Technology Tarena 达内科技

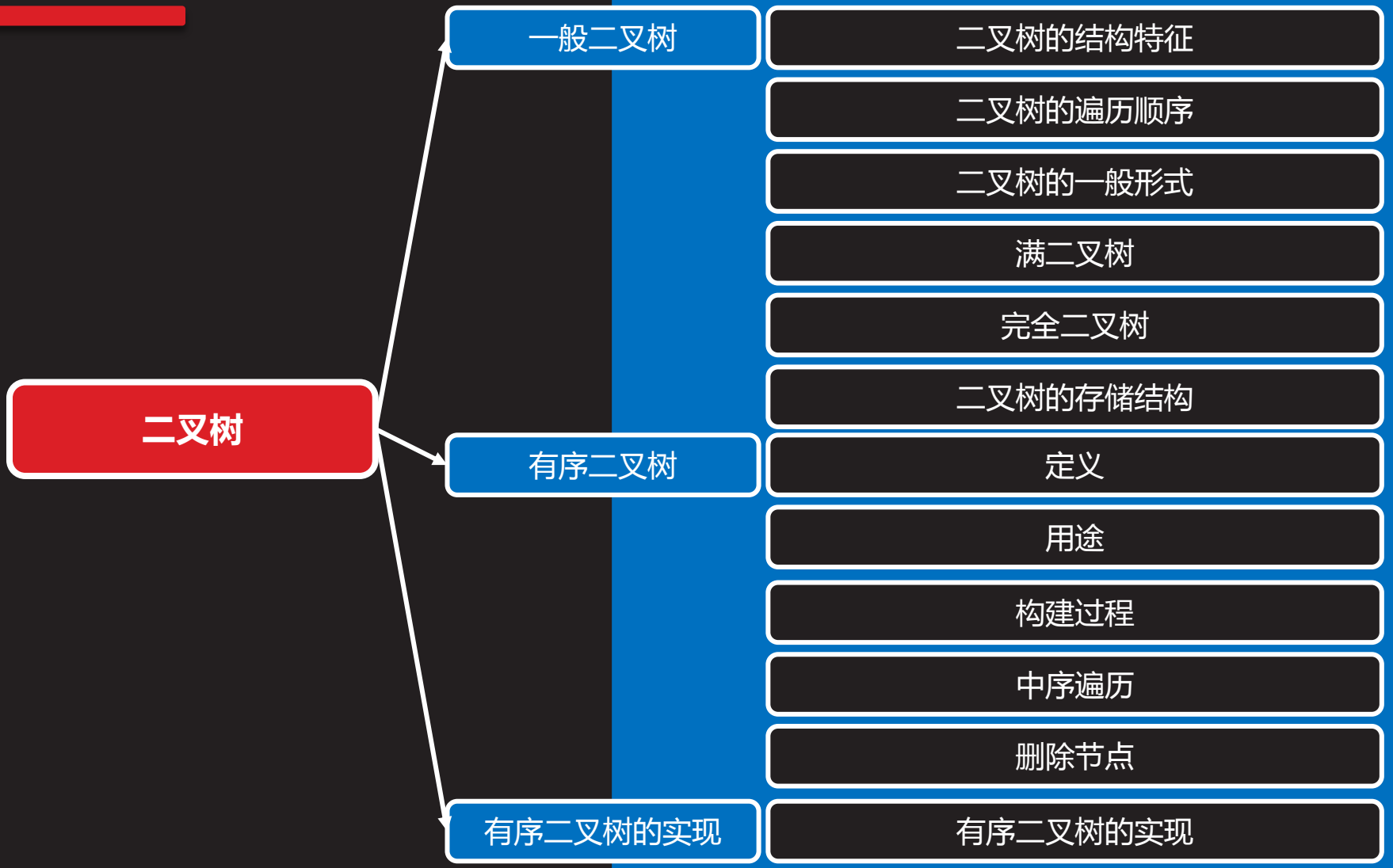
数据结构与算法

DATASTRUCTURE DAY03

内容

上午	09:00 ~ 09:30	作业讲解和回顾
	09:30 ~ 10:20	二叉树
	10:30 ~ 11:20	
	11:30 ~ 12:20	
下午	14:00 ~ 14:50	
	15:00 ~ 15:50	
	16:00 ~ 16:50	算法概述
	17:00 ~ 17:30	总结和答疑

二叉树

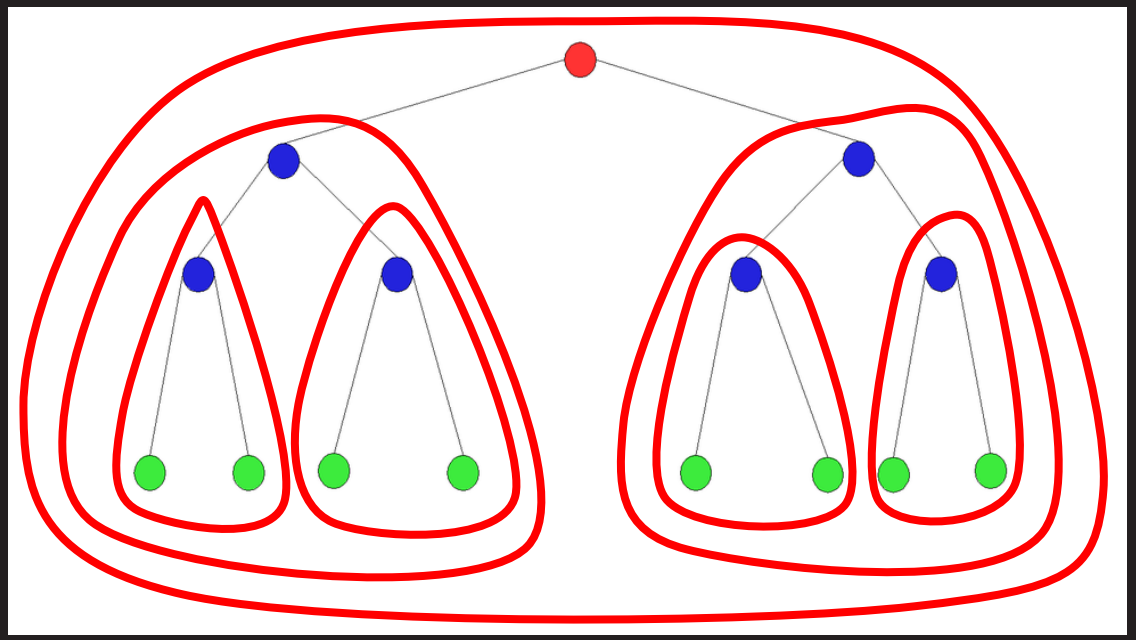

二叉树
一般二叉树
二叉树的结构特征
二叉树的遍历顺序
二叉树的一般形式
满二叉树
完全二叉树
二叉树的存储结构
有序二叉树
定义
用途
构建过程
中序遍历
删除节点
有序二叉树的实现
有序二叉树的实现

