

数据结构实验 (11)

总结

- 时间复杂度
 - 时间复杂度的常数项也很重要
 - 最大堆 VS 平衡二叉树 实现优先队列
 - 关注算法的局部并行化、全局并行化能力
 - 局部并行化 - 循环展开、SIMD 指令并行
 - 全局并行化 - 多线程
 - 平均时间复杂度、最坏时间复杂度
 - 快速排序、哈希表

- C/C++
 - 《Google 编程规范》
 - 例如：关于函数参数与返回值
 - 内存管理问题
 - 堆空间 / 栈空间
 - `int *a` 与 `int a[]`
 - 内存空间申请与释放
 - C++ Style
 - RAII
 - 智能指针
 - C Style
 - 谁使用，谁申请：两次调用，第一次返回 Buffer 大小，第二次传入 Buffer
 - 函数内部申请，提供显式的释放操作

- 现代计算机与操作系统
 - CPU
 - 矢量计算指令集（局部并行化）
 - x86 SIMD 指令
 - ARM NEON 指令
 - 多核（全局并行化）
 - *寄存器、Cache 与 内存
 - 机械磁盘与 SSD

- 分布式计算框架
 - Map-Reduce
 - 分布式计算解决什么问题?

系统知识

时间复杂度

C/C++

Modern Computer

MapReduce 等

- 线性表 / 链表
 - 注意内存管理的原则（用 RAII（智能指针）还是 C Styles ..）
 - 内存管理：
 - 连续 / 跳跃，堆空间 / 栈空间
 - `std::vector` 与 `std::list` 的内存管理方式
 - 常见问题：
 - 基本操作：插入、删除、合并、翻转、去重
 - 链表特殊操作：判断环（Loop）

基本数据结构

线性表

链表

系统知识

时间复杂度

C/C++

Modern Computer

MapReduce 等

- 分块思想
 - 大段维护、局部朴素
 - Bitmap 与 桶
 - 哈希表
 - 哈希冲突：
 - 拉链法
 - Cuckoo Hash：哈希效率改进
 - 字符串匹配与查找：Rabin-Karp
 - 高维数据：GeoHash、局部敏感哈希（LSH）
 - BloomFilter：大规模数据去重

专题问题

哈希与暴力查找

算法思想

分块思想

基本数据结构

线性表

链表

系统知识

时间复杂度

C/C++

Modern Computer

MapReduce 等

- 栈 (Stack)
 - 函数栈的概念与递归思想
 - 深度优先遍历

- 常见问题:
 - 特殊操作:
 - 反转栈元素
 - 栈增加 get_max() 操作
 - 双栈模拟队列
 - 括号匹配 等利用递归思想解决的问题
 - 递归程序改非递归
 - 快速排序非的递归实现
 - 树的各种遍历的非递归实现

专题问题

哈希与暴力查找

算法思想

分块思想

基本数据结构

线性表

链表

栈

系统知识

时间复杂度

C/C++

Modern Computer

MapReduce 等

- 队列 (Queue)
 - 队列与多线程编程
 - *Copy-On-Write Queue
 - 资源调度
 - 广度优先遍历
 - 优先队列
 - 优先队列的概念
 - 优先队列常见应用
 - 计时器
 - 许多算法的优化方法，例如：图的优先队列遍历框架

- 常见问题
 - 反转队列
 - 队列模拟栈
 - 优先队列相关的各种问题，如 Top K

专题问题

哈希与暴力查找

算法思想

分块思想

基本数据结构

线性表

链表

栈

队列

系统知识

时间复杂度

C/C++

Modern Computer

MapReduce 等

- 搜索思想
 - 深度优先遍历
 - 迭代加深深度优先搜索
 - 广度优先遍历
 - 图的优先队列遍历框架
 - 常见问题：
 - 树、图的遍历问题（见后文）
 - 状态空间的最优解、所有解
 - 问题与图遍历问题的转化，例如：八皇后问题、背包问题、朋友圈问题

