orgánica y consciente, despertando un asombro donde las matemáticas y el código se unen. Sin duda es lo más interesante de esta práctica.

El árbol R fue el más complejo de investigar debido a la escasa bibliografía y material de apoyo disponible. Este árbol llama mucho nuestra atención gracias a la forma en la que las ciencias de la computación pueden resolver problemas del mundo real después de complejas abstracciones y mucho código.

Después de investigar, resulta interesante considerar que las estructuras no lineales y jerárquicas se denominan "árboles" no solo por su similitud gráfica, sino por su comportamiento cuasi-orgánico y axiomático.

Referencias

Balanced Binary tree. (n.d.). <https://www.programiz.com/dsa/balanced-binary-tree>

*GeeksforGeeks. (2024, November 19). Binary Tree Data structure.
GeeksforGeeks. <https://www.geeksforgeeks.org/binary-tree-data-structure/>*

*GeeksforGeeks. (2024, August 8). AVL Tree Data Structure. GeeksforGeeks.
<https://www.geeksforgeeks.org/introduction-to-avl-tree/>*

GeeksforGeeks. (2023, April 26). Balanced Binary Tree definition & meaning in DSA. GeeksforGeeks. <https://www.geeksforgeeks.org/balanced-binary-tree-definition-meaning-in-dsa/>

*Guttman, A. (1984). R-trees: A dynamic index structure for spatial searching. In *Proceedings of the 1984 ACM SIGMOD international conference on Management of data (SIGMOD '84)* (pp. 47-57). Association for Computing Machinery. <https://doi.org/10.1145/602259.602266>*

*The R-Tree: A dynamic index structure for spatial searching. (n.d.).
<https://hpi.de/rabl/teaching/winter-term-2019-20/foundations-of-database-systems/the-r-tree-a-dynamic-index-structure-for-spatial-searching.html>*