Technology Tarena 达内科技

一般二叉树

二叉树的结构特征

- 二叉树是树形结构的最简模型，每个节点最多有两个子节点，理论上讲，任何树的问题总可以被转化为二叉树问题
- 二叉树中除去根节点以外的每个节点有且仅有一个父节点，整棵树只有一个根，枝节点有父亦有子，叶节点有父无子
- 二叉树具有递归嵌套式的空间结构特征，因此采用递归的方法处理二叉树问题，常常可以使算法变得更加简洁

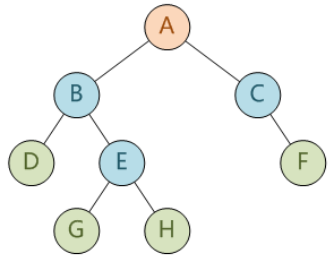


知识讲解

二叉树的遍历顺序

知识讲解

- 前序遍历
  - 对于从树根开始的每一棵子树，先处理根节点中的数据，然后处理它的左子树，最后处理它的右子树，简单表示为DLR
- 中序遍历
  - 对于从树根开始的每一棵子树，先处理它的左子树，然后处理根节点中的数据，最后处理它的右子树，简单表示为LDR
- 后序遍历
  - 对于从树根开始的每一棵子树，先处理它的左子树，然后处理它的右子树，最后处理根节点中的数据，简单表示为LRD



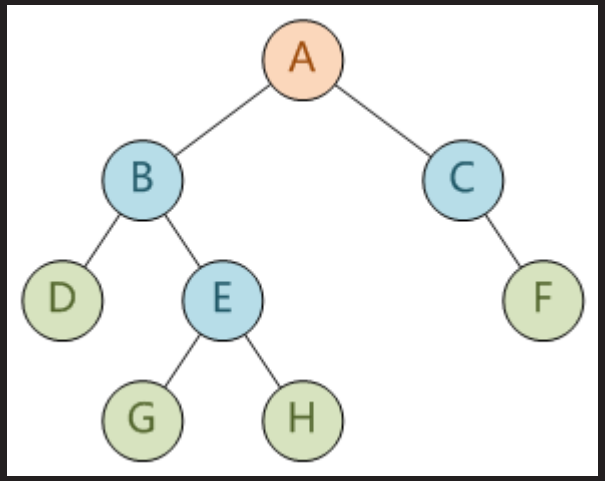
前序遍历：A-B-D-E-G-H-C-F

中序遍历：D-B-G-E-H-A-C-F

后序遍历：D-G-H-E-B-F-C-A

二叉树的一般形式

- 根节点、枝节点和叶节点
  - A是根节点，BCE是枝节点，DFGH是叶节点
- 父节点和子节点
  - A是B的父节点，D是B的子节点
- 左子节点和右子节点
  - D是B的左子节点，E是B的右子节点
- 左子树和右子树
  - BDEGH是A的左子树，
    CF是A的右子树
- 大小和高度(深度)
  - 大小是8，高度是4

