



# DeepSeek: 涉及到哪些数据结构?

## DeepSeek用到哪些数据结构

DeepSeek作为一个先进的AI模型或系统,其实现依赖于多种数据结构来高效处理数据和优化计算。以下是其可能涉及的关键数据结构:

### 1. 张量 (Tensor)

- **作用:** 深度学习的基础,用于表示输入数据、模型参数、梯度等。
- **场景:** 矩阵运算、卷积操作、激活函数计算等。

### 2. 计算图 (Computational Graph)

- **作用:** 以图结构描述神经网络的操作流程,节点为运算(如矩阵乘法),边为数据流动(张量)。
- **优化:** 支持自动微分和并行计算加速。

### 3. 键值对存储 (Key-Value Pairs)

- **作用:** 在Transformer的自注意力机制中,存储查询(Query)、键(Key)、值(Value)矩阵。
- **优化:** 通过高效的键值检索计算注意力权重(如缩放点积注意力)。

### 4. 缓存结构 (KV Cache)

- **作用:** 生成式模型(如GPT)在推理时缓存历史Key-Value对,避免重复计算。
- **实现:** 通常使用数组或链表管理,支持动态扩展。

### 5. 树结构 (Tree)

- **场景:** 决策树集成模型(如GBDT)、注意力机制中的层次化查询。
- **示例:** MoE(混合专家)模型中的专家选择可能依赖树结构加速路由。

### 6. 图结构 (Graph)

- **应用:** 若集成知识图谱,图数据库(如Neo4j)的邻接表或邻接矩阵用于关系推理。
- **场景:** 社交网络分析、推荐系统中的图神经网络(GNN)。

### 7. 队列与缓冲区 (Queue/Buffer)

- **数据加载:** 多线程/分布式训练中,队列缓冲训练数据(如 `tf.data` 管道)。
- **强化学习:** 经验回放缓冲区(Replay Buffer)存储状态转移数据。

### 8. 哈希表 (Hash Table)

- **分布式训练:** 参数服务器使用分布式哈希表同步模型参数。
- **特征工程:** 快速查找嵌入表(Embedding Table)中的类别特征。

### 9. 优先队列 (Priority Queue)

- **应用:** 在Beam Search算法中维护候选序列,按概率排序。
- **强化学习:** 优先经验回放(Prioritized Experience Replay)。

### 10. 稀疏矩阵 (Sparse Matrix)

- **场景:** 处理自然语言中的稀疏特征(如词袋模型)或推荐系统中的用户-物品交互矩阵。
- **优化:** 压缩存储格式(如CSR、CSC)节省内存。

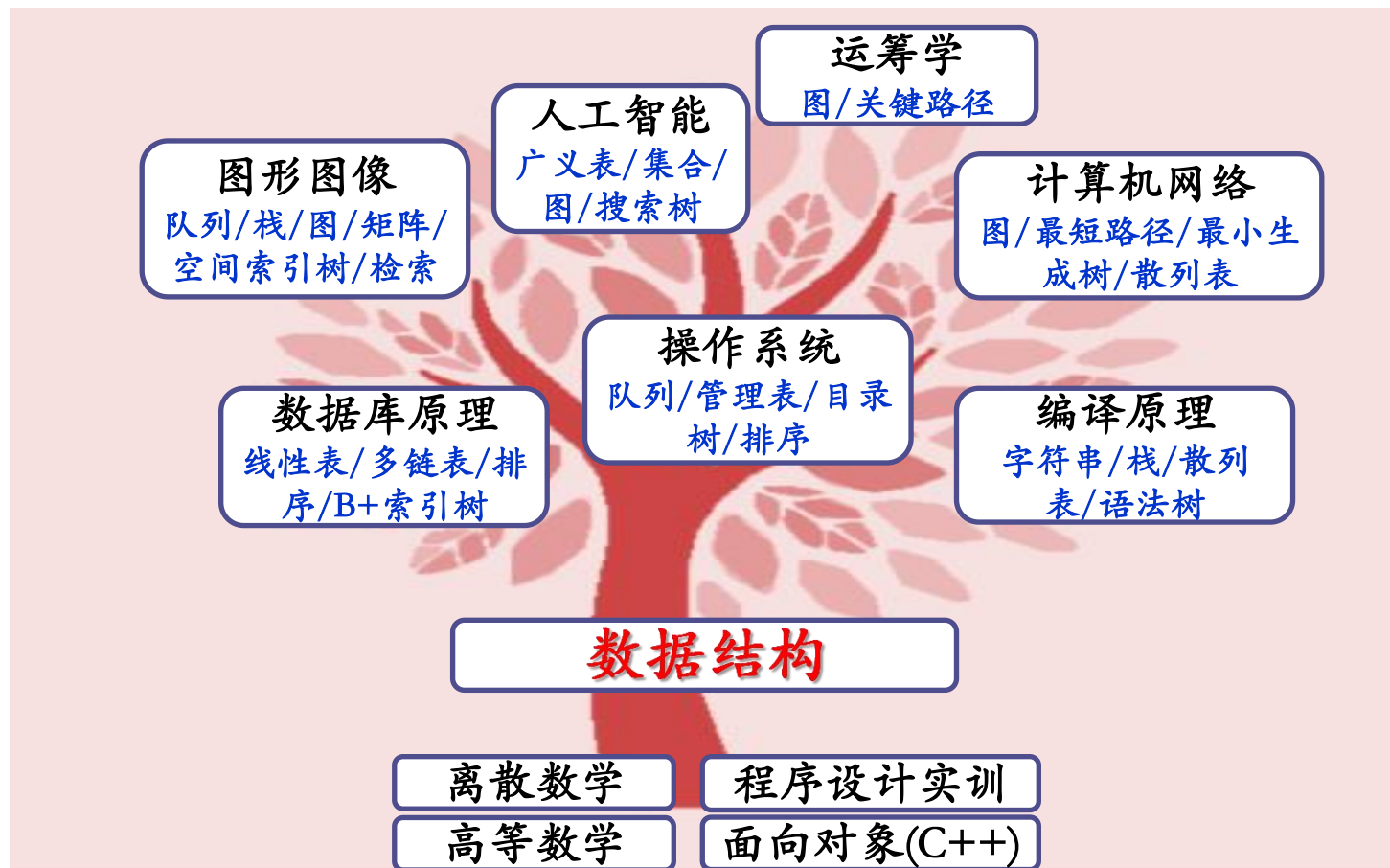
### 11. 堆 (Heap)

