## MODULE 4

## TREES \& GRAPHS

## TREE

- A tree is a nonlinear data structure. Data elements are related in a hierarchal manner (parent-child relationship)
- The first node of the tree is called Root node.
- Final nodes are called Leaf nodes.
- Leaf nodes are also called terminal nodes.
- A child has a single parent but a parent has more than one children. So we can say that, here, a one to many relationship is exist.
- In a general tree, A node can have any number of children nodes but it can have only a single parent.
- The following image shows a tree, where the node $A$ is the root node of the tree while the other nodes can be seen as the children of $A$.



## BASIC TERMINOLOGY

Root node - topmost node in the tree hierarchy
Sub tree - If the root node is not null, the tree T1, T2 and T3 is called sub-trees of the root node.
$>$ Degree - The number of sub trees of a node. Eg: The degree of $A$ is 3 Siblings - Children of the same parent. Eg: G, H are the siblings of $D$ $>$ Level - If a node is at level I, then its children are at level I+1
Depth - The height or depth of a tree is defined to be the maximum level of any node in the tree
Forest - If we remove the root of a tree, we get a forest. Eg: If we remove A, we get a forest with 3 trees.


## BINARY TREE

- A binary tree is either empty or consist of a root and 2 disjoint binary trees called the Left Sub tree and Right Sub tree.
- Each node can have at most two children
- B is the left child of $A$
- $C$ is the right child of $A$



## TYPES OF BINARY TREE

Strictly Binary Tree - All nodes must have a left child and a right child except leaf node

- Here B, F , G \& E are leaf nodes

Skewed Binary Tree - It is a special kind of binary tree. Two type skewed binary trees are Left skewed and Right skewed



Fig. Full Binary Tree

- Full Binary Tree - If each node of binary tree has either two children or no child at all, is said to be a Full Binary Tree. Full binary tree is also called as Strictly Binary Tree.


## Complete Binary Tree

- If all levels of tree are completely filled except the last level and the last level has all keys as left as possible, is said to be a Complete Binary Tree.
- Complete binary tree is also called as Perfect Binary Tree.
- In a complete binary tree, every internal node has exactly two children and all leaf nodes are at same level.
- For example, at Level 2, there must be $22=4$ nodes and at Level 3 there must be $23=8$ nodes.


Fig. Complete Binary Tree

- The maximum number of nodes on level " $i$ " of a binary tree is $2^{i-1}$
- The maximum number of nodes in a binary tree of depth $k$ is $2^{k}-1$


## TREE REPRESENTATION

Tree can be represented using two ways

1. Using Array
2. Using Linked List

If a complete binary tree with $n$ node, then depth $=\left\lfloor\log _{2} n\right\rfloor+1$, in an array representation,

- Parent (i) is at [i/2〕
- Lchild (i) is at 2 i
- Rchild (i) is at $2 \mathrm{i}+1$



## Eg:2



Size of array $=2^{d}=2^{3}=8$ In array representation, we do not use location " 0 ".

## Consider the following tree



Skewed representation of a tree can be implemented in an array. But this representation is inefficient. So we use linked list representation of the tree.

In linked list representation, a node is represented by

| L child | Data | R child |
| :--- | :--- | :--- |

Eg:


## BINARY TREE TRAVERSALS

> Traversing means, visiting the node at least one time. Three type of traversals are:

- In order (L D R)
- Pre order (D L R)
- Post order (L R D)


## *In order Traversal

- It follows "LDR" form. ( Left - Data - Right)


DBEAFC G




## Preorder using Recursion



Prepared By Mr.EBIN PM, AP, IESCE

## Postorder using Recursion

Algorithm postorder (current node)
\{
// current node is a pointer to node in binary tree if (current node $\neq$ nil) \{ postorder (current node . lchild); postorder (current node . rchild); write (current node . data);

## Ans:

ABC **/DeE +


Prepared By Mr.EBIN PM, AP, IESCE


## ALGORITHM TO CREATE A COPY OF A BINARY TREE

Algorithm copy (original tree) \{
// for a binary tree "original tree", copy Returns a pointer to an exact copy of the Original tree

```
if (original tree \not= nil)
    {
        new (temp tree);
```

        temp tree . lchild = copy(original tree. lchild);
        temp tree. rchild \(=\operatorname{copy}\) (original tree. rchild);
        temp tree . data \(=\) original tree . data;
        copy \(=\) temp tree;
    \}
    else
copy = nil;
\}


## THREADED BINARY TREE

- In normal case, binary tree is represented using linked list. In that case, if ' $n$ ' nodes present, then total number of links is $2 \times n$. Here $\mathrm{n}+1$ link is a null link.