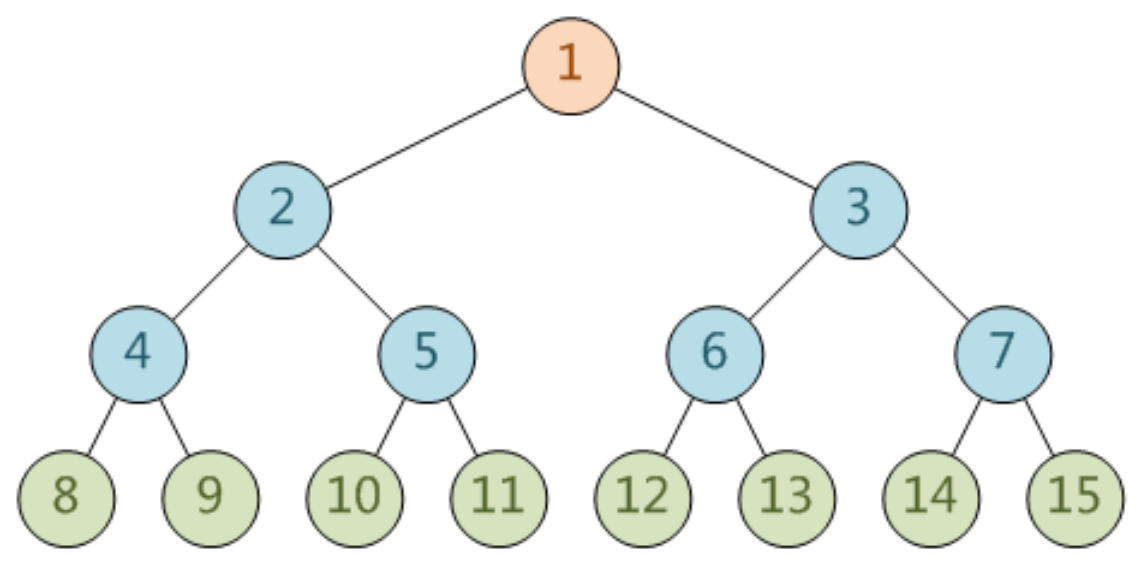


知识讲解

满二叉树

- 满足以下两个条件的二叉树即为满二叉树
  - 每层节点数均达到最大值
  - 所有枝节点均有左右子树

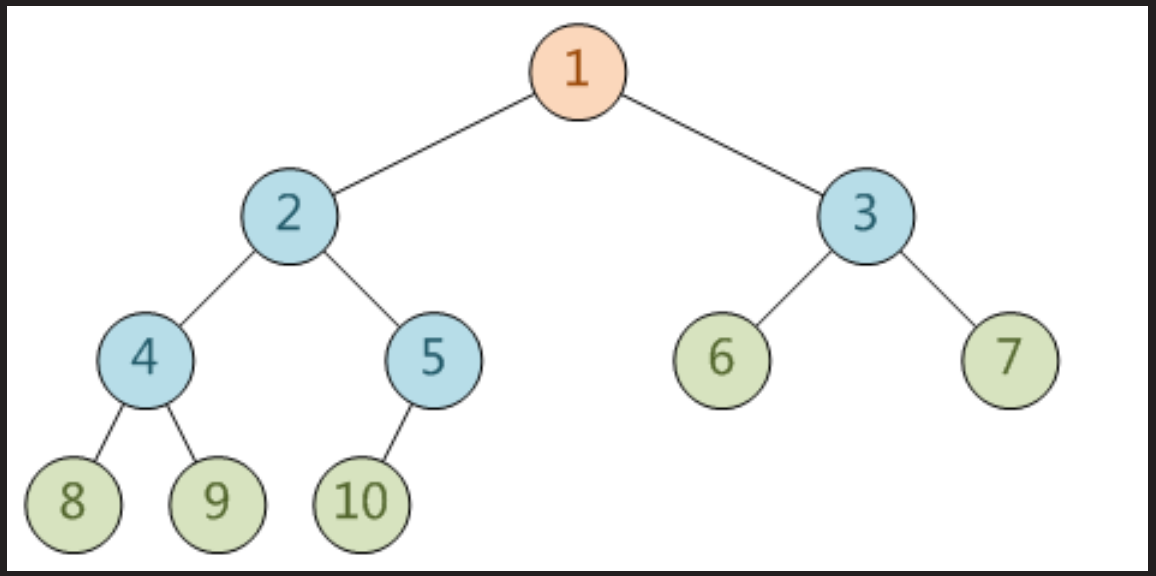
知识讲解



完全二叉树

- 满足以下两个条件的二叉树即为完全二叉树
  - 除最下层外，各层节点数均达到最大值
  - 最下层的节点都连续集中在左边

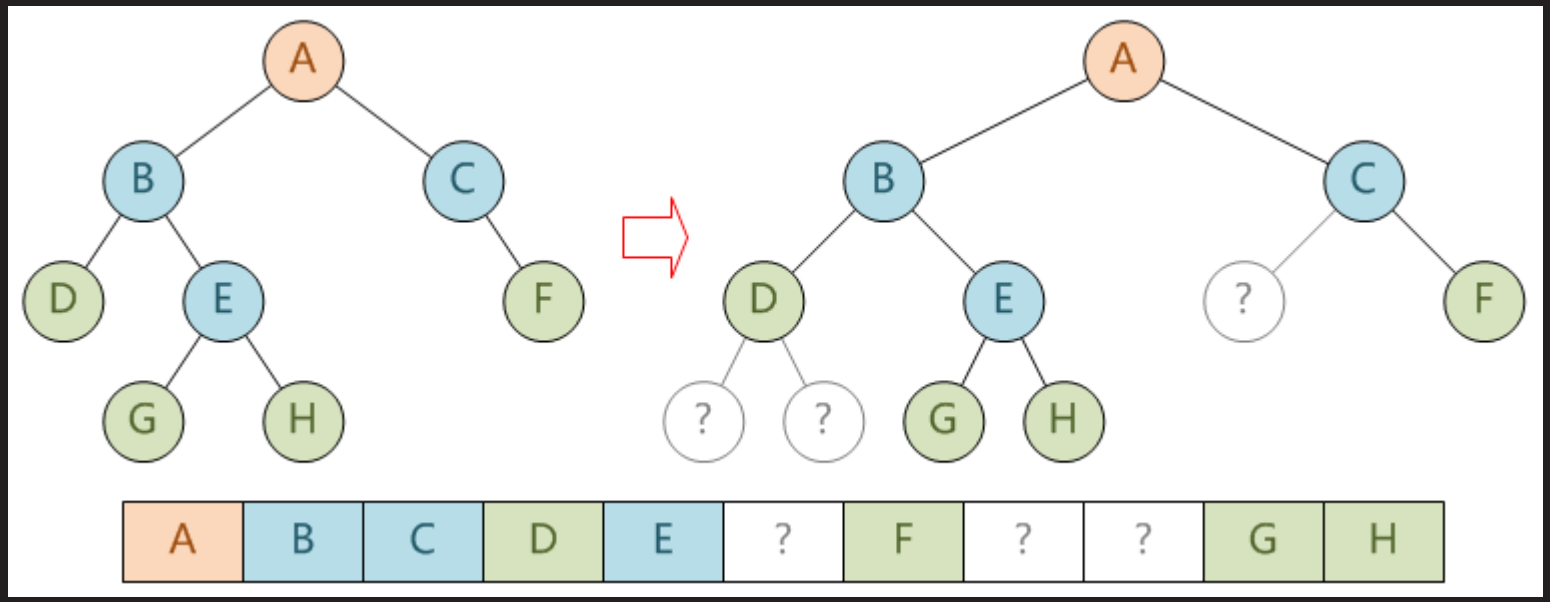
知识讲解



二叉树的存储结构

- 顺序存储
  - 从上到下、从左到右，依次存放
  - 非完全二叉树需用虚节点补成完全二叉树

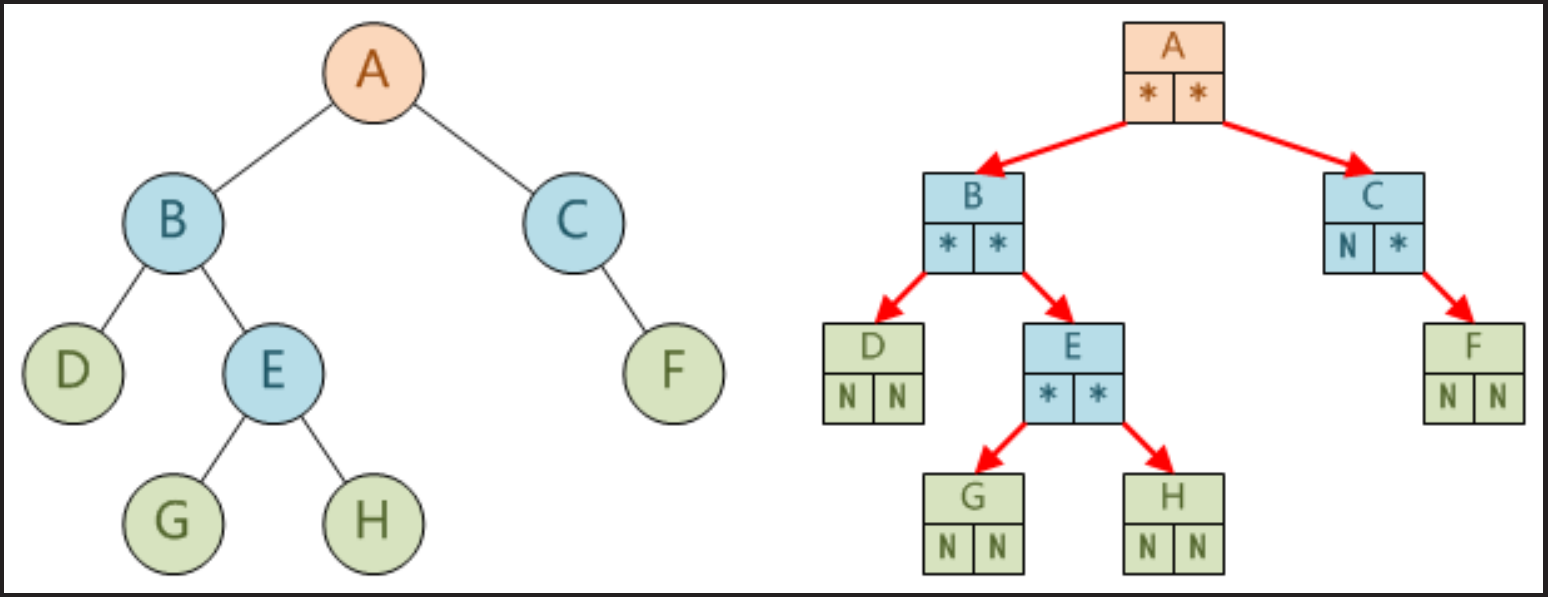
知识讲解



