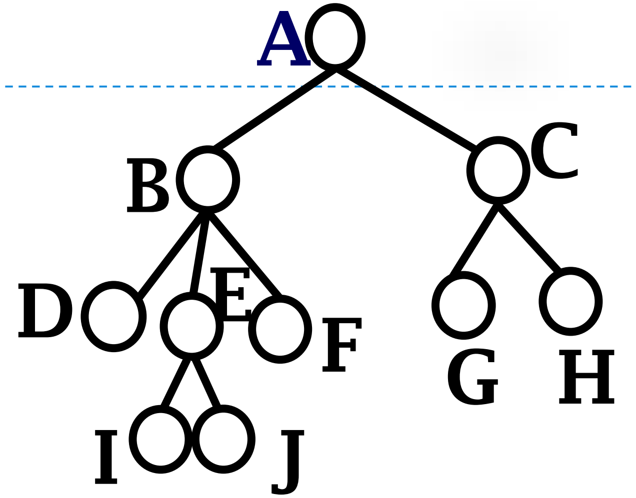
第六章　树

张铭　主讲

采用教材：张铭，王腾蛟，赵海燕 编写
高等教育出版社，2008. 6 （“十二五”国家级规划教材）

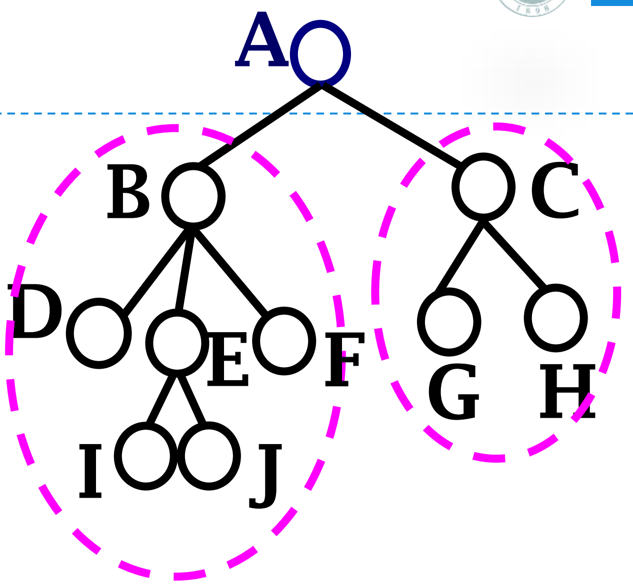
http://jpk.pku.edu.cn/course/sjjg/
https://www.icourse163.org/course/PKU-1002534001

第6章　树



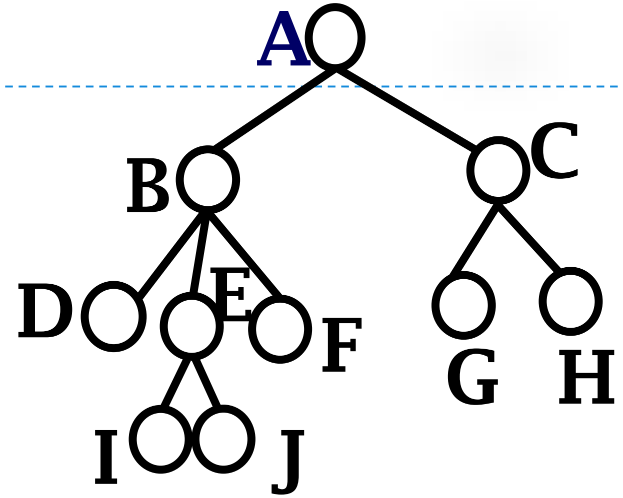
- **树的定义和基本术语**
  - **树和森林**
  - **森林与二叉树的等价转换**
  - 树的抽象数据类型
  - 树的遍历
- 树的链式存储结构
- 树的顺序存储结构
- K叉树

6.1 树的定义和基本术语



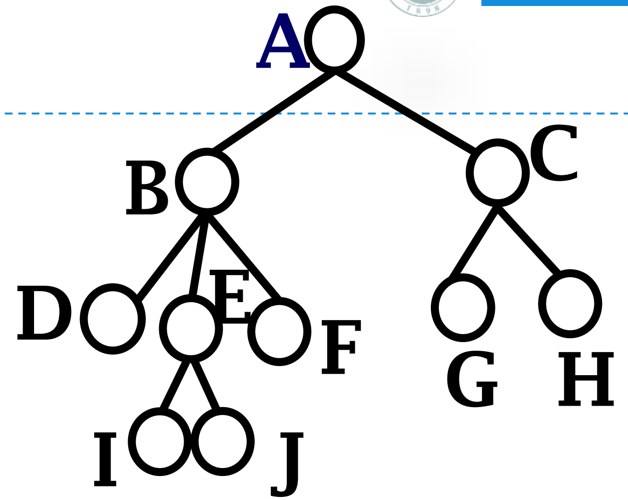
树和森林

- 树 (tree) 是包括 n 个结点的有限集合 T（n ≥ 1）：
  - 有且仅有一个特定的结点，称为 **根 (root)**
  - 除根以外的其他结点被分成 m 个 (m ≥ 0) 不相交的有限集合 T_1，T_2，…，T_m，而每一个集合又都是树，称为 T 的**子树 (subtree)**
  - 有向有序树：子树的相对次序是重要的
- 度为 2 的有序树并不是二叉树
  - 第一子结点被删除后
    第二子结点自然顶替成为第一



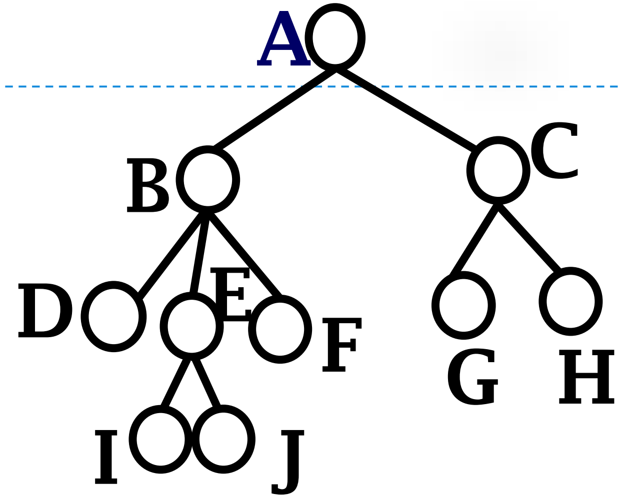
树的逻辑结构

- 包含n个结点的**有穷集合** K (n>0)，且在 K上定义了一个关系 r，关系 r 满足以下条件：
  - **有且仅有**一个结点 $k_0 \in K$，它对于关系 r 来说没有前驱。结点 k_0 称作树的 **根**
  - 除结点 k_0 外，K中的每个结点对于关系 r 来说都**有且仅有 一个前驱**
- 例如，
  - 结点集合 K={ A，B，C，D，E，F，G，H，I，J }
  - K 上的关系 r = { <A，B>，<A，C>，<B，D>，<B，E>，<B，F>，<C，G>，<C，H>，<E，I>，<E，J> }



树的相关术语

- **结点**
  - **子结点、父结点、最左子结点**
    - 若 <k，k′> ∈ r，则称 k 是 k′ 的父结点（或称“父母”），而 k′ 则是 k 的 子结点 (或“儿子”、“子女”)
  - **兄弟结点前兄弟、后兄弟**
    - 若有序对 <k，k′>及 <k，k″>∈ r，则称 k′和 k″互为兄弟
  - **分支结点、叶结点**
    - 没有子树的结点称作 叶结点（或树叶、终端结点）
    - 非终端结点称为 分支结点

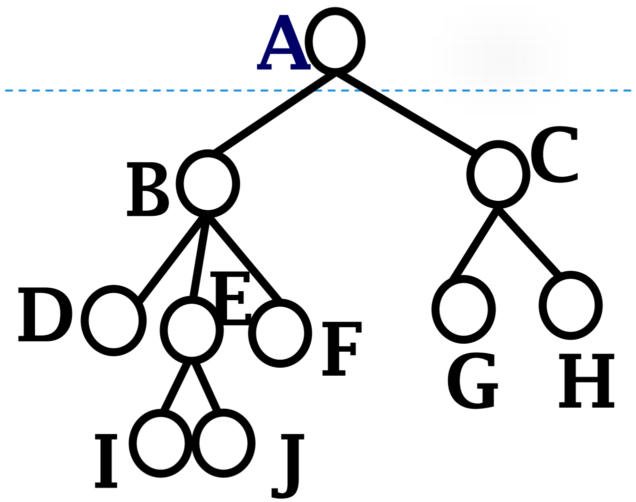


树的相关术语

- **边**
  - **两个结点的有序对**，称作边
- **路径、路径长度**
  - **除结点k_0外的任何结点 $k \in K$，都存在一个结点序列 k_0，k_1，…，k_s，使得 k_0 就是树根，且 $k_s=k$，其中有序对 $<k_{i-1}, k_i> \in r$（$1 \le i \le s$）。该序列称为从根到结点 k 的一条路径，其路径长度为 s (包含的边数)**
- **祖先、后代**
  - 若有一条由 **k 到达 k_s 的路径**，则称 k 是 k_s 的祖先，k_s 是 k 的子孙

树的相关术语

- **度数**：一个结点的子树的个数
- **层数**：根为第 0 层
  - 其他结点的层数等于其父结点的层数加 1
- **深度**：层数最大的叶结点的层数
- **高度**：层数最大的叶结点的层数加 1