专题问题

哈希与暴力查找

算法思想

分块思想

搜索思想

基本数据结构

线性表

链表

栈

队列

系统知识

时间复杂度

C/C++

Modern Computer

MapReduce 等

- 树
 - 树的存储方式
 - 线性表：完全二叉树
 - 链表：各种树
 - 优先队列的实现
 - 二叉堆
 - 其他高级的堆，解决堆合并、堆元素更新等问题

专题问题

哈希与暴力查找

算法思想

分块思想

搜索思想

系统知识

时间复杂度

C/C++

常见问题

- 二叉树、树的遍历相关问题：
 - 二叉树的前序、中序、后序遍历
 - 树的层次遍历
 - 判断是否为二叉树排序树、是否为二叉堆
 - 树的最大、最小深度，树的直径
 - 二叉树中和为某一值的路径数目
 - 节点的最近公共祖先
 - 树的直径
- 二叉树的变换：如树的反转（flip）
- 二叉树的构造：如前序+中序、前序+后序、中序+后序构造二叉树
- 二叉搜索树相关：
 - 第 K 大/小 元素

高级数据结构

树

基本数据结构

线性表

链表

栈

队列

Modern Computer

MapReduce 等

- 图
 - 图的存储方式
 - 各种操作的时间、空间复杂度
 - 优先队列遍历框架
 - Dijkstra 和 Prim 在该框架下的实现
 - 并查集与连通分量
 - 并查集的优化
 - 工业界中图论相关的三类问题：
 - 可达性问题
 - 中心化问题
 - 社群发现问题
 - MapReduce 框架下大规模图算法实现

专题问题

哈希与暴力查找

算法思想

分块思想

搜索思想

系统知识

时间复杂度

C/C++

Modern Computer

MapReduce 等

- 常见问题
 - 原始问题：
 - 图的连通性
 - (强) 连通分量 / 两点间连通性
 - 无向图、有向图环路检测
 - 拓扑排序
 - 图中顶点间的最短距离
 - 两点最短距离
 - 所有点对的最短距离
 - 图的最小生成树
 - 如何把问题转换成图论问题？
 - 例如：给定 N 个字符串，找出一个最短的字符串，使得这 N 个字符串都是其子串。子串不要求连续，要保持原始顺序。

高级数据结构

树

图

基本数据结构

线性表

链表

栈

队列

- 排序
 - 快速排序、归并排序与分治思想
 - 双调排序
- 二分查找

- 常见问题
 - 二分查找
 - 中位数 / Top K
 - 众数
 - 出现次数超过 $N/2$ ，出现次数最多
 - 动态更新的众数
 - 仅出现一次的数
 - 去重
 - 两路归并、多路归并
 - 复杂一些的：
 - 逆序对

专题问题

哈希与暴力查找

排序与二分查找

高级数据结构

树

图

算法思想

分块思想

搜索思想

分治思想

基本数据结构

线性表

链表

栈

队列

系统知识

时间复杂度

C/C++

Modern Computer

MapReduce 等

- 数据结构与算法是一门重要的课程

- 在一学期的课程里我们能学到什么：

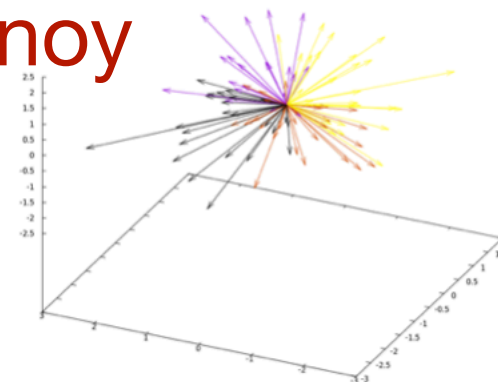
- 针对不同的问题规模，学会用不同的数据结构与算法，在确定的时间内解决问题

- Log-Based K-V Database

- 大段维护、局部朴素
- Annoy



现代文件系统在 磁盘=>SSD 的优化



K邻近问题与LSH

专题问题

哈希与暴力查找

排序与二分查找

高级数据结构

树

图

算法思想

分块思想

搜索思想

分治思想

基本数据结构

线性表

链表

栈

队列

系统知识

时间复杂度

C/C++

Modern Computer

MapReduce 等

- 寒假可以干什么？
 - 读书
 - 《数学之美》
 - 实践
 - 《集体智慧编程》
- 刷题：
 - LeetCode 一天1题？
 - 离 2021 年 4 月~ 5 月的暑期实习校园招聘还有 14 个月