二叉树的存储结构（续1）

- 链式存储
  - 二叉链表：每个节点包括三个域，一个数据域和两个分别指向其左右子节点的指针域

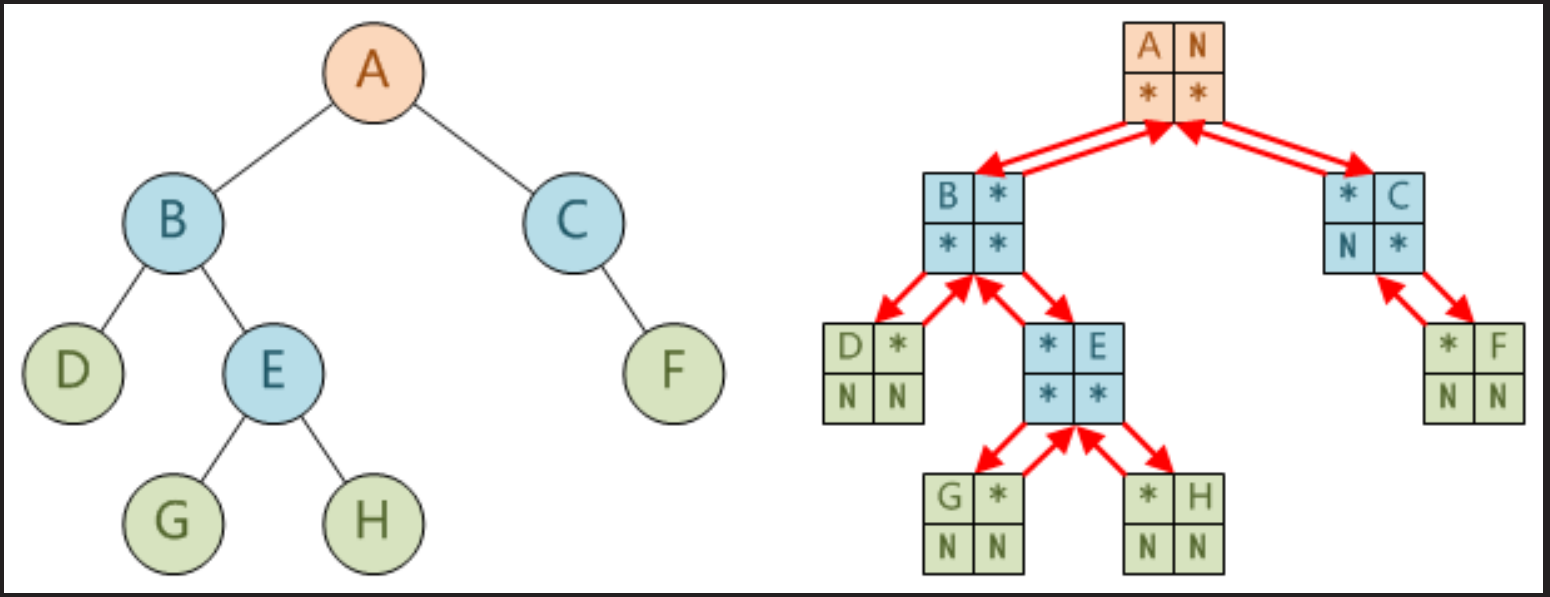
知识讲解



二叉树的存储结构（续2）

- 链式存储
  - 三叉链表：每个节点包括四个域，一个数据域、两个分别指向其左右子节点的指针域和一个指向其父节点的指针域

知识讲解



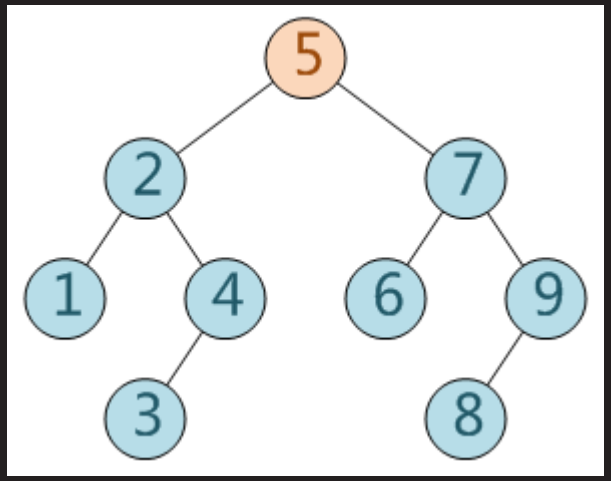
Technology Tarena 达内科技

有序二叉树

定义

- 满足以下三个条件的非空二叉树即为有序二叉树
  - 若左子树非空，则左子树上所有节点的值均小于等于根节点的值
  - 若右子树非空，则右子树上所有节点的值均大于等于根节点的值
  - 左右子树亦分别为有序二叉树