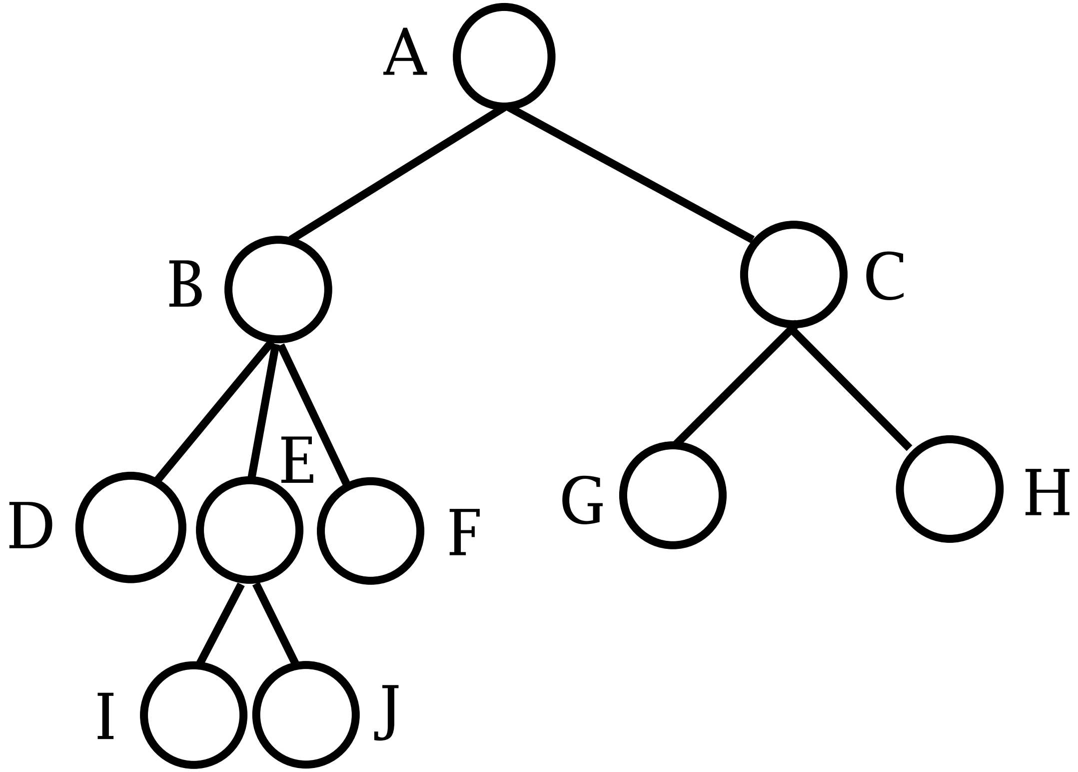


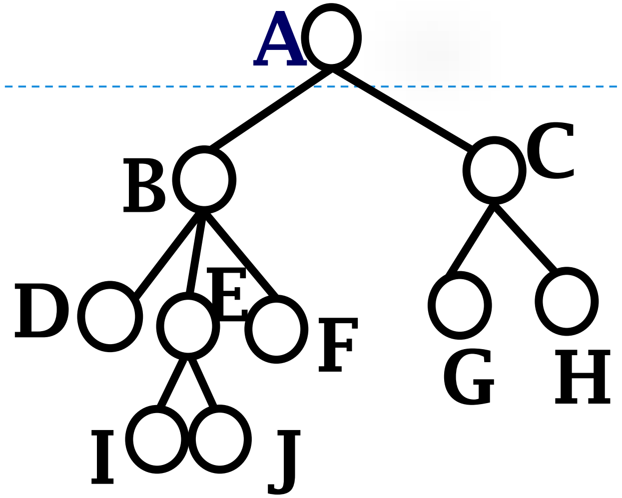
树形结构的各种表示法

- 树形表示法
- 形式语言表示法
- 文氏图表示法
- 凹入表表示法
- 嵌套括号表示法

6.1 树的定义和基本术语

树形表示法



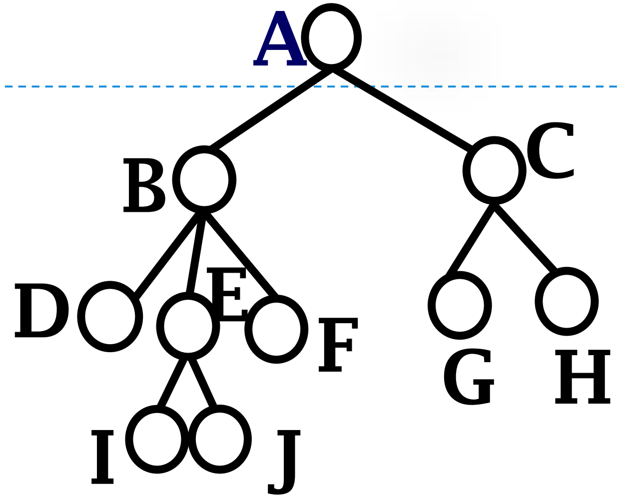


形式语言表示法

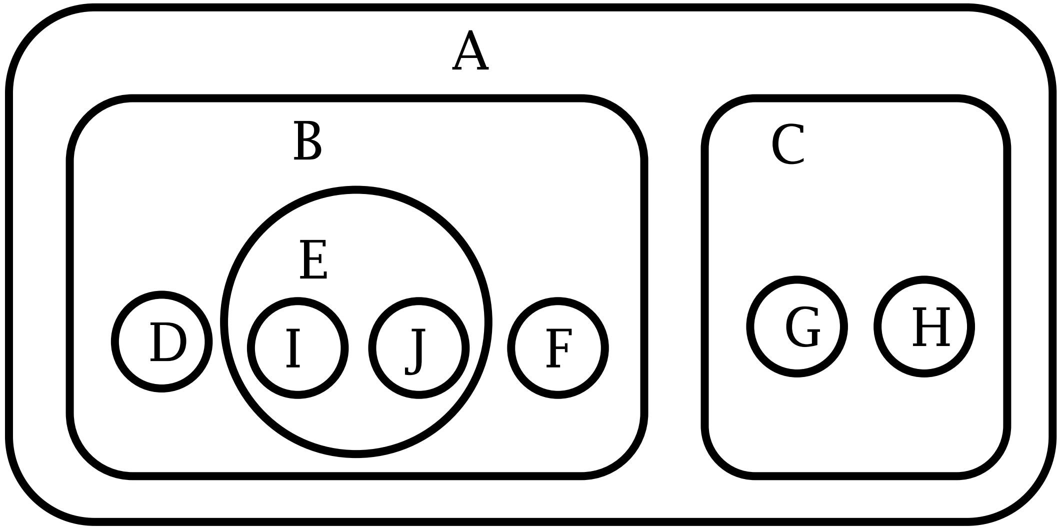
树的逻辑结构是：

结点集合K={A，B，C，D，E，F，G，H，I，J}

K上的**关系**N={<A，B>，<A，C>，<B，D>，
<B，E>，<B，F>，<C，G>，
<C，H>，<E，I>，<E，J>}

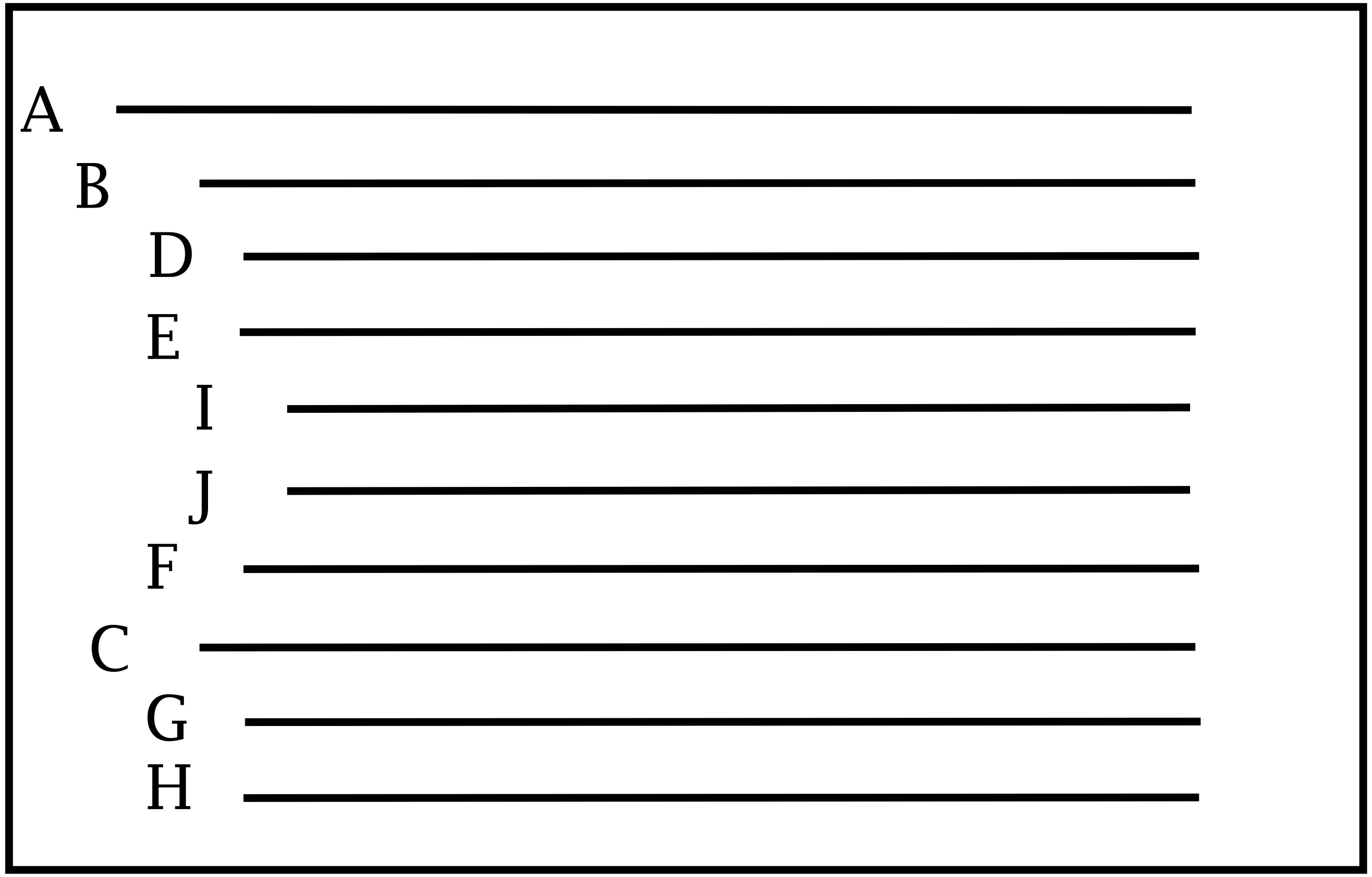


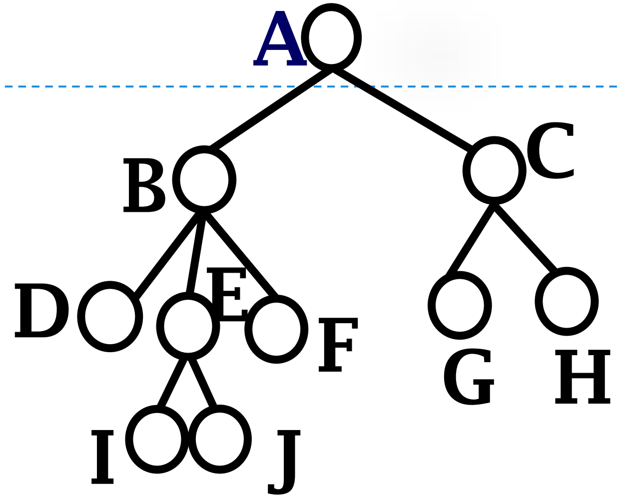
文氏图到嵌套括号表示的转化



(A(B(D)(E(I)(J))(F))(C(G)(H)))

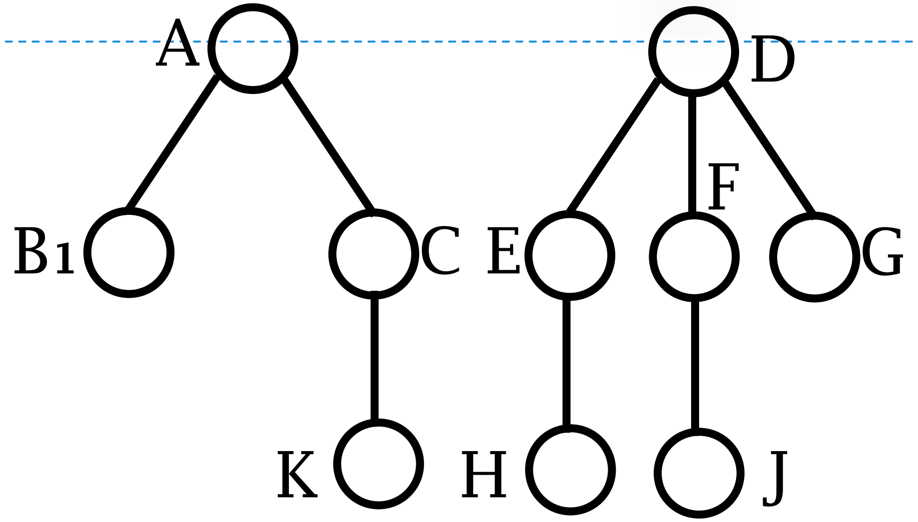
凹入表表示法





图书目录，杜威表示法

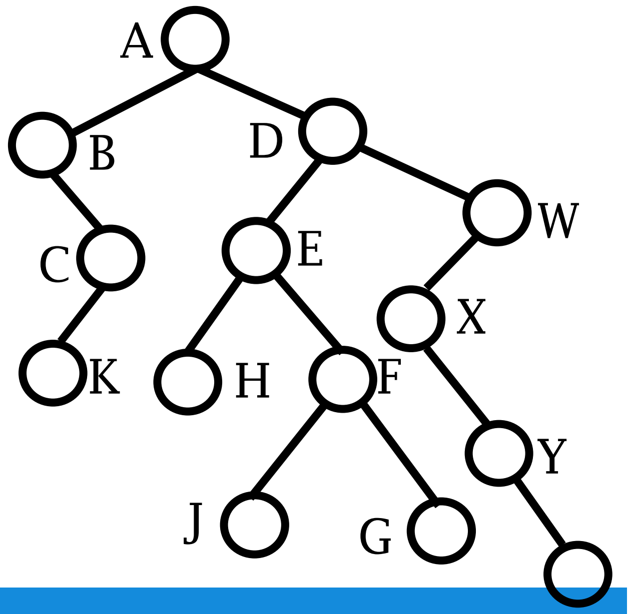
森林与二叉树的等价转换

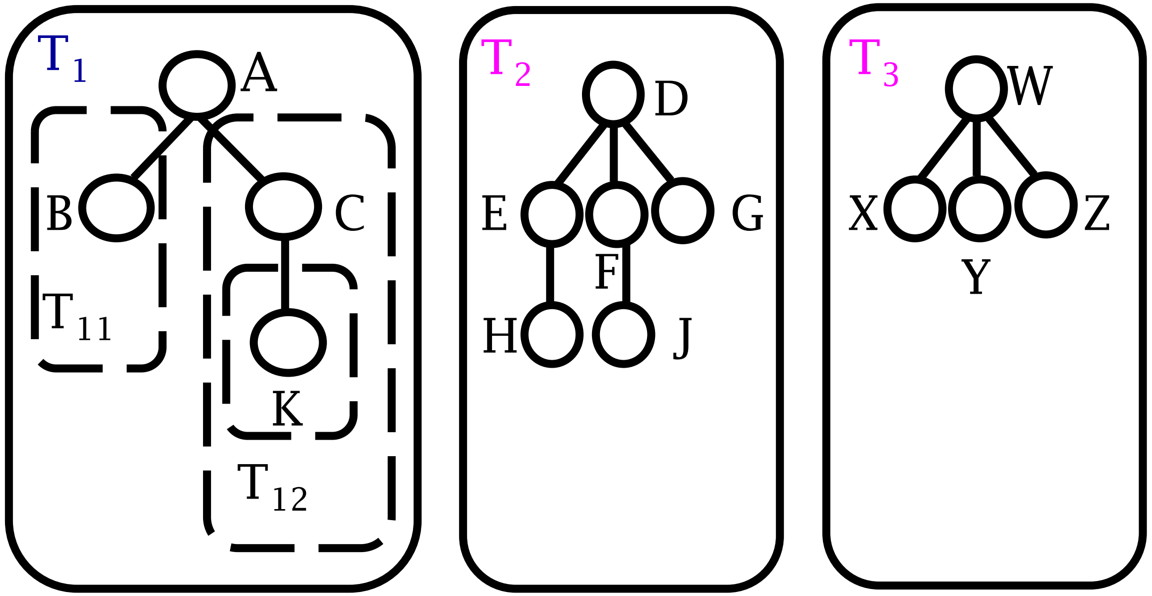


- **森林(forest)**：零棵或多棵 **不相交** 的树的集合（通常是有序）
- **树与森林的对应**
  - 一棵树，删除树根，其子树就组成了森林
  - 加入一个结点作为根，森林就转化成了一棵树
- **森林与二叉树之间可以相互转化**，而且这种转换是**一一对应**的
  - 森林的相关操作都可以转换成对二叉树的操作

森林转化成二叉树的形式定义

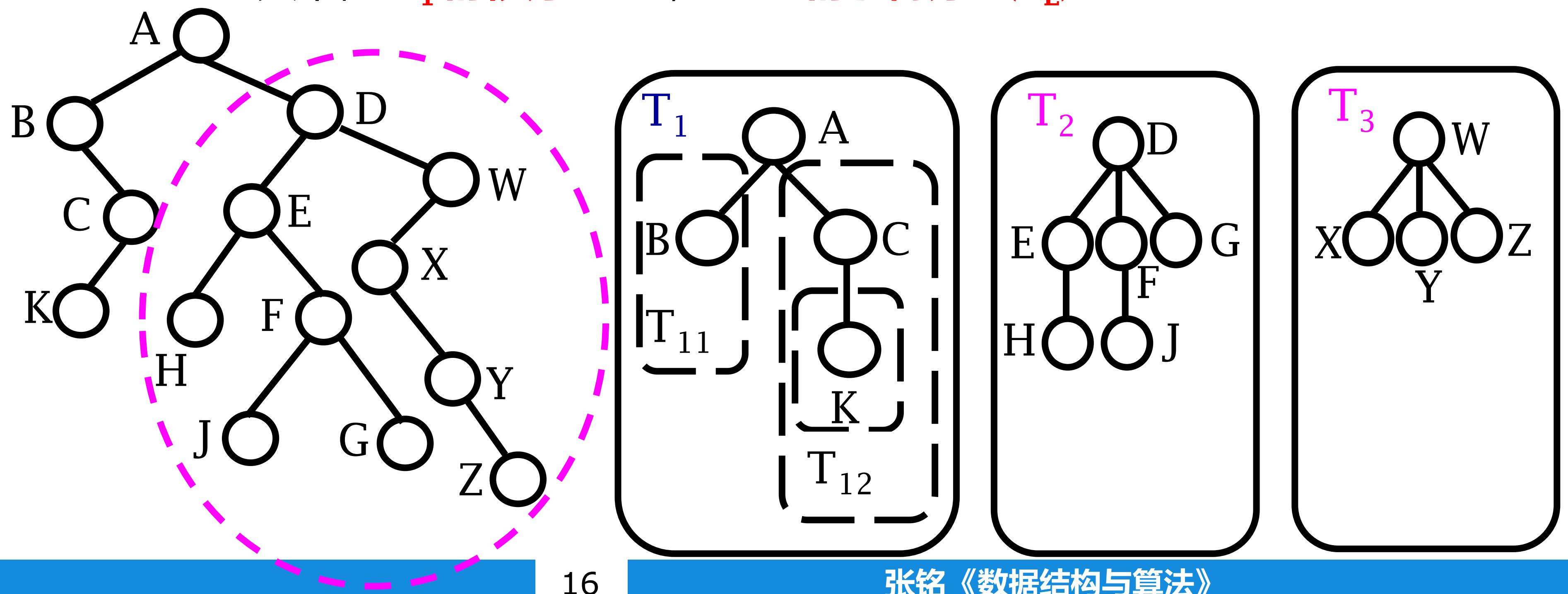
- 有序集合 F = {T_1, T_2, …, T_n} 是树 T_1, T_2, …, T_n 组成的森林，递归转换成二叉树B(F)：
  - 若 F 为空，即 n = 0，则 B(F) 为空。
  - 若 F 非空，即 n > 0，则 **B(F) 的根**是森林中**第一棵树 T_1 的根 W_1**，
    B(F) 的**左子树**是树 T_1 中**根结点 W_1的子树森林** F = {T_{11}, …, T_{1m}} **转换成的二叉树B(T_{11}, …, T_{1m})**；
    B(F)的**右子树是从森林 F’ = {T_2, …, T_n} 转换而成的二叉树**