For example,


Number of node ' n ' $=5$
Total no. of links $=2 \times n=2 \times 5=10$
Null link $=n+1=5+1=6$
This is a disadvantage. So for avoiding this we use threaded binary trees

The idea is to replace the null link by pointer called thread, to other nodes in the tree. Here,
If the Lchild of a node is empty, then we pointed this on an inorder predecessor. If Rchild is empty, then we pointed this on an inorder successor.



Inorder = 42513

- The disadvantage of threaded binary tree is the dangling pointer.
- The left pointer of leftmost child is a dangling pointer and the right pointer of rightmost child is a dangling pointer.
- In memory representation we must distinguish the normal pointer and thread pointer. So we create a new node structure.

| L thread | L child | Data | $R$ child | $R$ thread |
| :--- | :--- | :--- | :--- | :--- |

- Lthread and Rthread are Boolean variables. It can receive either True or False value
- If Lthread $=T$, it indicate that the Lchild is a thread pointer
- If Lthread = F, it indicate that Lchild pointer is a normal pointer. That is it points to the left child.
- For avoiding the dangling reference pointer, we include a head node to the tree.



## BINARY TREE REPRESENTATION OF TREES

- Every tree can be represented as a binary tree.
- To obtain a binary tree representation, we need a relationship between the nodes that can be characterized by at most two quantities.
- One such relationship is Left most child - Next - Right sibling
- Every node has at most one left most child and at most one next right sibling.
- In the following figure, the left most child of $B$ is $E$ and the next right sibling of $B$ is $C$. The node structure is given below


Node Structure




We tilt the figure roughly 45 degree clockwise and this is the binary tree representation.

## CREATION OF A BINARY TREE

## Preorder = ABCDEFGHI

Inorder = BCAEDGHFI
In preorder, the first one is the root node.ie, A.
Scanning is done from left to right, so we complete the left side first.
Preorder : A BCDEFGH I
Inorder: B C A E D G H F I



## CREATION OF A BINARY TREE

## Post order $=\mathrm{ABC} * / D-E F G+*+$

Inorder $=A / B * C-D+E * F+G$
In post order, the last one is the root node.ie, +
Scanning is done from right to left, so we complete the right side first.

Preorder: A B C ${ }^{*} / \mathrm{D}-\stackrel{\text { EFGG }}{\stackrel{\text { Scan }}{*}}+$
Inorder: A / B * C $-\mathrm{D}+\mathrm{E} * \mathrm{~F}+\mathrm{G}$



## BINARY SEARCH TREE (BST)

A Binary Search Tree is a binary tree ; either it is empty or each node in the tree contains an identifier and
a) All identifiers in the left sub-tree of T are less than the identifier in the root node T .
b) All identifiers in the right sub-tree of T are greater then the identifier in the root node $T$
c) The left and right sub-tree of $T$ are also binary search trees.

( Fig: Binary Search Tree )

## SEARCHING IN BINARY SEARCH TREE <br> Algorithm search ( $\mathrm{T}, \mathrm{x}, \mathrm{i}$ ) <br> \{ <br> $\mathrm{i}=\mathrm{T}$; <br> done $=$ false; <br> while ( $\mathbf{i} \neq 0$ and done $=$ false ) do <br> \{ <br> if ( x < IDENT(i)) then <br> $\mathrm{i}=$ Lchild( i ); <br> elseif ( x > IDENT(i)) then <br> $\mathrm{i}=$ Rchild( i ); <br> else <br> done=true; <br> \} <br> \}

## DELETION FROM A BINARY SEARCH TREE

## Original Tree When we delete 44 (No children)



Original Tree


When we delete 75 (one child)



## Create the binary search tree using the following data elements 43, 10, 79, 90, 12, 54, 11, 9, 50

1. Insert 43 into the tree as the root of the tree.
2. Read the next element, if it is lesser than the root node element, insert it as the root of the left sub-tree.
3. Otherwise, insert it as the root of the right of the right sub-tree.



## CONSTRUCTING AN EXPRESSION TREE

Eg: $\mathbf{a} \mathbf{b}+\mathbf{c} \mathbf{d e + * *}$ (postfix expression)

- The first two symbols are operands and we create nodes and push pointers to the stack.

- Next " + " is read. So 2 pointers from stack are popped and new tree is formed and pointer to it is pushed on to the stack

- Next "c" , "d" and "e" are read. Node is created and its pointers push into the stack.