知识讲解



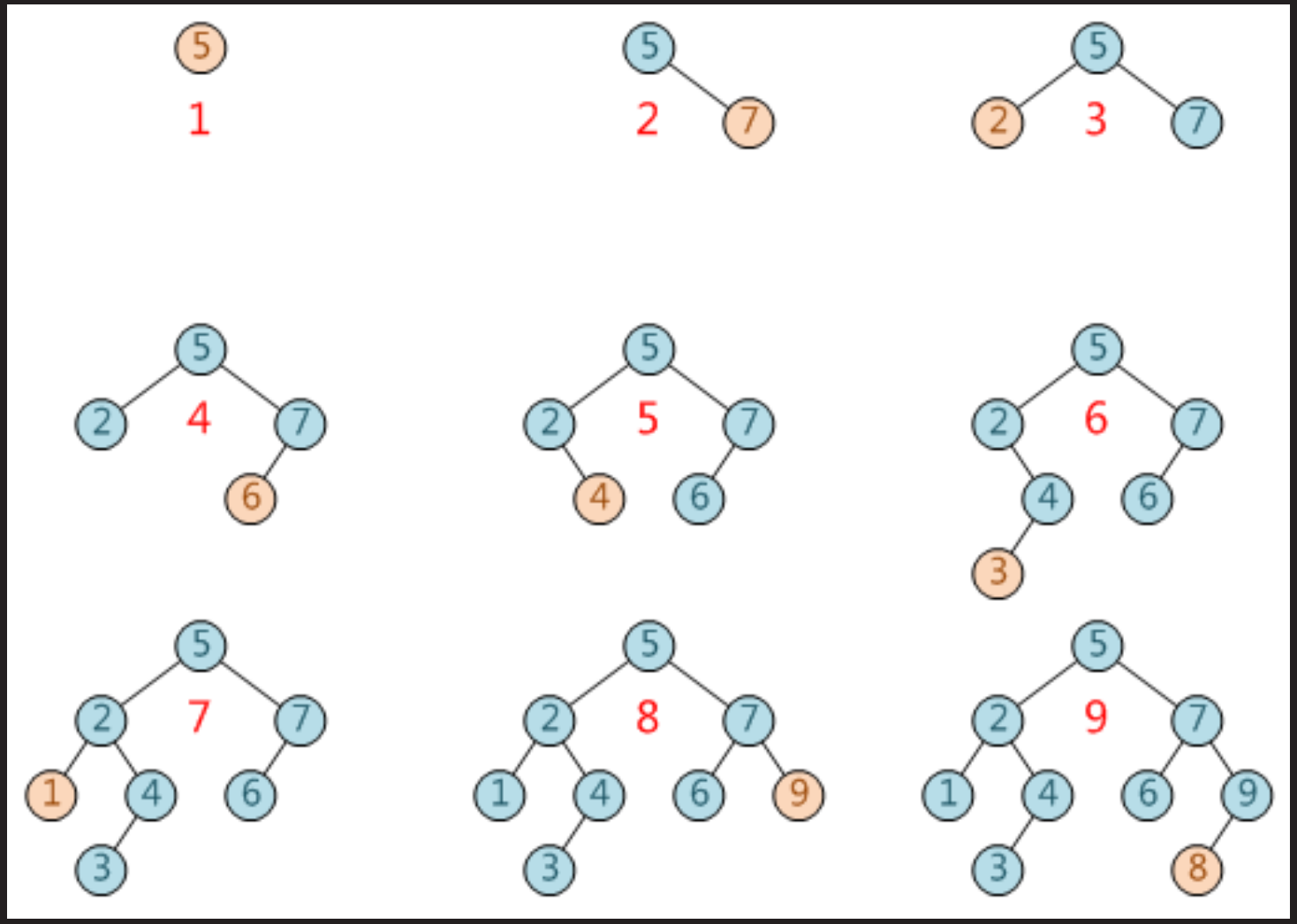
用途

- 排序
  - 无论以何种顺序构建有序二叉树，其中序遍历的结果一定是一个有序序列
- 搜索
  - 若搜索目标与根节点的值相等，则搜索成功，否则用搜索目标和根节点的值比较大小
  - 若搜索目标小于根节点的值，则在根节点的左子树中继续搜索，否则在根节点的右子树中继续搜索
  - 以递归的方式重复以上过程，直到搜索成功，或因子树不存在而宣告失败
  - 基于有序二叉树的搜索，可达到对数级的平均时间复杂度