二叉树转化成森林或树的形式定义

- 设B是一棵二叉树，**root**是 B 的根，**B_L** 是 root 的左子树，**B_R** 是 root 的右子树，则对应于二叉树B的森林或树 F(B) 的形式定义是：
  - 若 B 为空，则 F(B) 是空的森林
  - 若 B 不为空，则 F(B)是一棵树T_1加上森林 F(B_R)，其中树 **T_1 的根为 root**，**root 的子树为 F(B_L)**



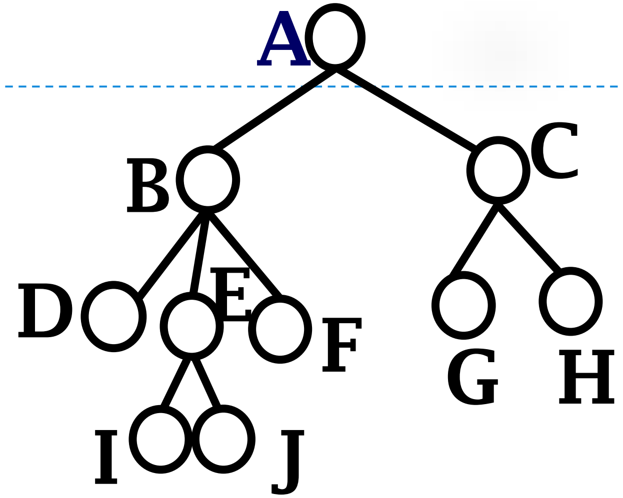
思考

- 1. 树也是森林吗？

- 2. 为什么要建立二叉树与森林的对应关系？

第6章　树

- 树的定义和基本术语
  - 树和森林
  - 森林与二叉树的等价转换
  - 树的抽象数据类型
  - 树的遍历
- 树的链式存储结构
- 树的顺序存储结构
- K叉树



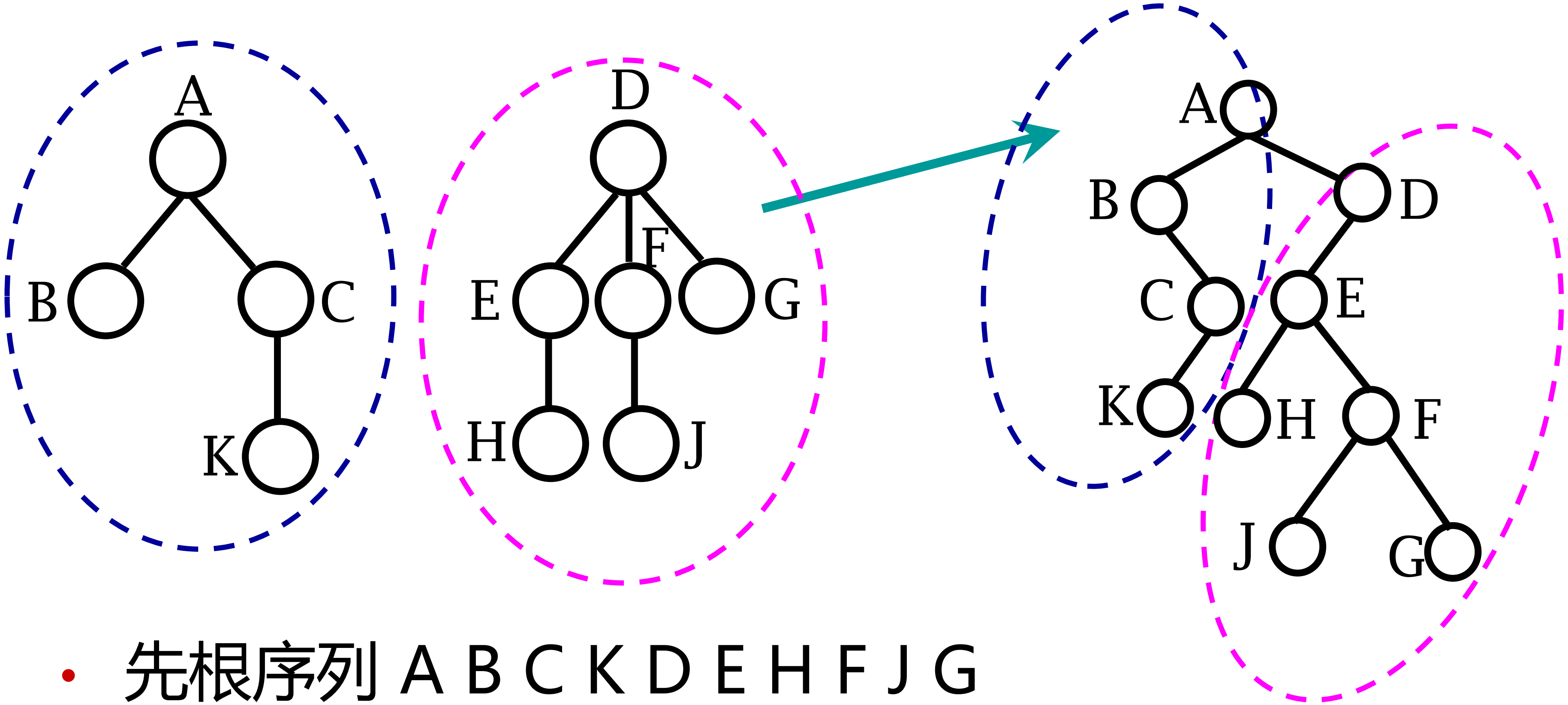
树的抽象数据类型

```
template<class T>
class TreeNode  {                                    // 树结点的ADT
public:
  TreeNode(const T& value);                          // 拷贝构造函数
  virtual ~TreeNode() {};                            // 析构函数
  bool isLeaf();                                     // 判断当前结点是否为叶结点
  T Value();                                         // 返回结点的值
  TreeNode<T> *LeftMostChild();                      // 返回第一个左孩子
  TreeNode<T> *RightSibling();                       // 返回右兄弟
  void setValue(const T& value);                     // 设置当前结点的值
  void setChild(TreeNode<T> *pointer);               // 设置左孩子
  void setSibling(TreeNode<T> *pointer);             // 设置右兄弟
  void InsertFirst(TreeNode<T> *node);               // 以第一个左孩子身份插入结点
  void InsertNext(TreeNode<T> *node);                // 以右兄弟的身份插入结点
};
```

树的抽象数据类型

```
template<class T>
class Tree {
public:
    Tree();                                          // 构造函数
    virtual ~Tree();                                 // 析构函数
    TreeNode<T>* getRoot();                          // 返回树中的根结点
    void CreateRoot(const T& rootValue);             // 创建值为rootValue的根结点
    bool isEmpty();                                  // 判断是否为空树
    TreeNode<T>* Parent(TreeNode<T> *current);       // 返回父结点
    TreeNode<T>* PrevSibling(TreeNode<T> *current); //返回前一个兄弟
    void DeleteSubTree(TreeNode<T> *subroot);        // 删除以subroot子树
    void RootFirstTraverse(TreeNode<T> *root);       // 先根深度优先遍历树
    void RootLastTraverse(TreeNode<T> *root);        // 后根深度优先遍历树
    void WidthTraverse(TreeNode<T> *root);           // 广度优先遍历树
};
```

森林的遍历



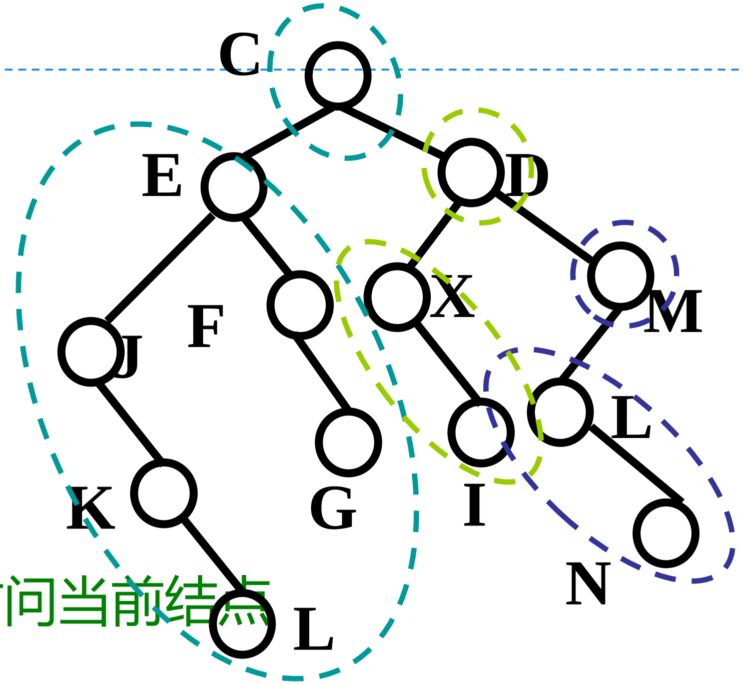
- 先根序列 A B C K D E H F J G
- 后根序列 B K C A H E J F G D

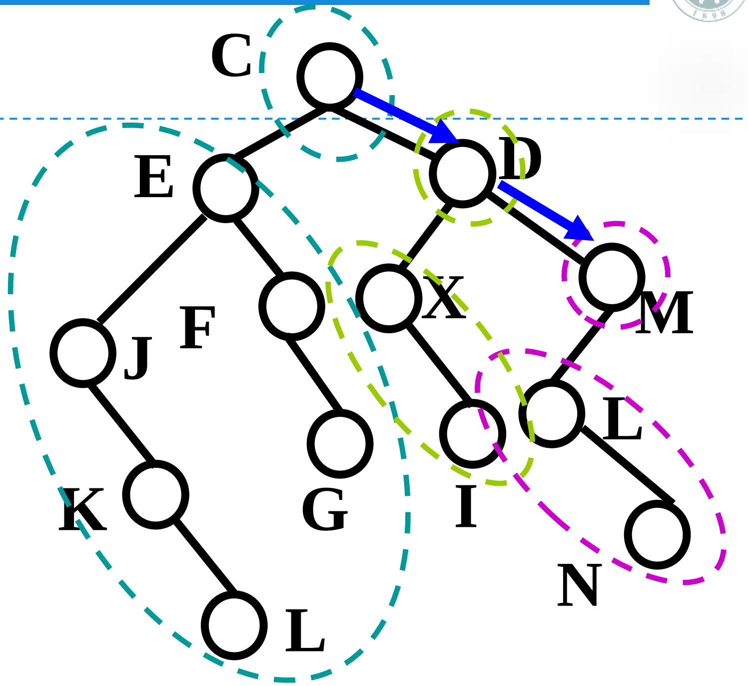
遍历森林vs遍历二叉树

- **先根**次序遍历森林
  - **前序法**遍历二叉树
- **后根**次序遍历森林
  - 按**中序法**遍历对应的二叉树
- **中根**遍历?
  - 无法明确规定根在哪两个子结点之间

先根深度优先遍历森林

```
template<class T>
void Tree<T>::RootFirstTraverse(
        TreeNode<T> * root)  {
    while (root != NULL) {
        Visit(root->Value());                    // 访问当前结点
        // 遍历第1棵树根的子树森林(树根除外)
        RootFirstTraverse(root->LeftMostChild());
        root = root->RightSibling();             // 遍历其他树
    }
}
```





后根深度优先遍历森林

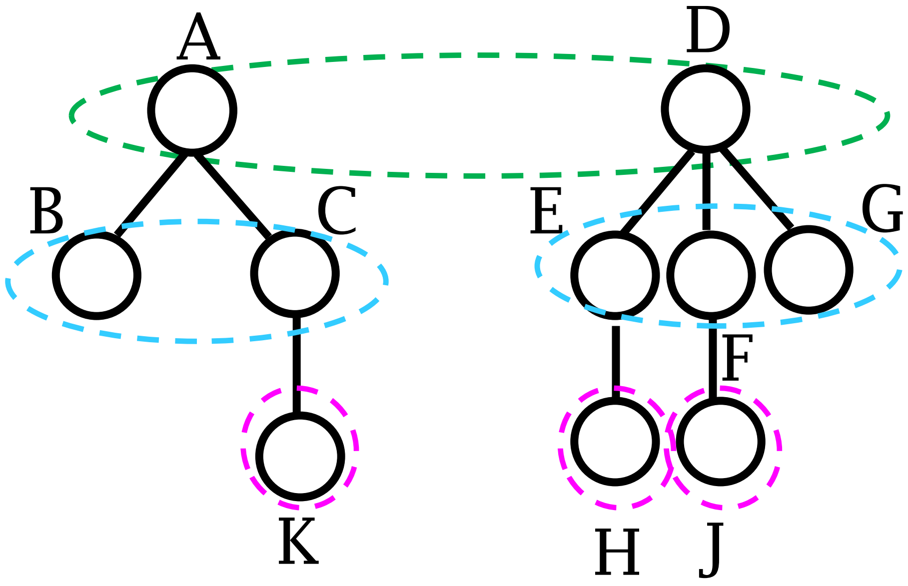
```
template<class T>
void Tree<T>::RootLastTraverse(
        TreeNode<T> * root)  {
    while (root != NULL) {
        // 遍历第一棵树根的子树森林
        RootLastTraverse(root->LeftMostChild());
        Visit(root->Value());            // 访问当前结点
        root = root->RightSibling();  // 遍历其他树
    }
}
```