## GRAPH

- Graph is a nonlinear data structure.
- Graph is shown as a tuple $\mathbf{G}=(\mathbf{V}, E)$

$$
\begin{aligned}
& \mathrm{V}=\text { set of vertex }=\left\{\mathrm{v}_{1}, \mathrm{v}_{2}, \mathrm{v}_{3}, \ldots \ldots . v_{n}\right\} \\
& \mathrm{E}=\text { set of edges }=\left\{\mathrm{e}_{1}, \mathrm{e}_{2}, \mathrm{e}_{3}, \ldots \ldots . e_{n}\right\}
\end{aligned}
$$

- Graph has no hierarchal relationship.
- If a tree contains a closed loop or circuit , it is a graph.
- Tree is an acyclic graph


## Eg:



- Edges $\mathrm{E}=\left\{\mathrm{e}_{1}, \mathrm{e}_{2}, \mathrm{e}_{3}, \ldots \ldots . . e_{7}\right\}$
- Vertices $\mathrm{V}=\left\{\mathrm{v}_{1}, \mathrm{v}_{2}, \mathrm{v}_{3}, \mathrm{v}_{4}, \mathrm{v}_{5}\right\}$


## End Vertices

- The two vertices associated with any edge are called end vertices

$v_{i}$ and $v_{j}$ are the end vertex of $e_{k}$
$>$ Self Loop
- Any edge that have only one vertex are called self loop



## Parallel Edges

- In some cases, two edges are present associated with 2 vertex. This edges are called parallel edges.


Simple Graph

- A graph with neither self loop nor parallel edges are called simple graph.


## >Isomorphic Graph

- One graph is able to represent several styles is called isomorphic graph.


Fig (a)


Fig (b)

- Graph (a) and (b) are same , because edges and vertices are same in both graphs.


## Degree (valency) of graph

- The number of edges incident on a vertex $\mathrm{v}_{i}$, with self loops counted twice is called the degree $d\left(v_{i}\right)$ of vertex $v_{i}$


$$
d\left(v_{1}\right)=d\left(v_{3}\right)=d\left(v_{4}\right)=3
$$

$d\left(v_{2}\right)=4$ (self loop counted twice) $d\left(v_{5}\right)=1$

Theorem -1 : The sum of degree of all vertex in a graph G is twice the number of edges in the graph $G$

## That is $\sum_{i=1}^{n} d\left(v_{i}\right)=2 e$



The number of edges $=7$
Sum of degree of all vertices $=$

$$
\begin{array}{lc}
\mathrm{d}\left(\mathrm{v}_{1}\right)=3 & \\
\mathrm{~d}\left(\mathrm{v}_{2}\right)=4 & =(3+4+3+3+1)=14 \\
\mathrm{~d}\left(\mathrm{v}_{3}\right)=3 & \sum_{i=1}^{n} d\left(v_{i}\right)=14=2 \mathrm{e} \\
\mathrm{~d}\left(\mathrm{v}_{4}\right)=3 & \text { ( One edge will contribute } \\
\mathrm{d}\left(\mathrm{v}_{5}\right)=1 & \text { for two degree counts) }
\end{array}
$$

Theorem -2 : The number of vertices of odd degree in a graph is always even.
Consider the following graph


The total number of vertex which have odd degree is 4.
That is, $\mathrm{v}_{1}, \mathrm{v}_{3}, \mathrm{v}_{4}, \mathrm{v}_{5}$
4 is an even number

## Regular Graph

- If the degree of all vertex is same, it is called regular graph.


Degree of all vertices are 3 . So it is a regular graph

## $>$ Isolated Vertex

- A vertex having no incident edge or no edge will be incident to a vertex is called isolated vertex

$v_{6}$ and $v_{7}$ are isolated vertices Degree of isolated vertex is zero


## Pendant Vertex

- A vertex of degree one is called pendant vertex or an end vertex

$\mathrm{v}_{5}$ is pendant vertex; degree $=1$
$>$ Null Graph
- A graph without any edge is called null graph. In other words every vertex in a null graph is an isolated vertex

$$
\stackrel{\mathrm{v}_{1}}{\mathrm{v}_{2} \bullet \quad \bullet \mathrm{v}_{3}} \text { is a null graph }
$$

## Applications of Graphs

1. Graph is used to represent communication network
2. It is also used to represent electrical networks
3. Seating problem solving
4. Graph coloring
5. Project planning
6. Solving puzzles


## GRAPH TRAVERSALS

## $>2$ type traversals are performed on a graph. They are

- Depth First Search (DFS)
- Breadth First Search (BFS)
- For performing DFS, we use stack and queue for BFS


## DFS

Once at a vertex V , we scan all edges incident on V and then move to an adjacent vertex W. At W, we scan all edges incident on W. This process is continued till all edges in the graph are scanned.