知识讲解

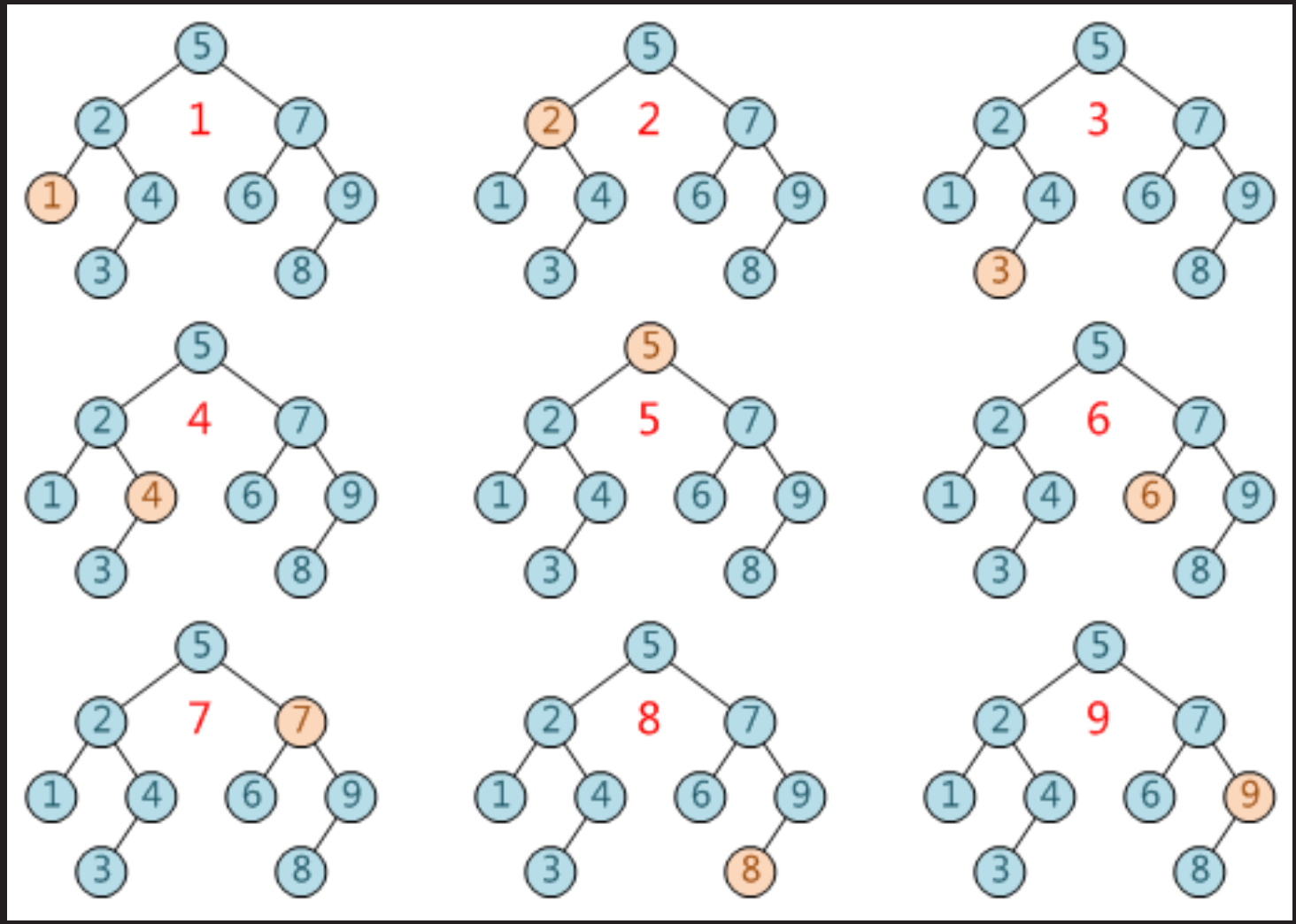
构建过程

知识讲解



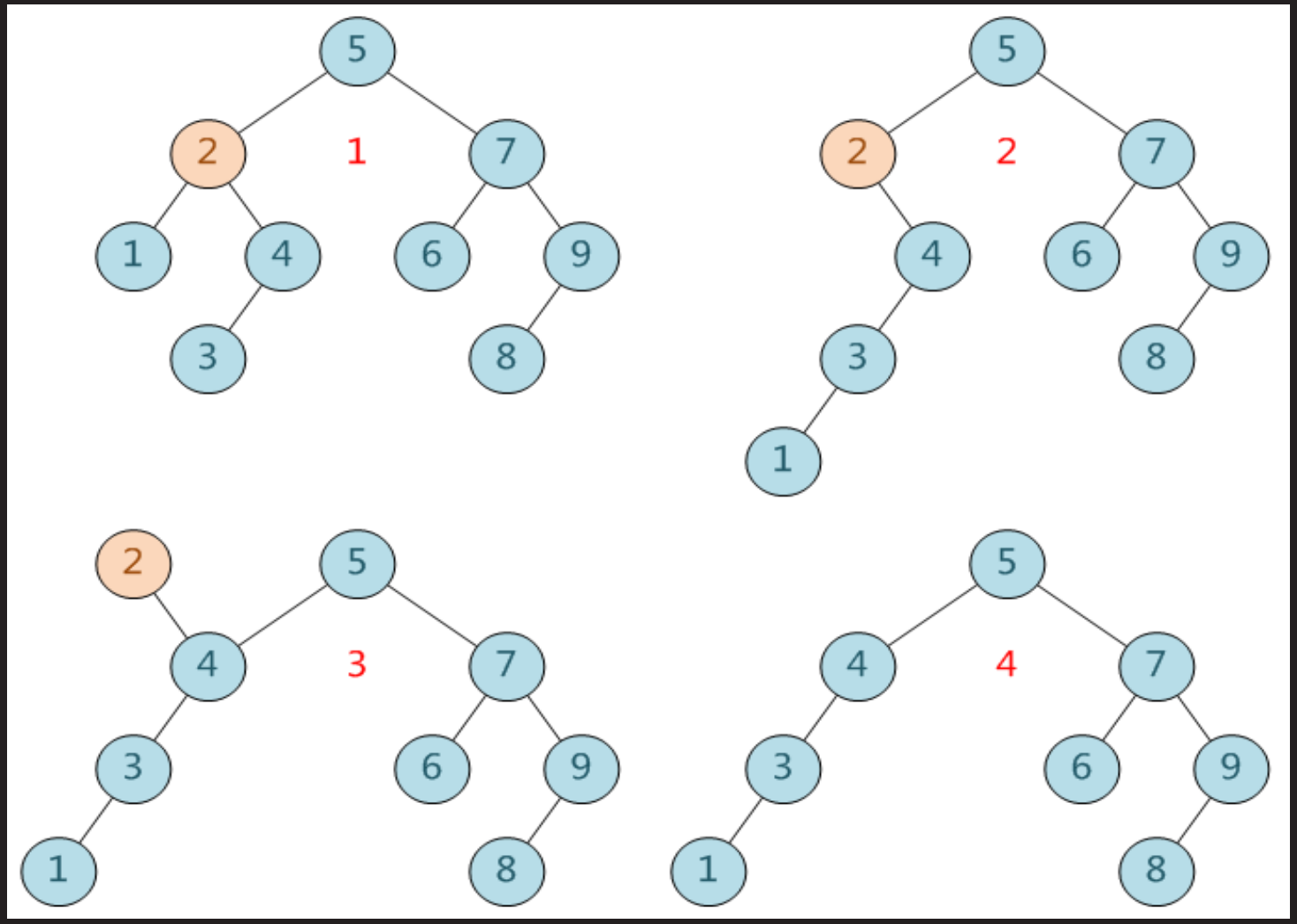
中序遍历

知识讲解



删除节点

知识讲解



Technology Tarena 达内科技

有序二叉树的实现

有序二叉树的实现

- 按照有序规则，插入数据元素
- 删除第一个匹配数据元素
- 删除所有的匹配数据元素
- 清空整棵树
- 在不破坏有序规则的前提下，更新特定数据元素的值
- 判断特定数据元素是否存在
- 中序遍历
- 获取树的大小和高度
- 打印树中全部数据元素

知识讲解

Technology Tarena 达内科技

基于二叉链表的有序二叉树

【参见：TTS COOKBOOK 】

课堂练习

- 基于二叉链表的有序二叉树

算法概述


算法概述
算法的基本概念
算法的基本特征
算法的基本要素
算法的评定标准
算法的思想方法
算法的分类
算法的描述
算法的基本概念
算法的基本特征
算法的基本要素
算法的评定标准
算法的思想方法
算法的分类
算法的描述