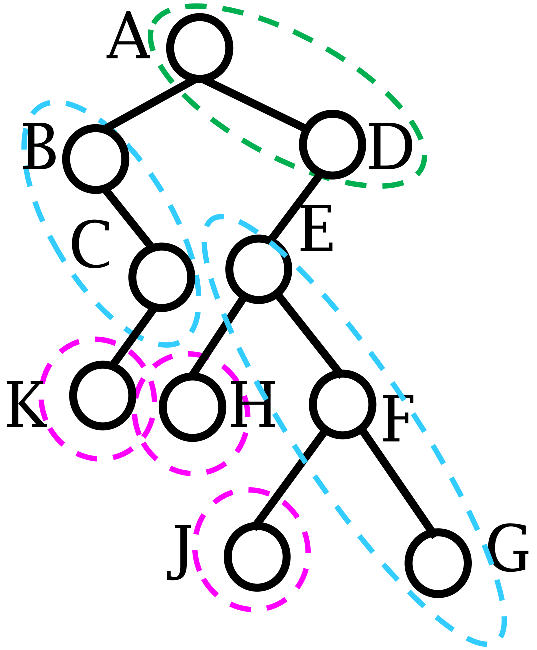
宽度优先遍历森林

- 宽度优先遍历
  - 也称广度优先遍历
  - 或称层次遍历

- a) 首先依次访问层数为0的结点
- b) 然后依次访问层数为1的结点
- c) 直到访问完最下一层的所有结点

广度优先遍历森林





- 森林广度优先：A D B C E F G K H J
- **看二叉链存储结构的右斜线**

广度优先遍历森林

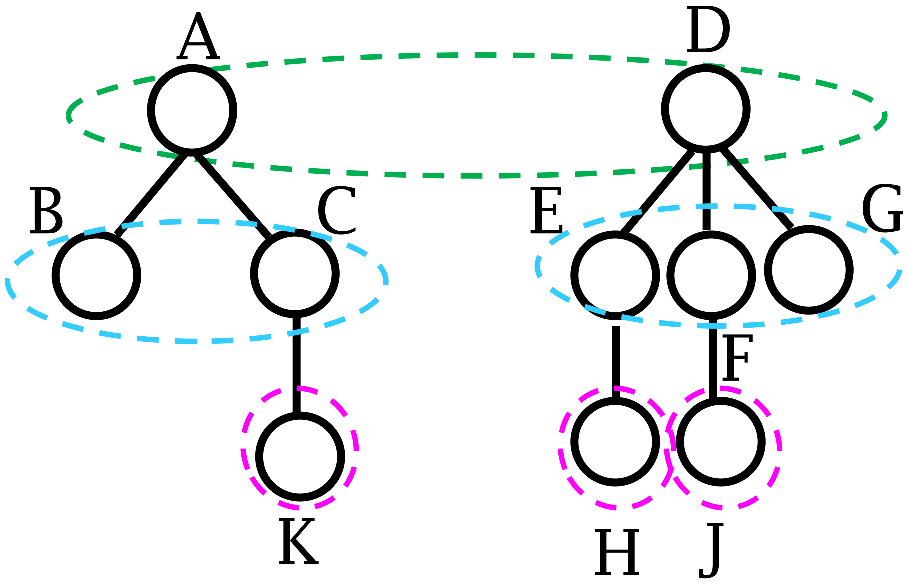
```
template<class T>
void Tree<T>::WidthTraverse(TreeNode<T> * root)  {
   using std::queue;                                   // 使用STL队列
   queue<TreeNode<T>*> aQueue;
   TreeNode<T> * pointer = root;
   while (pointer != NULL) {
      aQueue.push(pointer);                            // 当前结点进入队列
      pointer = pointer->RightSibling();               // pointer指向右兄弟
   }
 while (!aQueue.empty()) {
      pointer = aQueue.front();                        // 获得队首元素
      aQueue.pop();                                    // 当前结点出队列
      Visit(pointer->Value());                         // 访问当前结点
      pointer = pointer-> LeftMostChild();  // pointer指向最左孩子
      while (pointer != NULL) {                        // 当前结点的子结点进队列
            aQueue.push(pointer);
            pointer = pointer->RightSibling();
  }  }  }
```

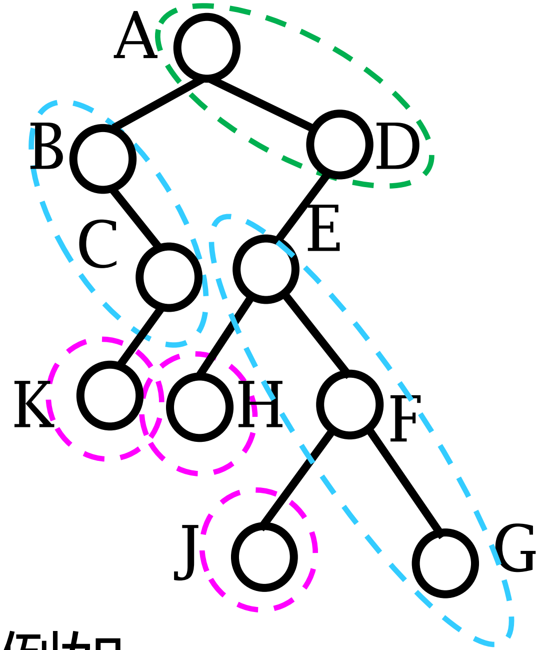
思考

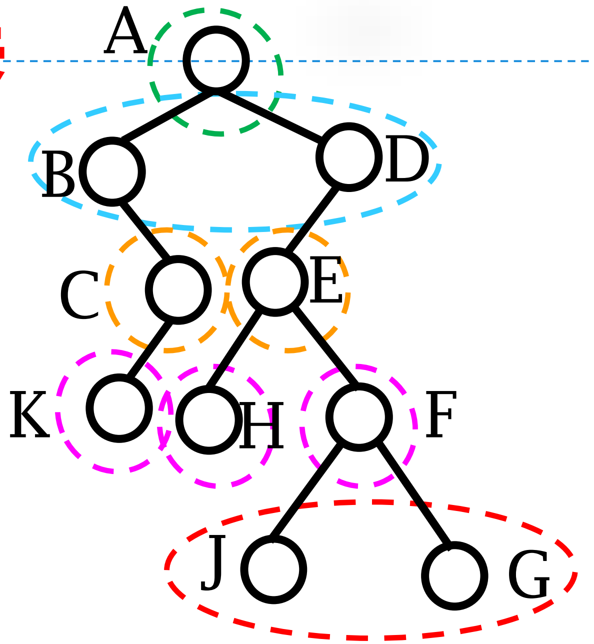
- 1. 能否直接用二叉树前序遍历框架来编写森林的先根遍历?
- 2. 能否直接用二叉树中序遍历框架来编写森林的后根遍历?
- 3. 森林的非递归深搜框架?

错误

思考：宽搜的各种观点

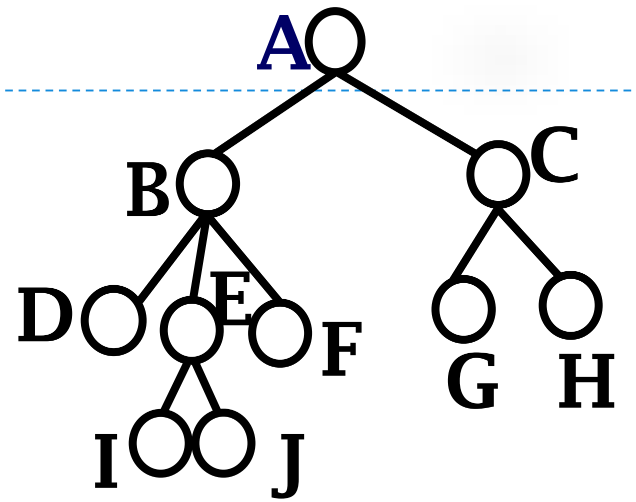






- 不能用二叉树的广度遍历模板。例如，
  - 上左图，森林广度优先：A D B C E F G K H J
    - **看二叉树的右斜线**
  - 上右图，二叉树广度： A B D C E K H F J G
    - **看平行横线**

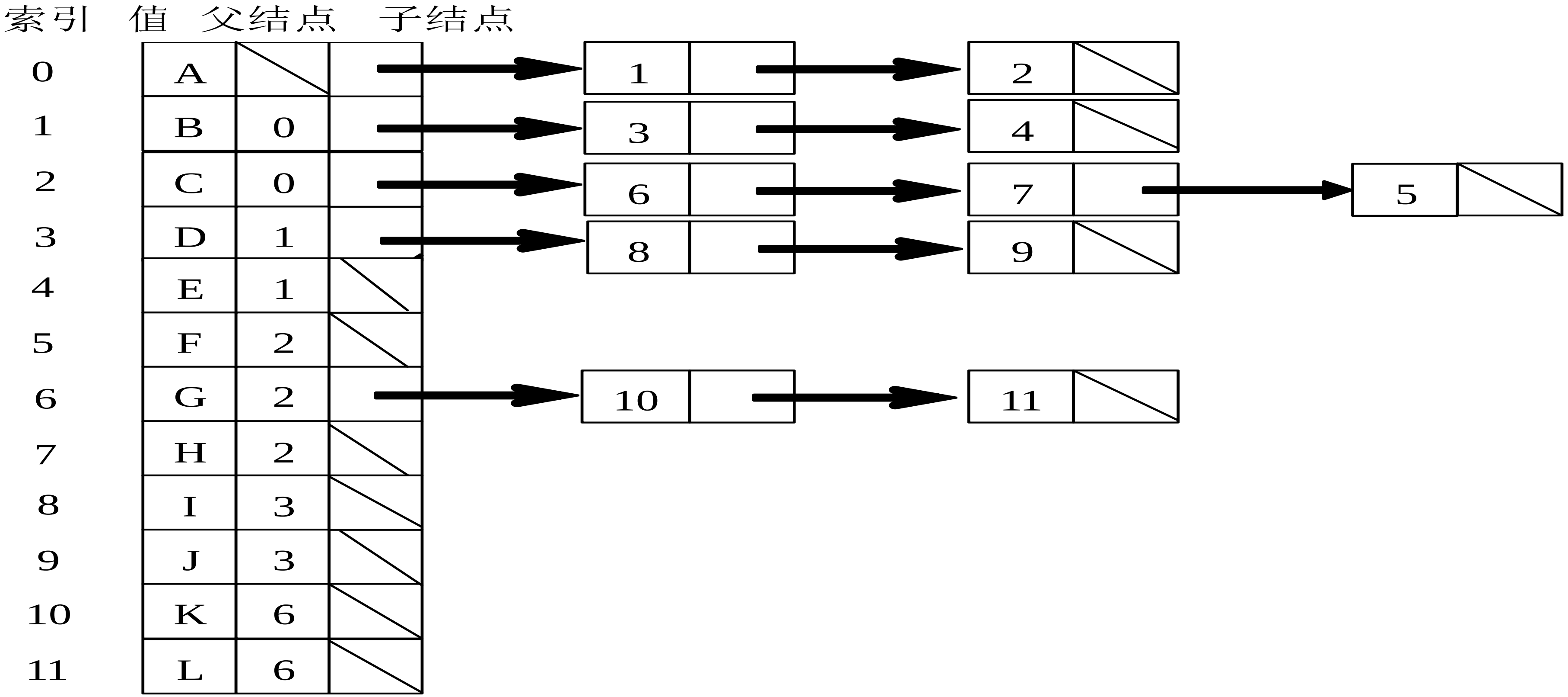
第6章　树



- 树的定义和基本术语
- **树的链式存储结构**
  - “子结点表”表示方法
  - 静态“左孩子/右兄弟”表示法
  - 动态表示法
  - 动态“左孩子/右兄弟”表示法
  - 父指针表示法及其在并查集中的应用
- 树的顺序存储结构
- K叉树

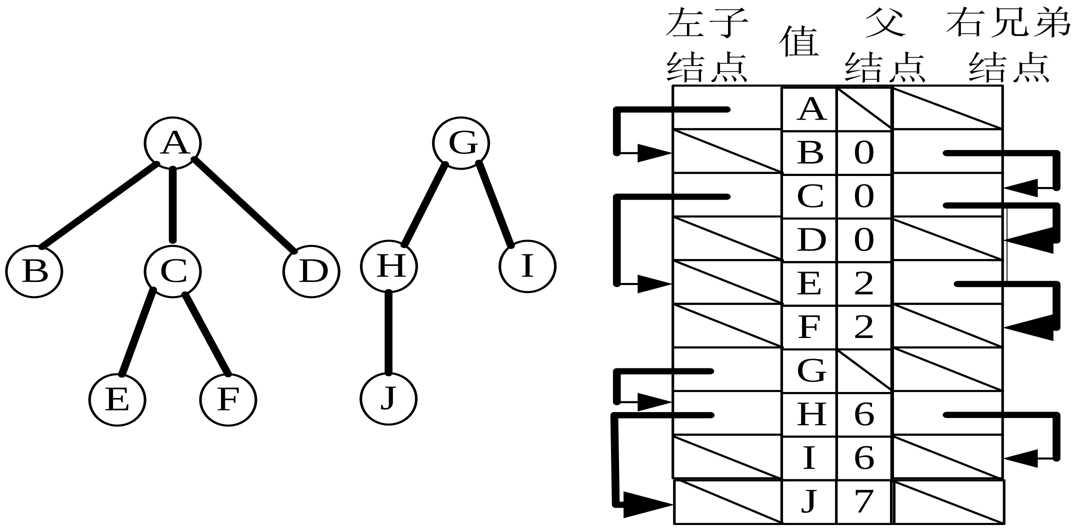
“子结点表”表示方法

list of children，就是图的邻接表。



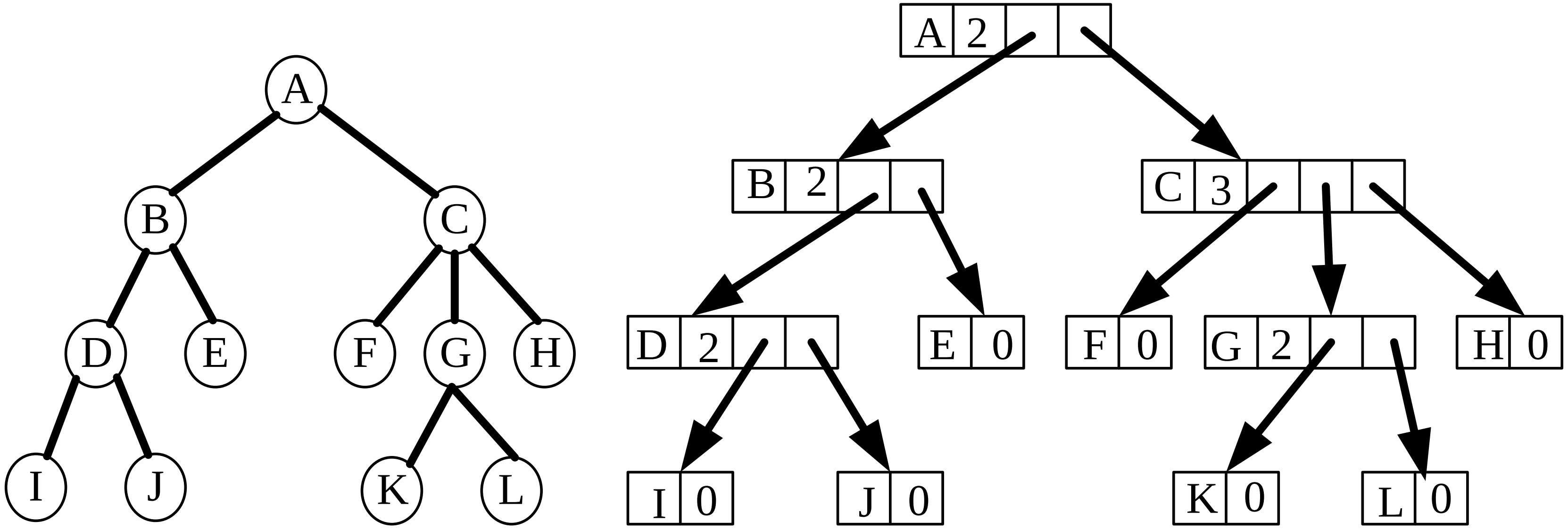
静态“左孩子/右兄弟”表示法

- 在数组中存储的“子结点表”