## GRAPH REPRESENTATION IN MEMORY

The most commonly used representations are

- Adjacency Matrix
- Adjacency List
- Adjacency Multilist
- Sequential Representation


## 1. Adjacency Matrix

- The graph can be represented in a matrix form.
- Let $\mathrm{G}=(\mathrm{V}, \mathrm{E})$ be a graph with " n " vertices. The size of the adjacency matrix $G$ is $n \times n$.
- $\mathrm{A}[\mathrm{i}, \mathrm{j}]=1$ indicate that an edge is present between the vertex $v_{i}$ and $v_{j}$.
- $\mathrm{A}[\mathrm{i}, \mathrm{j}]=0$ indicate that an edge is not present in between the vertex $v_{i}$ and $v_{j}$.


Degree of vertex $\mathrm{v}_{1}=$ number of ones in row or column in 1 . That is 3

- If the graph is undirected, the obtaining matrix is a symmetric matrix. So we can store the upper portion or the lower portion of the matrix in memory.
Consider the following directed graph


The matrix is not symmetric. The row sum of the matrix indicate the "out degree" and the column sum indicate "in degree"

## Out degree of V1 = 1

In degree of V1 = 1
Out degree of V2 $=2$
In degree if V2 = 1

- If we represented a graph (Directed or undirected) in the adjacency matrix form, the diagonal elements are always zero. So we can avoid the searching of diagonal elements.
- Number of search $=n^{2}-n$
- Time complexity $=0\left(\mathrm{n}^{2}\right)$
- The searching time can be further reduced as $\mathrm{O}(\mathrm{n}+\mathrm{e})$ where, n is number of vertices and $e$ is number of edges. For reducing the time complexity of adjacency matrix, we use adjacency list.


## Adjacency List

- In adjacency list, head nodes and list nodes are present. The number of head node is equal to the number of vertex.
- Head node can be represented in a sequential manner or in an array form. Each head node contains a set of list nodes.

- V1 is the head node. The above figure represents, from V1, there is an edge present to $\mathrm{V} 2, \mathrm{~V} 3$ and V 4

- Consider the directed graph. Here 2 separate lists are used for representing "in degree" and "out degree"



Out degree representation using adjacency list


In degree representation using adjacency list

- Adjacency list representation is efficient when the graph is sparse.
- Consider an undirected graph of $n$ vertices

Number of head nodes $=\mathrm{n}$
Number of list nodes $=2 \mathrm{e}$
For representing each node, we need the space of log n. For representing list node, the space needed is (log $n+\log e)$
So, Total space requirement is
$n \log n+2 e(\log n+\log e)$
In the above figure, there are 3 list nodes present with V1. So the degree of V1 is 3 .

## Adjacency Multilist

- In the case of adjacency list, the total number of list node is $2 e$, because one edge contribute 2 nodes.
- For reducing the number of list nodes, we introduce adjacency Multilist. The list node structure is

| N | Vertex 1 | Vertex 2 | Path 1 | Path 2 |
| :---: | :--- | :--- | :--- | :--- |

- So each head node contains only 'e' number of list nodes, not 2e.
- $N$ is a mark bit used for analyze the edge.

- If 1,2 is included , 2,1 is also embedded. So no need to write 2,1 separately.
- The corresponding path is

$$
\begin{aligned}
& \mathrm{V} 1=\mathrm{N} 1 \rightarrow \mathrm{~N} 2 \rightarrow \mathrm{~N} 3 \\
& \mathrm{~V} 2=\mathrm{N} 1 \rightarrow \mathrm{~N} 4 \rightarrow \mathrm{~N} 5 \\
& \mathrm{~V} 3=\mathrm{N} 2 \rightarrow \mathrm{~N} 4 \rightarrow \mathrm{~N} 6 \\
& \mathrm{~V} 4=\mathrm{N} 3 \rightarrow \mathrm{~N} 5 \rightarrow \mathrm{~N} 6
\end{aligned}
$$

## Sequential Representation

- Consider a graph of ' $n$ ' vertex and ' $e$ ' edges. The total space needed to represent sequentially is, $n+2 e+1$


Total space needed is $n+2 e+1$
$=8+(2 \times 7)+1$
$=23$