Technology Tarena 达内科技

算法的基本概念

算法的基本概念

- 算法是指对解题方案准确而完整的描述，是一系列解决问题的清晰指令，算法代表着用系统的方法描述解决问题的策略机制
- 算法中的指令描述的是一个计算过程，在它的作用下，系统从初始状态和初始输入(也可能没有)开始，经历一系列有限且被明确定义的中间状态，最终产生所要求的输出，并停止于终止状态
- 同一个问题，不同的算法，可能会在时间、空间等方面表现出明显的差异，一个算法的优劣可以用时间复杂度和空间复杂度来衡量

知识讲解

Technology Tarena 达内科技

算法的基本特征

算法的基本特征

- 有穷性
  - 算法必须能在执行有限个步骤后终止
- 确切性
  - 算法的每一个步骤都必须有确切的定义
- 输入项
  - 算法有规范化的输入，以表示运算对象的初始状态
- 输出项
  - 算法至少有一个输出，以反映处理的最终结果
- 可行性
  - 算法中的每个执行步骤都必须能在有限时间内完成

知识讲解

Technology Tarena 达内科技

算法的基本要素

算法的基本要素

- 运算和操作
  - 计算机可以执行的基本操作是以指令的形式描述的，包括如下四类
    - 算术运算：加、减、乘、除、模等
    - 关系运算：大于、小于、等于、不等于等
    - 逻辑运算：与、或、非等运算
    - 数据传输：输入、输出、赋值等
- 流程和控制
  - 算法的功能不仅取决于所选用的操作，还与各操作之间的执行顺序有关

知识讲解

Technology Tarena 达内科技

算法的评定标准

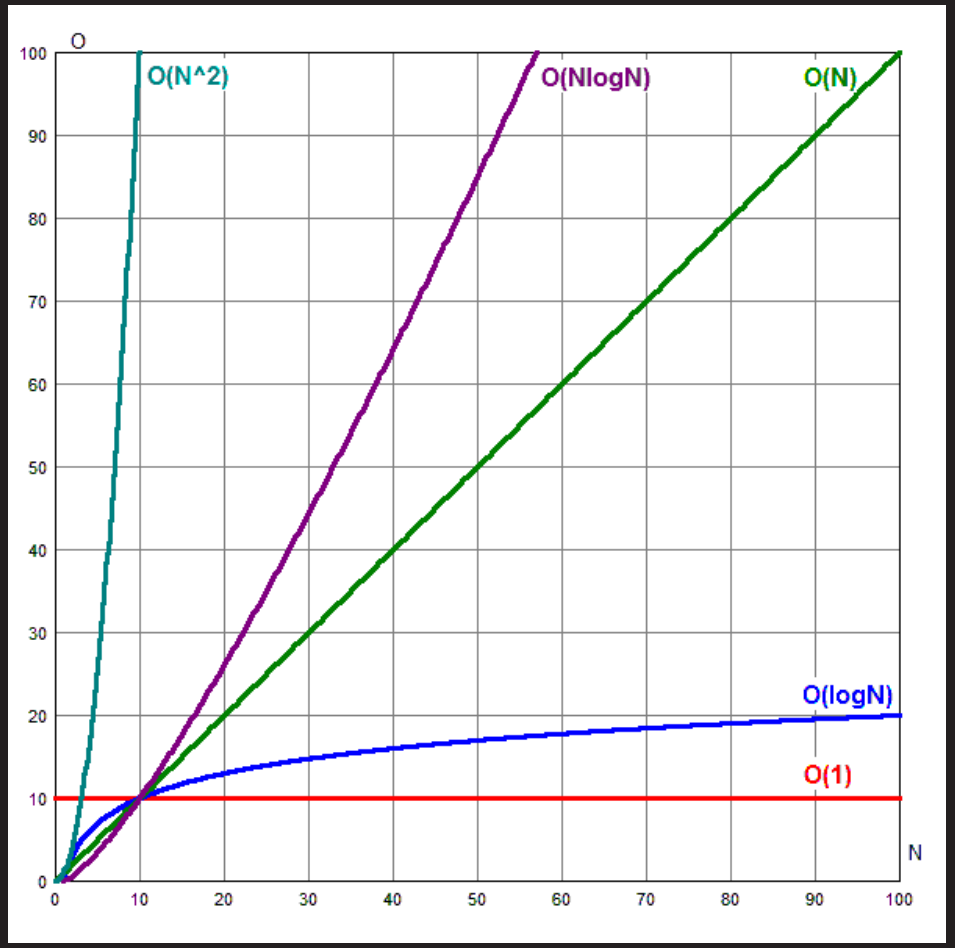
算法的评定标准

- 时间复杂度
  - 算法的时间消耗与问题规模之间的函数关系：T(N)=O(F(N))
- 空间复杂度
  - 算法的空间消耗与问题规模之间的函数关系：S(N)=O(F(N))
- 正确性
  - 执行算法的结果是否满足要求
- 可读性
  - 算法本身可供人们阅读的难易程度
- 健壮性
  - 算法对非正常输入的反应和处理能力，亦称容错性

知识讲解

算法的评定标准（续1）

- 常见时空复杂度
  - 常数级复杂度
    - O(1)
  - 对数级复杂度
    - O(logN)
  - 线性级复杂度
    - O(N)
  - 线性对数级复杂度
    - O(NlogN)
  - 平方级复杂度
    - $O(N^2)$



- 复杂度曲线越平越好，越陡越差，常数级复杂度最为理想

知识讲解

Technology Tarena 达内科技

算法的思想方法

算法的思想方法

- 递推法
  - 通过计算前面的一些项来得出序列中指定项的值
  - 其思想是把一个复杂而庞大的计算过程转化为简单过程的多次重复，充分发挥计算机速度快且不知疲倦的特点
- 递归法
  - 把大问题转化为与原问题相似的小问题来求解
  - 递归的神奇之处在于用有限的语句来定义无限的对象集合
  - 递归需要有终止条件、递归前进段和递归返回段，若终止条件不满足则递归前进，否则递归返回
- 穷举法
  - 对于要解决的问题，列举出它所有的可能性，逐个判断其中哪些符合问题所要求满足的条件，从而得到问题的解