动态表示法

- 每个结点分配**可变的存储空间**
  - 子结点数目发生变化，需要重新分配存储空间

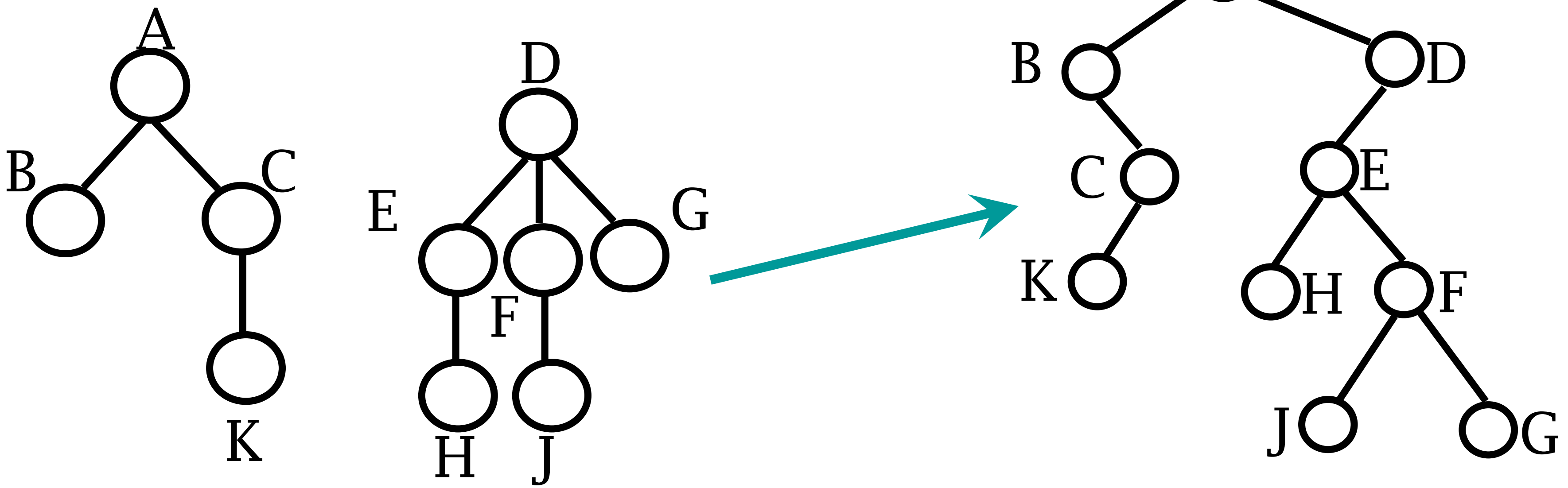


(a) 树　　(b) 树的实现

动态“左子/右兄”二叉链表示法

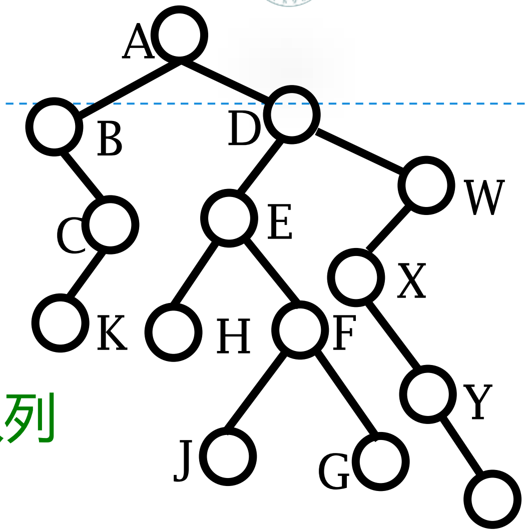
- 左孩子在树中是结点的最左子结点，右子结点是结点原来的右侧兄弟结点
- 根的右链就是森林中每棵树的根结点

pChild	info	pSibling



动态二叉链表树的关键实现细节

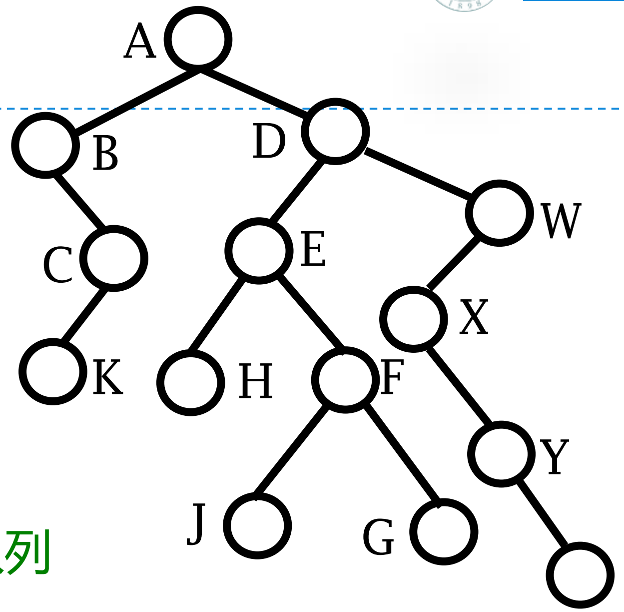
```
// 在TreeNode的抽象类中增加以下私有数据成员
private:
T m_Value;                      // 树结点的值
TreeNode<T> *pChild;            // 第一个左孩子指针
TreeNode<T> *pSibling;          // 右兄弟指针
```



寻找当前结点的父结点

```
template<class T>
TreeNode<T>* Tree<T>::Parent(TreeNode<T> *current)  {
    using std::queue;                                       // 使用STL队列
    queue<TreeNode<T>*> aQueue;
    TreeNode<T> *pointer = root;
    TreeNode<T> *father = upperlevelpointer  = NULL;        // 记录父结点
    if (current != NULL && pointer != current)  {
        while (pointer != NULL) {                           // 森林中所有根结点进队列
            if (current == pointer)                         // 森林中所有第一层根的父为空
                break;
            aQueue.push(pointer);                           // 当前结点进队列
            pointer=pointer-> RightSibling();               // 指针指向右
}
```

6.2 树的链式存储结构



寻找当前结点的父结点

```
  while (!aQueue.empty()) {
    pointer = aQueue.front();                     // 取队列首结点指针
    aQueue.pop();                                 // 当前元素出队列
    upperlevelpointer  = pointer;                 // 指向上一层的结点
    pointer = pointer-> LeftMostChild();          // 指向最左孩子
    while (pointer) {                             // 当前结点的子结点进队列
        if (current == pointer) {
           father = upperlevelpointer;            // 返回父
           break;}
        else {
           aQueue.push(pointer); pointer = pointer->RightSibling();}
     }
  }
 }
 aQueue. clear( );                                // 清空队列，也可以不写（局部变量）
 return father;
}
```

删除以root为代表的森林的所有结点

```
template <class T>
void Tree<T>::DestroyNodes(TreeNode<T>* root)  {
    if (root) {
        DestroyNodes(root->LeftMostChild());//递归删除第一子树
        DestroyNodes(root->RightSibling());   //递归删除其他子树
        delete root; //删除根结点
    }
}
```

思考：删除以subroot为根的子树

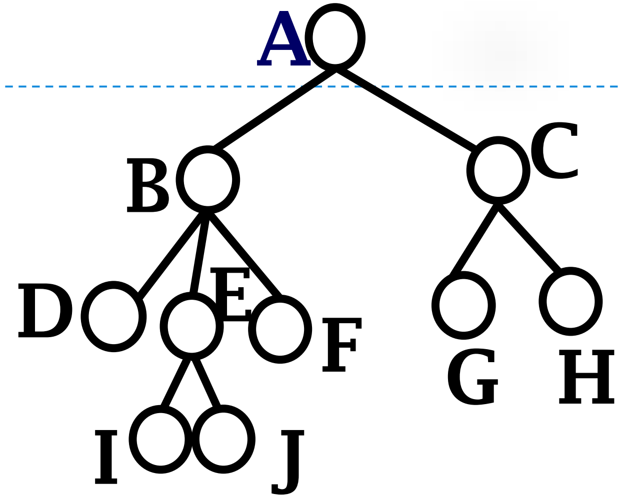
请注意待删除的子树是否为空、subroot有无父指针等情况的判断。

考虑删除以后各项相关链接的修改顺序。

删除以subroot为根的子树

```
template<class T>
void Tree<T>::DeleteSubTree(TreeNode<T> *subroot)  {
    if (subroot == NULL) return;                   // 若待删除的子树为空则返回
    TreeNode<T> *pointer = Parent (subroot);       // 找subroot的父结点
    if (pointer == NULL) {                         // subroot没有父，则是某个树根
      pointer =  root;
      while (pointer->RightSibling() != subroot) // 顺森林各树根的右链找左邻树根
        pointer = pointer->RightSibling();
      pointer->setSibling(subroot->RightSibling());      // 前后挂接，脱链
    }
    else if (pointer->LeftMostChild() == subroot)        // subroot为最左子
      pointer->setChild(subroot->RightSibling());        // 挂新的最左
    else {                                               // subroot有左兄弟的情况
      pointer = pointer->LeftMostChild();                // 下降到最左兄弟
      while (pointer->RightSibling() != subroot)         // 顺右链找左邻兄弟
        pointer = pointer->RightSibling();
      pointer->setSibling(subroot->RightSibling());      // 前后挂接，脱链
    }
    subroot->setSibling(NULL);                           // 非常重要，丢了会出错
    DestroyNodes(subroot); }                             // 删除以subroot代表的子森林的所有结点
```

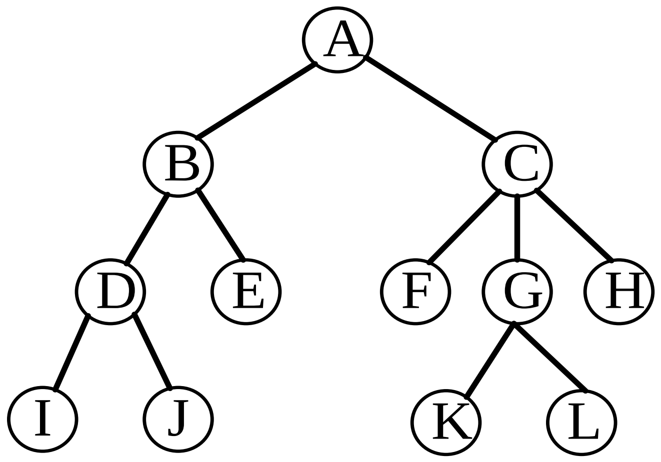
第6章 树



- 树的定义和基本术语
- **树的链式存储结构**
  - “子结点表”表示方法
  - 静态“左孩子/右兄弟”表示法
  - 动态表示法
  - 动态“左孩子/右兄弟”表示法
  - **父指针表示法及其在并查集中的应用**
- 树的顺序存储结构
- K叉树

父指针表示法

- 只需要知道父结点的应用
- 只需要保存一个指向其父结点的指针域，称为 父指针(parent pointer)表示法
- 用数组存储树结点，同时在每个结点中附设一个指针指示其父结点的位置



结点索引	0	1	2	3	4	5	6	7	8	9	10	11
值	A	B	C	D	E	F	G	H	I	J	K	L
父结点索引	╲	0	0	1	1	2	2	2	3	3	6	6

父指针表示法：算法

- 查询**结点的根**
  - 从一个结点出发找出一条向上延伸到达根的祖先路径
    - O(k)，k为树高
- 判断两个结点是否**同属一棵树**
  - 两个结点根结点相同，它们一定在同一棵树中
  - 如果其根结点不同，那么两个结点就不在同一棵树中

并查集

并查集 是一种特殊的集合，由一些不相交子集构成，合并并查集的基本操作是：

- **Find**： **查询**结点所在集合
- **Union**： **归并**两个集合

- 并查集是重要的抽象数据类型
  - 应用于求解**等价类**等等问题

等价关系

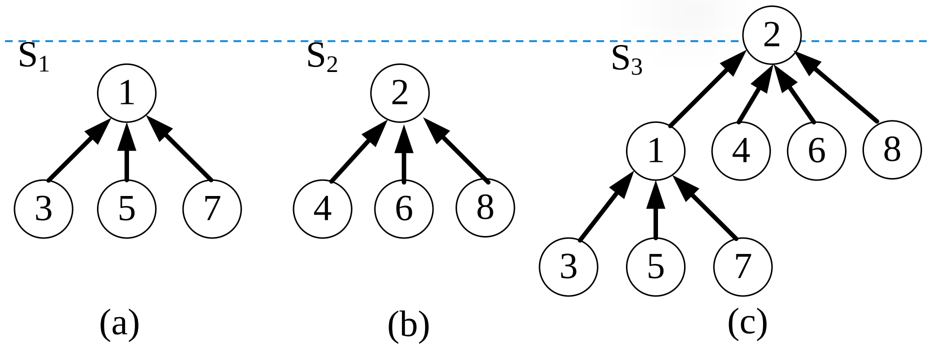
- 一个具有 n 个元素的集合 S，另有一个定义在集合 S 上的 r 个关系的关系集合 R。x，y，z表示集合中的元素
- 若关系 R 是一个 **等价关系**，当且仅当如下条件为真时成立：
  - (a) 对于所有的 x，有 (x，x)∈R（即关系是**自反**的）
  - (b) 当且仅当 (x，y)∈R 时 (y，x)∈R（即关系是**对称**的）
  - (c) 若 (x，y)∈R 且 (y，z)∈R，则有 (x，z)∈R（即关系是**传递**的）
- 如果(x，y)∈R，则元素x和y是等价的

等价类(equivalence classes)

- 等价类是指相互**等价的元素**所组成的**最大集合**。所谓最大，就是指不存在类以外的元素，与类内部的元素等价

- 由$x \in S$生成的一个R等价类
  - $[x]_R = \{y \mid y \in S \wedge xRy\}$
  - R将S划分成为r个不相交的划分S_1，S_2，…S_r，这些集合的并为S

用树来表示等价类的并查



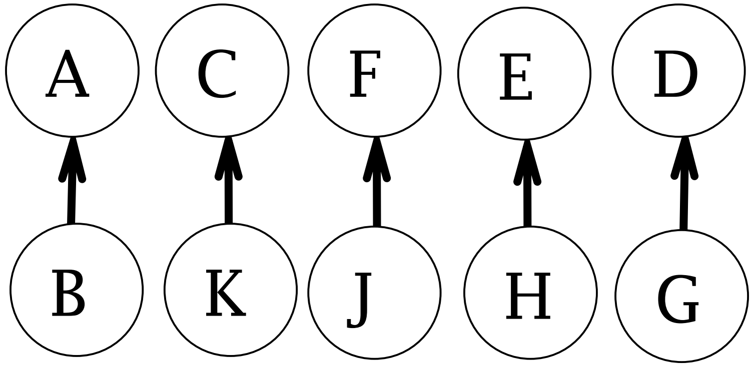
- 用一棵树代表一个集合
  - 集合用父结点代替
  - 若两个结点在**同一棵树**中，则它们处于**同一个集合**
- 树的实现
  - 存储在**静态指针数组**中
  - 结点中仅需保存**父指针**信息

UNION/FIND算法示例(1)

对这5个等价对进行处理 (A,B)、(C,K) 、(J,F) 、(H,E) 、(D,G)

	0		2				4	5	6
A	B	C	K	D	E	F	G	H	J
0	1	2	3	4	5	6	7	8	9

(A,B)(C,K)(J,F)(E,H)(D,G)

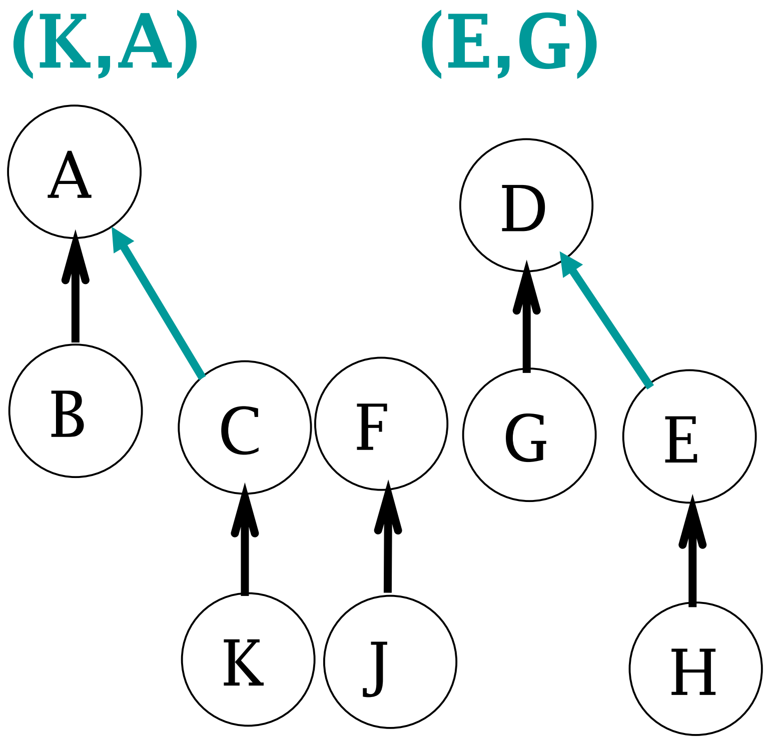


UNION/FIND算法示例(1)

然后对两个等价对（K，A）和（E，G）进行处理

K所在树的根为C，A自己是根结点，A！=C，所以两棵树要合并

	0	0	2		4		4	5	6
A	B	C	K	D	E	F	G	H	J
0	1	2	3	4	5	6	7	8	9



树的父指针表示与Union/Find算法实现

```
template<class T>
class ParTreeNode  {                              //树结点定义
private:
  T value;                                        //结点的值
  ParTreeNode<T>* parent;                         //父结点指针
  int nCount;                                     //集合中总结点个数
public:
  ParTreeNode();                                  //构造函数
  virtual ~ParTreeNode(){};                       //析构函数
  T getValue();                                   //返回结点的值
  void  setValue(const T& val);                   //设置结点的值
  ParTreeNode<T>*  getParent();                   //返回父结点指针
  void  setParent(ParTreeNode<T>* par);           //设置父指针
  int  getCount();                                //返回结点数目
  void  setCount(const int count);                //设置结点数目
};
```

树的父指针表示与Union/Find算法实现

```
template<class T>
class ParTree  {                                    // 树定义
public:
  ParTreeNode<T>* array;                            // 存储树结点的数组
  int Size;                                         // 数组大小
  ParTreeNode<T>*
  Find(ParTreeNode<T>* node) const;                 // 查找node结点的根结点
  ParTree(const int size);                          // 构造函数
  virtual ~ParTree();                               // 析构函数
  void Union(int i,int j);                          // 把下标为i，j的结点合并成一棵子树
  bool Different(int i,int j);                      // 判定下标为i，j的结点是否在一棵树中
};
```

树的父指针表示与Union/Find算法实现

```
template <class T>
ParTreeNode<T>*
ParTree<T>::Find(ParTreeNode<T>* node) const
{
  ParTreeNode<T>* pointer=node;
  while ( pointer->getParent() != NULL )
    pointer=pointer->getParent();
  return pointer;
}
```

树的父指针表示与Union/Find算法实现

```
template<class T>
void ParTree<T>::Union(int i,int j) {
ParTreeNode<T>* pointeri = Find(&array[i]);     //找到结点i的根
ParTreeNode<T>* pointerj = Find(&array[j]);     //找到结点j的根
if (pointeri != pointerj) {
  if(pointeri->getCount() >= pointerj->getCount()) {
      pointerj->setParent(pointeri);
      pointeri->setCount(pointeri->getCount() +
                              pointerj->getCount());
  }
  else {
      pointeri->setParent(pointerj);
      pointerj->setCount(pointeri->getCount() +
                              pointerj->getCount());
   }
 }
}
```

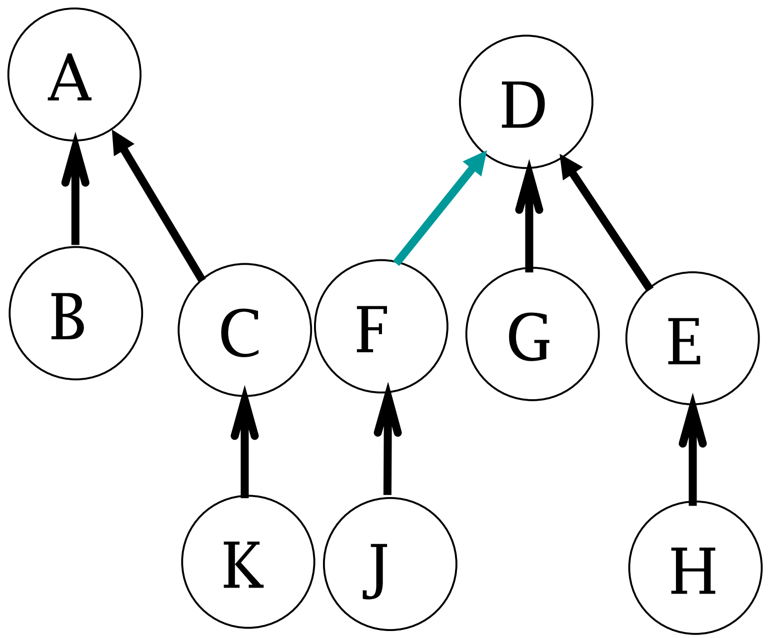
UNION/FIND算法示例(2)

最后按**权重**（或“**秩**”）处理等价对（H,J）

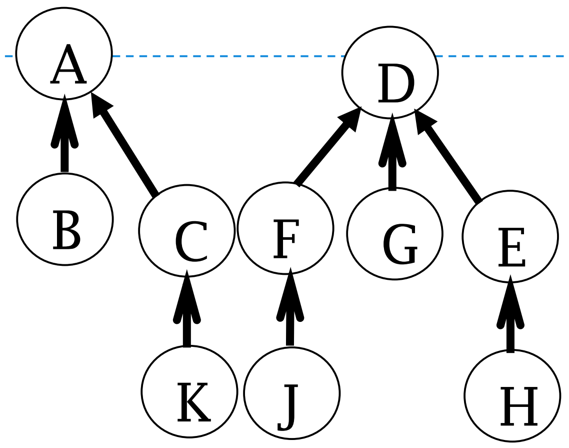
依据加权合并规则，以F为根的树结点个数少，故将F指向D

	0	0	2		4	4	4	5	6
A	B	C	K	D	E	F	G	H	J
0	1	2	3	4	5	6	7	8	9

(H,J)



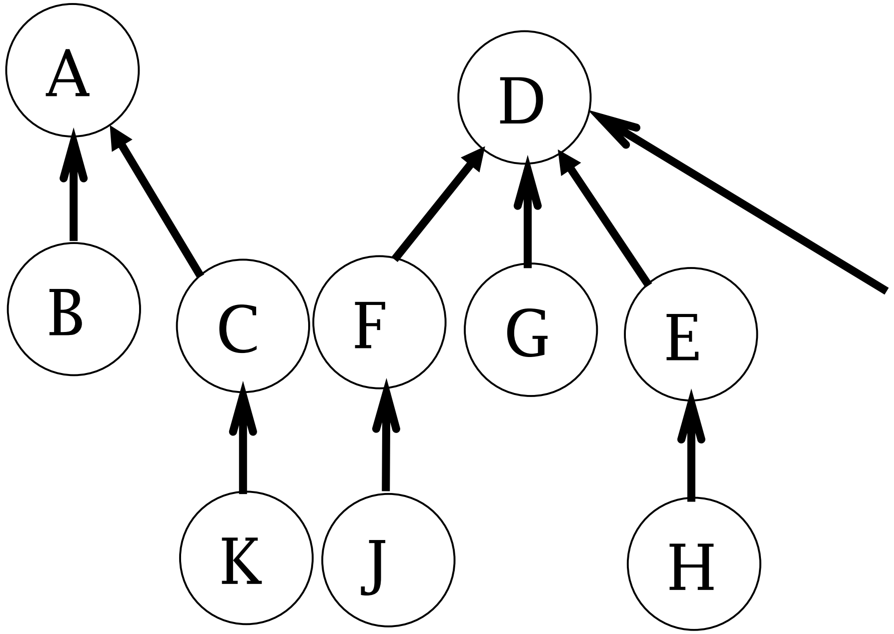
路径压缩



- 查找X
  - 设X最终到达**根R**
  - 顺着**由X到R的路径**把每个结点的**父指针域**均设置为**直接指向R**
- **产生极浅树**

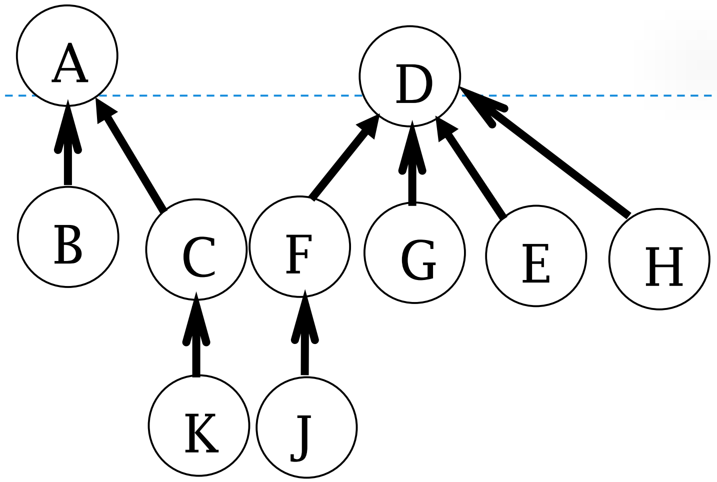
UNION/FIND算法示例(3)

使用路径压缩规则处理Find(H)



	0	0	2		4	4	4	~~5~~ 4	6
A	B	C	K	D	E	F	G	H	J
0	1	2	3	4	5	6	7	8	9

路径压缩




```
template <class T>
ParTreeNode<T>*
ParTree<T>::FindPC(ParTreeNode<T>* node) const
{
  if (node->getParent() == NULL)
    return node;
  node->setParent(FindPC(node->getParent()));
  return node->getParent();
}
```

Ackermann 函数

$$A(m,n)=\begin{cases} n+1 & \text{若m=0} \\ A(m-1,1) & \text{若m>0且n=0} \\ A(m-1,A(m,n-1)) & \text{若m>0且n>0} \end{cases}$$

$A(m, n)$ 的值

$m \backslash n$	0	1	2	3	4	n
0	1	2	3	4	5	$n+1$
1	2	3	4	5	6	$n+2$
2	3	5	7	9	11	$2\cdot(n+3)-3$
3	5	13	29	61	125	$2^{(n+3)}-3$
4	13	65533	$2^{65536}-3$	$A(3, 2^{65536}-3)$	$A(3, A(4, 3))$	$\underbrace{2^{2^{\cdot^{\cdot^{2}}}}}-3$ (n+3个数字2)
5	65533	$A(4, 65533)$	$A(4, A(5, 1))$	$A(4, A(5, 2))$	$A(4, A(5, 3))$	
6	$A(5, 1)$	$A(5, A(5, 1))$	$A(5, A(6, 1))$	$A(5, A(6, 2))$	$A(5, A(6, 3))$	

路径压缩使Find开销接近于常数

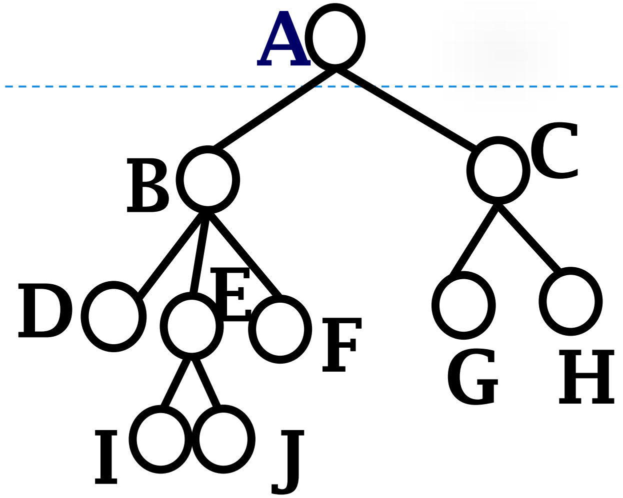
- 权重 + 路径压缩
- 对n个结点进行**n次Find操作的开销为O(nα(n))**，约为Θ($n\log^* n$)
  - α(n)是单变量反Ackermann函数，找最大的整数m使得Ackermann(m,m)≤n
  - $\log^* n$ 是在 n = logn ≤ 1 之前要进行的对 n 取对数操作的次数
  - $\log^* 65536 = 4$（4次log操作）
- Find至多需要一系列**n个Find操作**的开销非常**接近于Θ(n)**
  - 在实际应用中，**α(n)往往小于4**

思考

- 可否使用动态指针方式实现父指针表示法?
- 查阅各种并查集权重和路径压缩优化方法，并讨论各种方法的异同和优劣

第6章　树

- 树的定义和基本术语
- 树的链式存储结构
- **树的顺序存储结构**
- K叉树



树的顺序存储结构

- 带右链的先根次序表示
- 带双标记的先根次序表示
- 带双标记的层次次序表示
- 带度数的后根次序表示

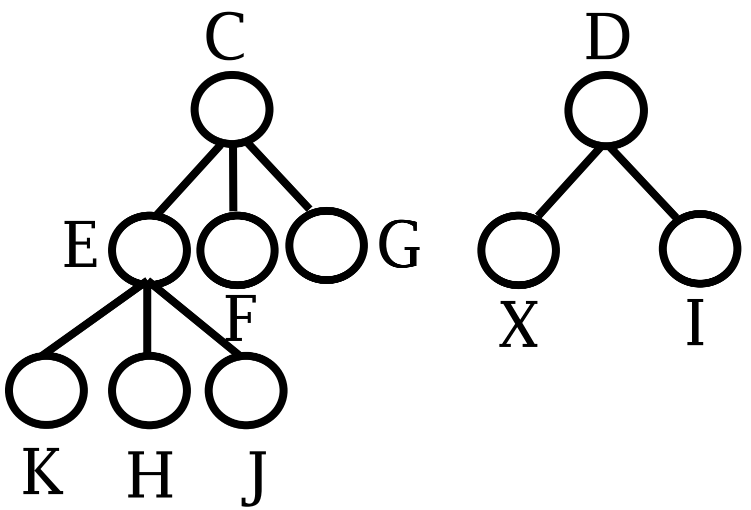
带右链的先根次序表示

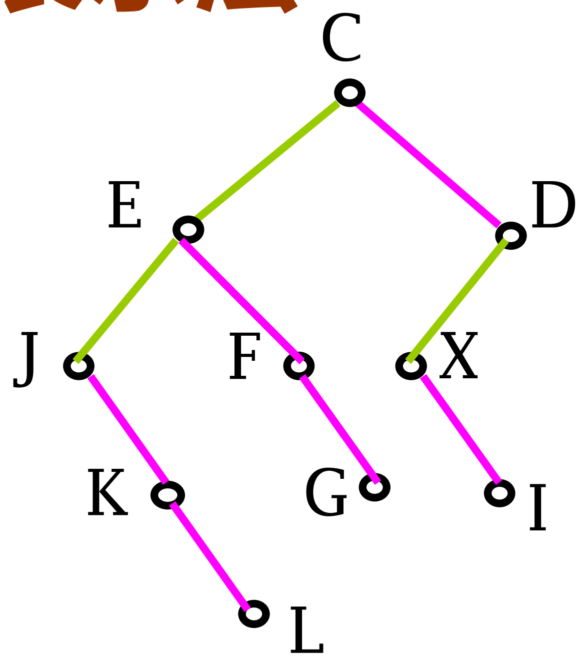
- 结点按先根次序顺序连续存储

ltag	info	rlink

- info：结点的数据
- rlink：右指针
  - 指向结点的下一个兄弟、即对应的二叉树中结点的右子结点
- ltag：标记
  - 树结点没有子结点，即二叉树结点没有左子结点，ltag为 1
  - 否则为0