知识讲解

算法的思想方法（续1）

- 贪心算法
  - 自顶向下，以迭代的方式做出相继的贪心选择，每做一次贪心选择，就将所求问题简化为一个规模更小的子问题，通过每一步贪心选择，可得到相应子问题的一个最优解，虽然每一步都能够保证局部解最优，但最终得到的全局解未必最优
- 分治法
  - 把一个复杂问题分成两个或更多相同或相似的子问题，再把子问题分成更小的子问题……，直到最后的子问题可以简单地直接求解，原问题的解即子问题解的合并
- 动态规划法
  - 将原问题分解为相似的子问题，在求解的过程中通过子问题的解求出原问题的解

知识讲解

算法的思想方法（续2）

- 迭代法
  - 按照一定的规则，不断地用旧值推算出新值，直到满足某种特定条件为止
  - 利用计算机速度快、擅于简单重复的特点，让计算机重复执行若干步骤，并在每次执行这些步骤时，都从旧值算出新值
- 分支界限法
  - 把全部可行的解空间不断分割为越来越小的子集，谓之分支，并为每个子集计算一个下界或上界，谓之定界
  - 在每次分支后，对界限超出已知可行解的子集不再做进一步分支，搜索范围迅速收敛
  - 这一过程一直进行到找出可行解为止，该可行解的值不大于任何子集的界限，因此这种算法一般可以求得全局最优解

知识讲解

算法的思想方法（续3）

- 回溯法
  - 在包含问题所有解的解空间树中，从根节点出发，按深度优先搜索解空间树，当搜索到某一节点时，先判断该节点是否包含问题的解，若包含则从该节点出发继续搜索，否则逐层向其父节点回溯
  - 若希望求得问题的所有解，则必须回溯到根，且根节点的所有可行子树都必须被搜索到才能结束，否则只要搜索到问题的一个解即可终止

知识讲解

Technology Tarena 达内科技

算法的分类

算法的分类

- 算法按其执行期限可被分为
  - 有限算法
    - 算法过程无论长短，总能在有限的时间内终止
  - 无限算法
    - 算法过程因无终止条件或终止条件无法得到满足，而永不停息
- 算法按其解的确定性可被分为
  - 确定性算法
    - 对于确定的输入，算法总能得到确定的结果
  - 非确定算法
    - 对于确定的输入，算法得到的结果并不唯一确定

知识讲解

算法的分类（续1）

- 算法按其理论和应用的专门领域可被分为
  - 基本算法
  - 排序算法
  - 搜索算法
  - 代数算法
  - 几何算法
  - 数论算法
  - 图论算法
  - 随机算法
  - 并行算法
  - 分类算法
  - 数值分析算法
  - 数据结构算法
  - 动态规划算法
  - 压缩解压算法
  - 加密解密算法
  - 数据摘要算法
  - 数据挖掘算法
  - 随机森林算法
  - 人工智能算法
  - 模式识别算法

知识讲解

Technology Tarena 达内科技

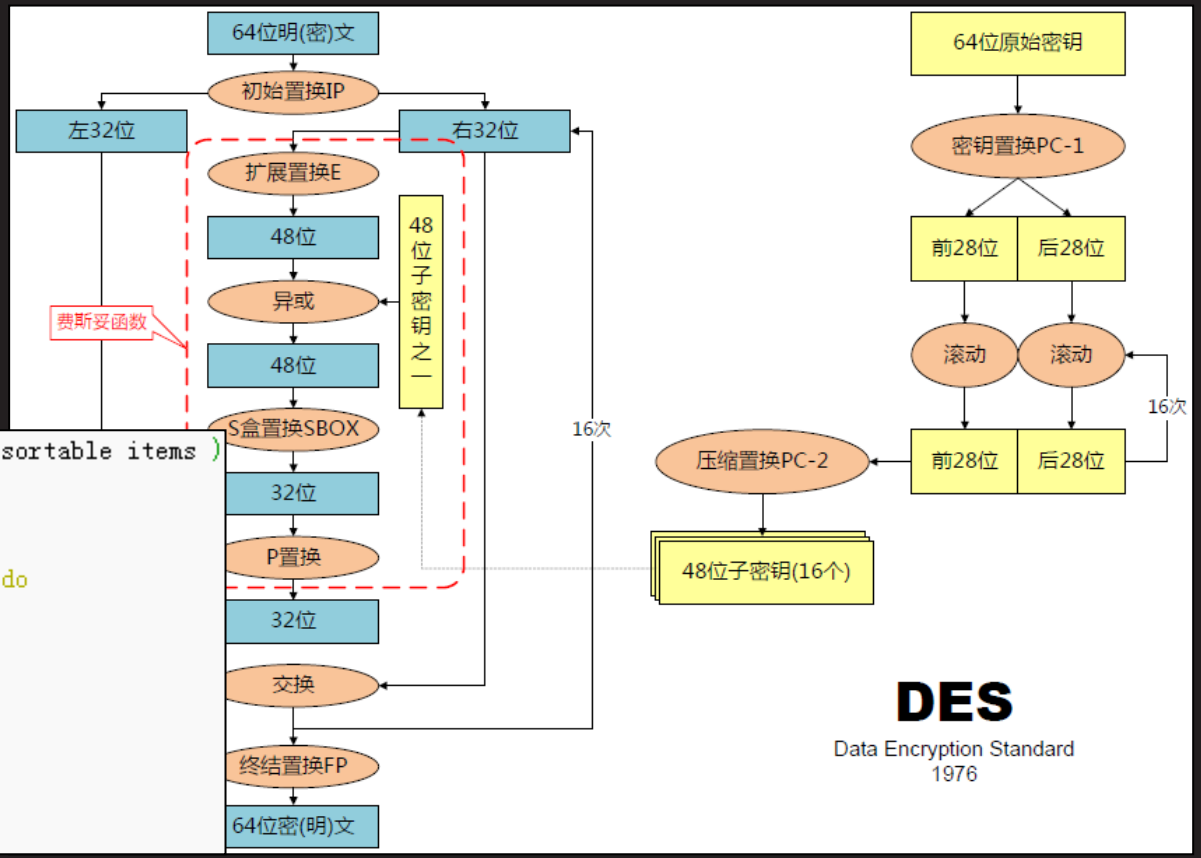
算法的描述

算法的描述

- 常见的算法描述方式包括如下几种
  - 自然语言
  - 伪代码
  - 结构化流程图
  - PAD图

知识讲解

```
procedure bubbleSort( A : list of sortable items )
    n = length(A)
    repeat
       swapped = false
       for i = 1 to n-1 inclusive do
          if A[i-1] > A[i] then
             swap(A[i-1], A[i])
             swapped = true
          end if
       end for
       n = n - 1
    until not swapped
end procedure
```



Technology Tarena 达内科技

总结和答疑