带右链的先根次序表示法

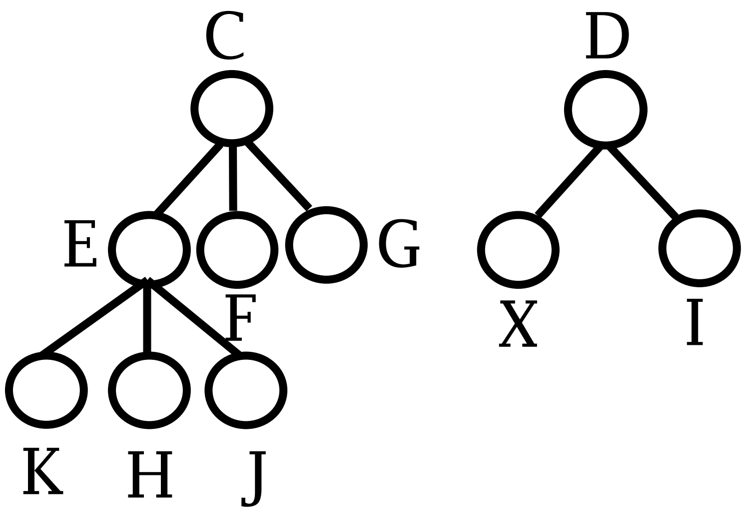


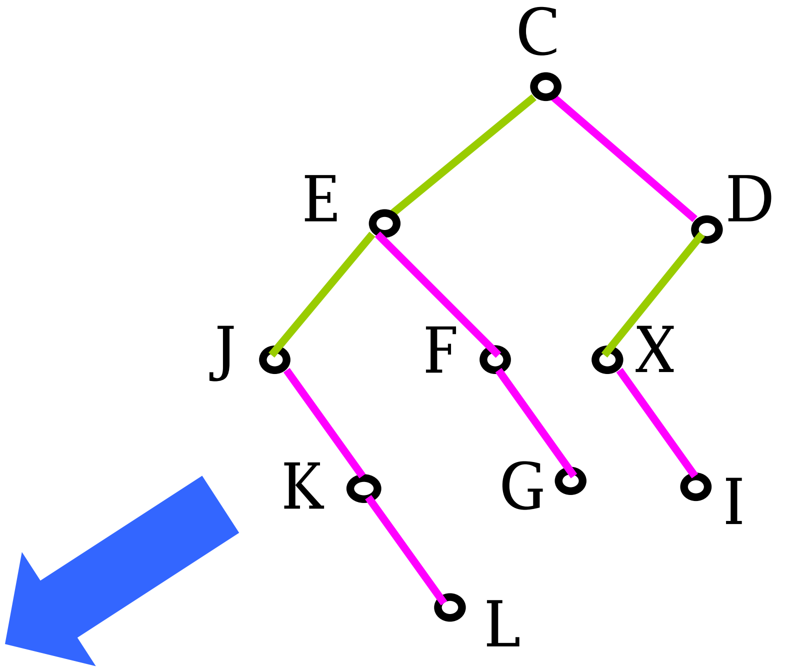


下标	0	1	2	3	4	5	6	7	8	9
rlink	7	5	3	4	-1	6	-1	-1	9	-1
info	C	E	J	K	L	F	G	D	X	I
ltag	0	0	1	1	1	1	1	0	1	1

从先根rlink-ltag到树

下标	0	1	2	3	4	5	6	7	8	9
rlink	7	5	3	4	-1	6	-1	-1	9	-1
info	C	E	J	K	L	F	G	D	X	I
ltag	0	0	1	1	1	1	1	0	1	1





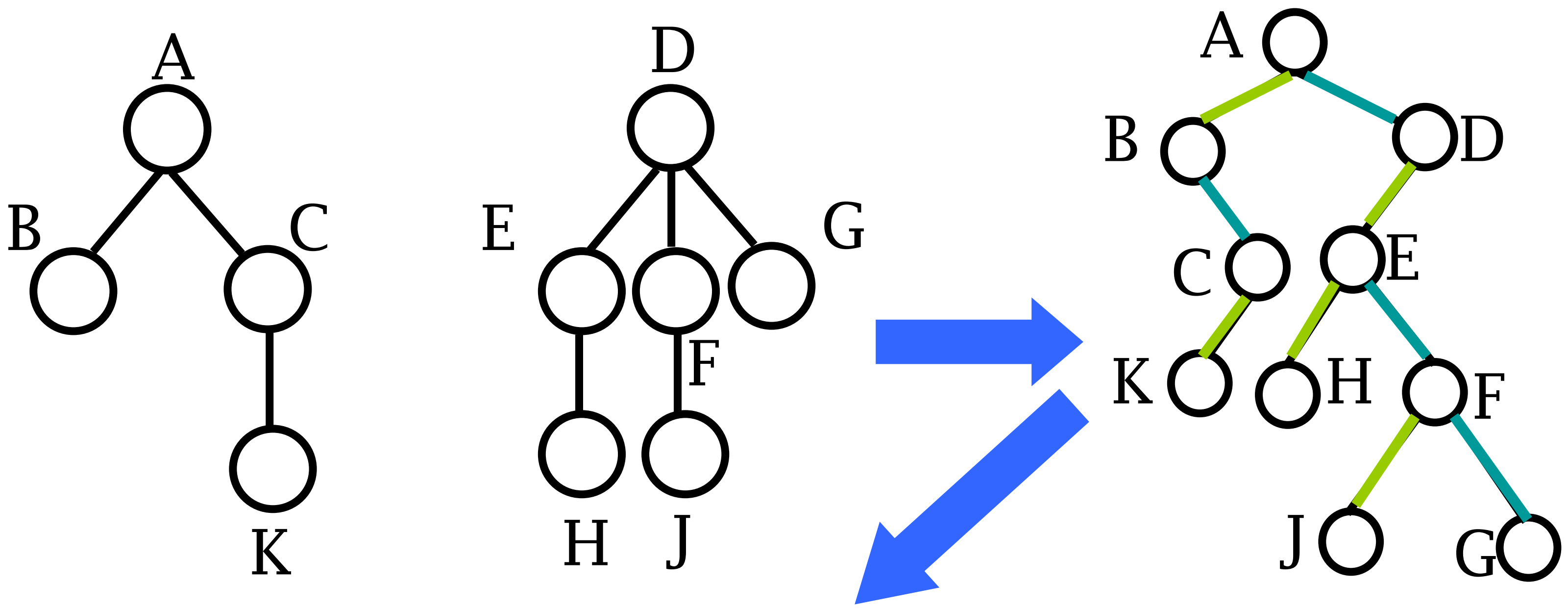
带双标记的先根次序表示

- 带右链的先根次序表示”中**rlink也有冗余**，可以把rlink指针替换为一个标志位rtag，成为“带双标记的先根次序表示”。其中，每个结点包括结点本身数据，以及两个标志位ltag和rtag，其结点的形式为：

ltag	info	rtag

由结点的先根次序以及**ltag、rtag两个标志位**，就可以确定树“左孩子/右兄弟”链表中结点的llink和rlink值。其中llink的确定与带右链的先根次序表示法相同。

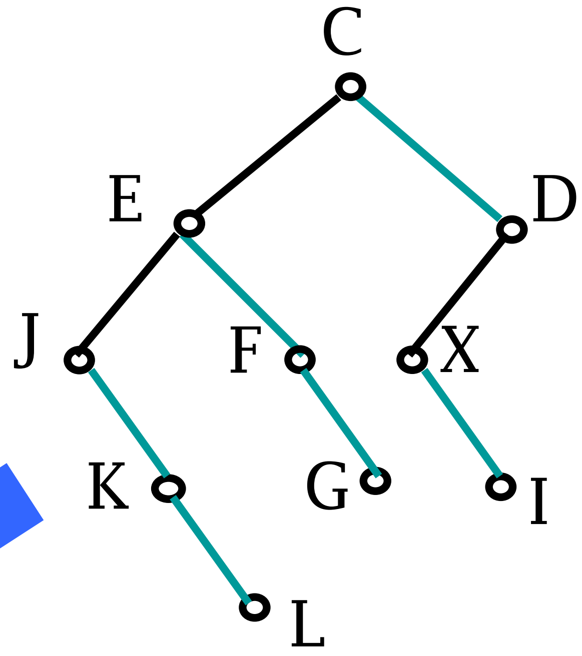
带双标记位的先根次序表示法

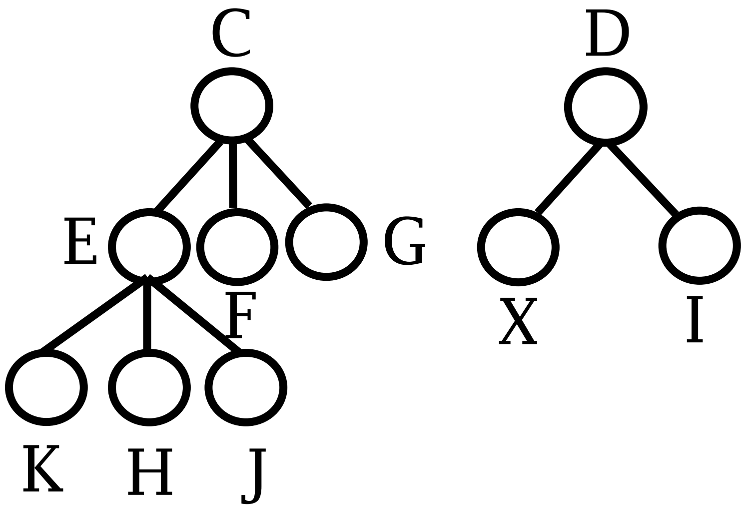


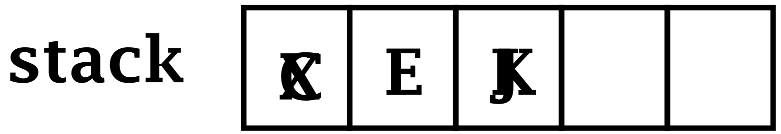
rtag	0	0	1	1	1	0	1	0	1	1
info	A	B	C	K	D	E	H	F	J	G
ltag	0	1	0	1	0	0	1	0	1	1

从rtag-ltag先根序列到树

下标	0	1	2	3	4	5	6	7	8	9
rtag	0	0	0	0	1	0	1	1	0	1
info	C	E	J	K	L	F	G	D	X	I
ltag	0	0	1	1	1	1	1	0	1	1







从双标记的先根次序恢复树

```
template<class T>
class DualTagTreeNode  {                         // 双标记位先根次序树结点类
public:
    T info;                                      // 结点数据信息
    int ltag, rtag;                              // 左、右标记
    DualTagTreeNode();                           // 构造函数
    virtual ~DualTagTreeNode(); };

template <class T>
Tree<T>::Tree(DualTagTreeNode<T> *nodeArray, int count)  {
     // 利用带双标记位的先根次序表示构造左孩子右兄弟表示的树
    using std::stack;                                    // 使用STL中的栈
    stack<TreeNode<T>* > aStack;
    TreeNode<T> *pointer = new TreeNode<T>;   // 准备建立根结点
    root = pointer;
```

6.3 树的顺序存储结构

```
for (int i = 0; i < count-1; i++) {                    // 处理一个结点
  pointer->setValue(nodeArray[i].info);                // 结点赋值
  if (nodeArray[i].rtag == 0)                          // 若右标记为0则将结点压栈
     aStack.push(pointer);
  else pointer->setSibling(NULL);                      // 右标记为1，则右兄弟指针为空
  TreeNode<T> *temppointer = new TreeNode<T>;          // 预先准备下一个
  if (nodeArray[i].ltag == 0)                          // 左标记为0，则设置孩子结点
     pointer->setChild(temppointer);
  else {                                               // 若左标记为1
     pointer->setChild(NULL);                          // 孩子指针设为空
     pointer = aStack.top();                           // 取栈顶元素
     aStack.pop();
     pointer->setSibling(temppointer); }               // 为栈顶设置一个兄弟结点
  pointer = temppointer; }
pointer->setValue(nodeArray[count-1].info); // 处理最后一个结点
pointer->setChild(NULL);  pointer->setSibling(NULL); }
```

带双标记的层次次序表示法

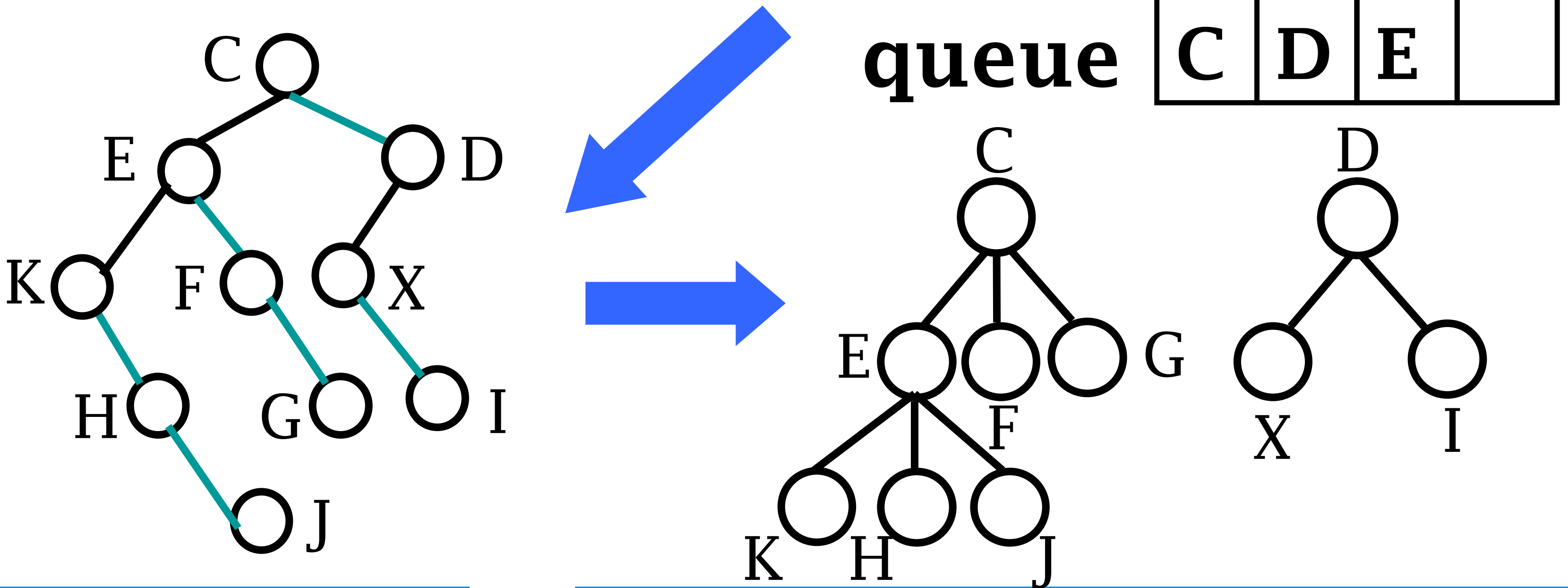
- 结点按 层次次序顺序 存储在连续存储单元

ltag	info	rtag

- info是结点的数据
- ltag是一个一位的左标记，当结点没有子结点，即对应的二叉树中结点没有左子结点时， ltag为1，否则为0
- rtag是一个一位的右标记，当结点没有下一个兄弟，即对应的二叉树中结点没有右子结点时， rtag为1，否则为0

带双标记的层次次序转换为树

ltag	0	0	0	1	1	1	1	1	1	1
info	C	D	E	F	G	X	I	K	H	J
rtag	0	1	0	0	1	0	1	0	0	1



带双标记位的层次次序构造

```
template <class T>
Tree<T>::Tree(DualTagWidthTreeNode<T>* nodeArray, int
count)  {
  using std::queue;                                    // 使用STL队列
  queue<TreeNode<T>*> aQueue;
  TreeNode<T>* pointer=new TreeNode<T>; // 建立根
  root=pointer;
  for(int i=0;i<count-1;i++) {                         // 处理每个结点
    pointer->setValue(nodeArray[i].info);
    if(nodeArray[i].ltag==0) aQueue.push(pointer);   // 入队
       else pointer->setChild(NULL);                   // 左孩子设为空
    TreeNode<T>* temppointer=new TreeNode<T>;
```

```
    if(nodeArray[i].rtag == 0)
      pointer->setSibling(temppointer);
    else {
        pointer->setSibling(NULL);     // 右兄弟设为空
        pointer=aQueue.front();        // 取队列首结点指针
        aQueue.pop();                  // 队首元素出队列
        pointer->setChild(temppointer);
    }
    pointer=temppointer;
  }
  pointer->setValue(nodeArray[count-1].info); // 最后一个结点
  pointer->setChild(NULL);  pointer->setSibling(NULL);
}
```

带度数的后根次序表示

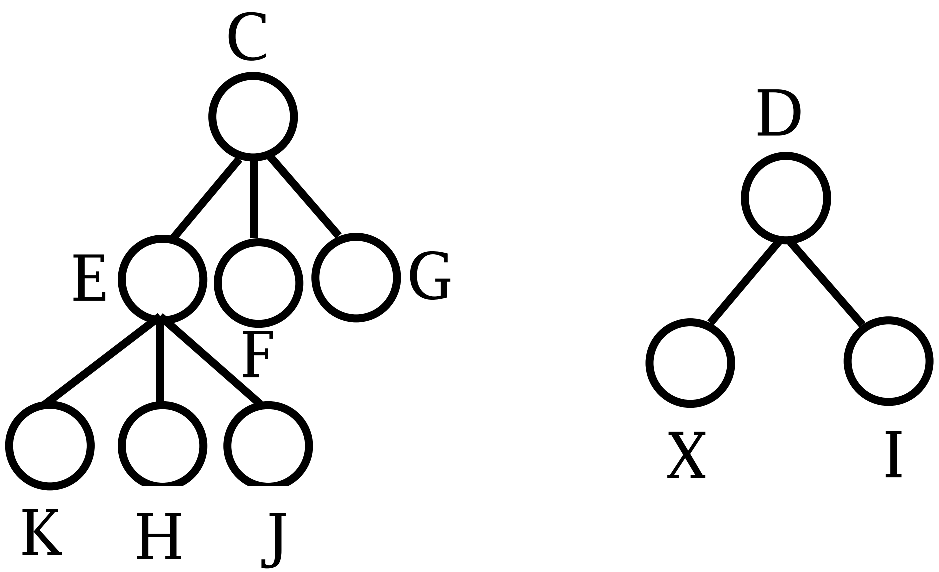
- 在带度数的后根次序表示中，结点按后根次序顺序存储在一片连续的存储单元中，结点的形式为

info	degree

- 其中info是结点的数据，degree是结点的度数

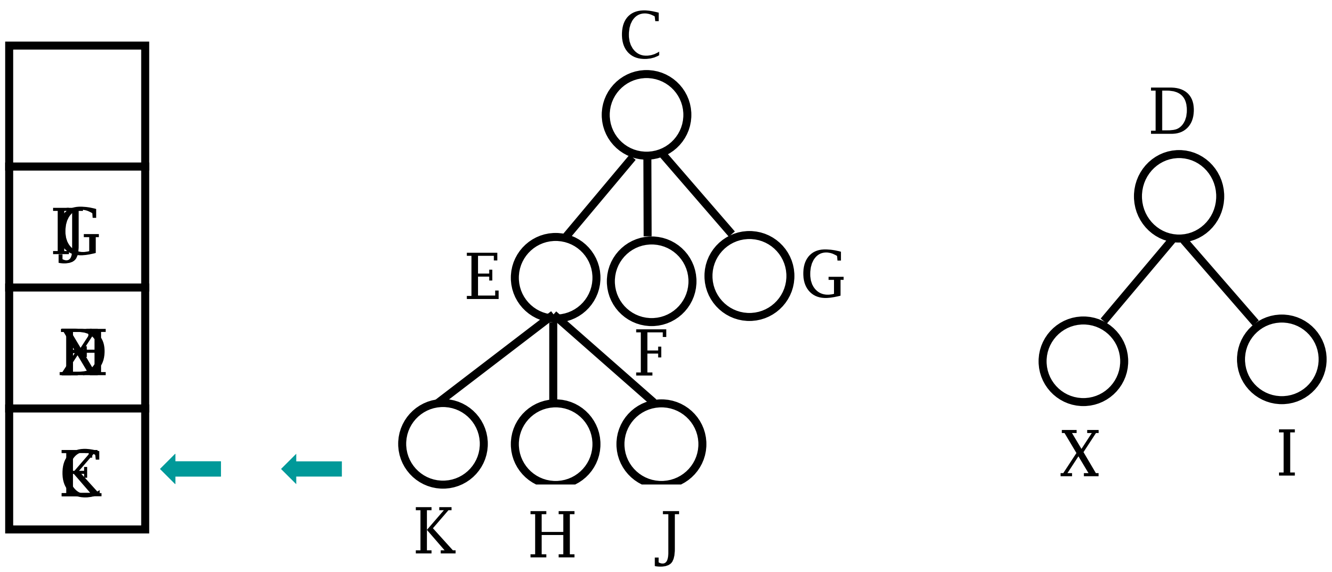
带度数的后根次序表示法

degree	0	0	0	3	0	0	3	0	0	2
info	K	H	J	E	F	G	C	X	I	D



带度数的后根次序变成树

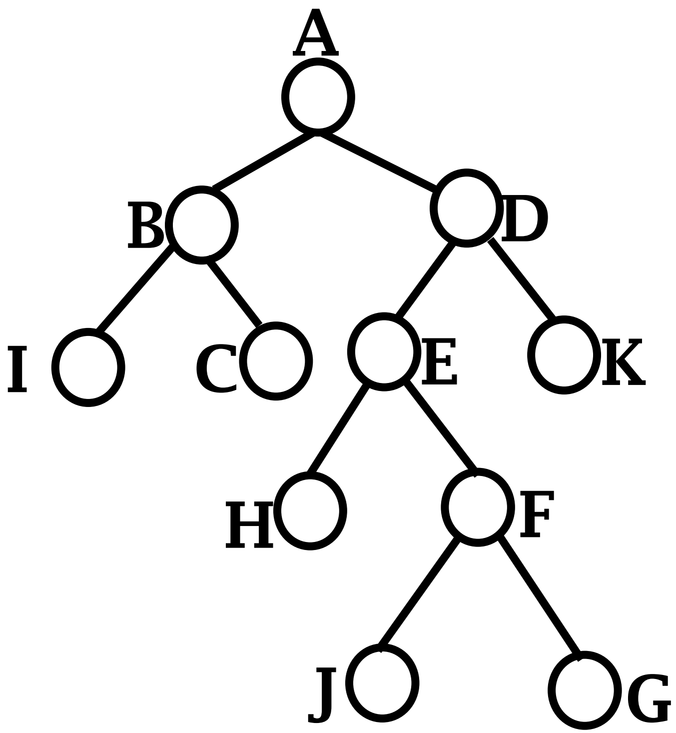
degree	0	0	0	3	0	0	3	0	0	2
info	K	H	J	E	F	G	C	X	I	D



6.3 树的顺序存储结构

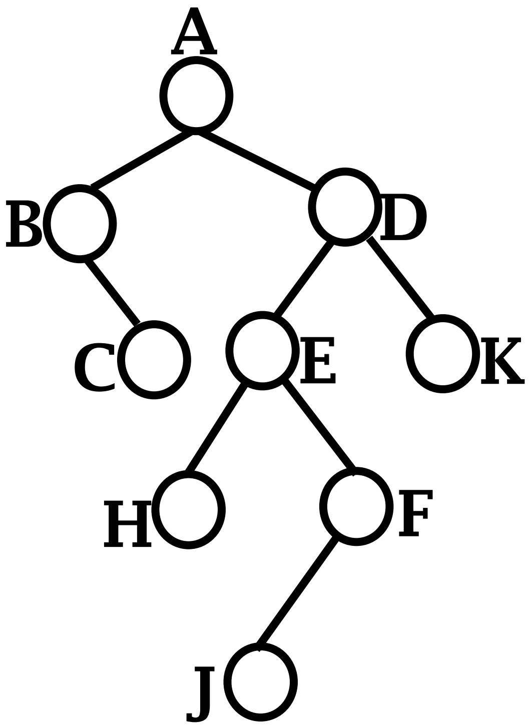
- 带标记的满二叉树前序序列

  A' B' I C D' E' H F' J G K



- 带标记的伪满二叉树前序序列

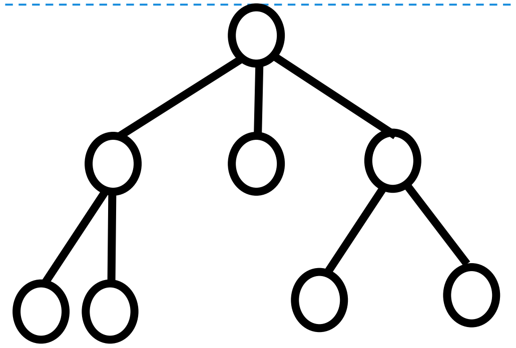
  A' B' / C D' E' H F' J / K



思考：森林的顺序存储

- 信息冗余问题
- 树的其他顺序存储
  - 带度数的先根次序?
  - 带度数的层次次序?
- 二叉树的顺序存储?
  - 二叉树与森林对应，但语义不同
    - 带右链的二叉树前序
    - 带左链的二叉树层次次序

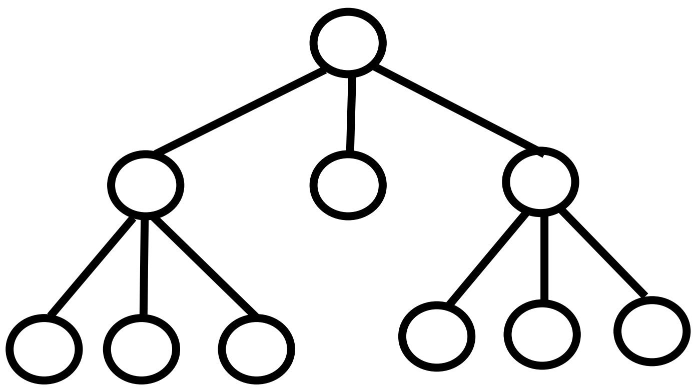
K 叉树定义



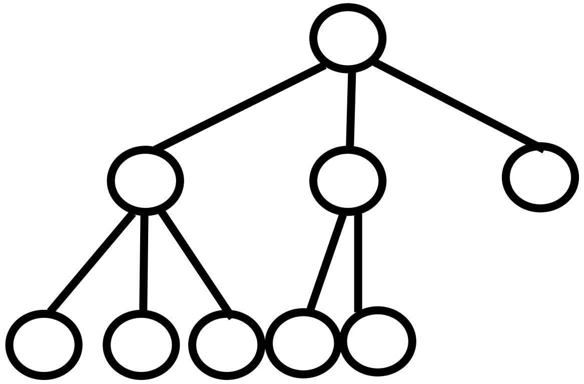
- K 叉树 T 是具有下列性质的有限结点集：
  - (a) 集合可以为空；
  - (b) 非空集合是由**一个根结点** root 及 **K 棵互不相交的 K 叉树**构成。
- 其余结点被划分成 T_0， T_1， …， T_{K-1} （K ≥ 1） 个子集，每个子集都是 K 叉树，使得 $T = \{R, \ T_0, \ T_1, \ \ldots, \ T_{K-1}\}$。
- K 叉树的各分支结点都有 K 个子结点

满 K 叉树和完全 K 叉树

- K 叉树 (K-ary Tree) 的结点有 K 个子结点
- **二叉树**的许多性质可以**推广**到 K 叉树
  - 满K叉树和完全K叉树与满二叉树和完全二叉树是类似的
  - 也可以把完全K叉树存储在一个数组中



满3叉树



完全3叉树

例题：Dragon Balls (HDU 3635)

（来源： http://acm.hdu.edu.cn/showproblem.php?pid=3635）

- 描述：在这个神奇的世界上，存在着n个城市和n个龙珠。在世界诞生之初，神威龙主将第i个龙珠放在第i个城市里。随着时间的推移，其中的一些龙珠被转移到了其他城市。而且转移的时候，某一个城市的所有龙珠都会一起被移动到另一个城市。
- 输入：第一行是N，M，N是城市的个数，也是龙珠的个数

  接着输入M行指令，这M行指令都遵循着两种格式：

  T A B：表示A球所在城市中的所有龙珠都被移动到B球的城市

  Q A：查询标号为A的龙珠目前所在城市X，X城中目前的龙珠

数量Y，A被转移的次数Z
- 输出： 对于输入中的每一条Q指令，输出对应的X Y Z

```
3 4
T 1 2
Q 1
T 1 3
Q 1
```

```
2 2 1
3 3 2
```

例题：Cube stacking

- 有N(N<=30,000)堆方块，开始每堆都是一个方块。方块编号1 ~ N. 有两种操作：

 M x y ： 表示把方块x所在的堆，拿起来叠放到y所在的堆上。

 C x : 问方块x下面有多少个方块。

- 操作最多有 P (P<=100,000)次。对每次C操作，输出结果。

并查集解题的思路

- 如何构造集合
- 需要维护哪些信息
- 信息在什么时候更新

一些相关的计数问题

- {1, 2, …, n} 顺序进栈、出栈所得序列数目
- n 个结点的不同形态二叉树的数目 b_n
- 前序序列为 $\{a_1, a_2, \ldots\ldots, a_n\}$ 的二叉树数目
- n 个结点的标号 BST 树数目
- n+1 个结点的不同形态树（不是森林）的数目 t_{n+1}

数据结构与算法

感谢倾听

国家精品课“数据结构与算法”

http://jpk.pku.edu.cn/course/sjjg/

https://www.icourse163.org/course/PKU-1002534001

张铭，王腾蛟，赵海燕

高等教育出版社，2008. 6。“十二五”国家级规划教材