- **场景:** Top-K采样(如生成文本时选择概率最高的Token)。

## 总结

DeepSeek的设计综合了深度学习模型的通用数据结构和特定优化策略。其核心依赖于**张量**和**计算图**实现高效计算,同时结合**键值缓存**、**树/图结构**等处理复杂任务。不同模块可能针对性能需求(如低延迟、高吞吐)选择特定结构,例如哈希表加速查询,稀疏矩阵减少内存占用。具体实现细节可能因模型版本和应用场景而异。

# 为什么要学习数据结构？ - 专业知识体系



# 为什么要学习数据结构？ - 求职

## 腾讯（软件开发-后台开发方向）岗位要求：

### 必须具备的：

扎实的编程能力；

熟练掌握C/C++/Java/Go等其中一门开发语言；

TCP/UDP网络协议及相关编程、进程间通讯编程；

专业软件知识，包括**算法**、操作系统、软件工程、设计模式、**数据结构**、数据库系统、网络安全等。

### 有一定了解的：

Python、Shell、Perl等脚本语言；

MySQL及SQL语言、编程；

NoSQL, Key-value存储原理。

### 可以加分的：

分布式系统设计与开发、负载均衡技术，系统容灾设计，高可用系统等知识；

对云原生相关技术有所有了解。

# 为什么要学习数据结构？ - 求职

## 阿里（研发工程师C/C++）职位要求：

### 必备项：

热爱计算机及互联网技术，精通但不限于C/C++等编程语言中的一种或几种，有良好的编程习惯；

熟悉网络编程和多线程编程，对TCP/IP，HTTP等网络协议有一定的理解；

熟悉Unix/Linux/Win32环境下编程，熟练使用调试工具，并熟悉某种脚本语言；

对**数据结构**、算法有一定了解；

学习能力强，不畏挑战，喜欢钻研，有良好的沟通能力和团队协作能力，善于独立思考并反思总结。

### 加分项：

有互联网公司实习经验，参与过互联网软件等相关产品开发项目；

在相关领域国际顶级会议、期刊发表论文，或相关学术会议组织的权威比赛中获奖；

热衷于ACM，在校期间参与过大学生数学建模竞赛，“挑战杯”，机器人足球比赛等，或作为骨干参与过学生网站的建设和开发。



# 为什么要学习数据结构？ - 求职

## 字节跳动（ 算法工程师-国际电商） 职位要求：

计算机或相关专业，本科及以上学历；

具备扎实的代码功底，具备Linux环境研发能力，精通Python、Go或C++之一；

具备**扎实的数据结构/算法基础**，熟练掌握机器学习/深度学习理论知识，并具有丰富的实践经验；

熟悉机器学习、自然语言处理、计算机视觉、多模态、文本/数据挖掘中的1-2个领域；

优秀的分析和解决问题的能力，对挑战性问题充满激情；

具备良好的团队合作精神，较强的沟通能力；

有平台治理或内容安全经验者优先。

# 为什么要学习数据结构？ - 面试

## 学长对数据结构重要性的看法：

**谢同学（字节）：**数据结构属于面试时必问知识点，尤其是在招聘本科生时，由于本科生在专业深度上不如研究生，所以尤为侧重对基础知识的考查，往坏了说，如果在面试的时候回答不出数据结构方面的知识，会严重影响到面试者的评分。

**李同学（字节）：**数据结构是一门必须掌握的很基础的面试入门知识，至少需要对其基本的都能够有所掌握，此外还考察红黑树、二叉树、还有算法题。

**蒋同学（百度）：**程序=数据结构+算法，在面试时考查的比例大约有30%-40%，除此之外面试的时候很考察你的抽象能力，而数据结构与算法正好适合作为考察抽象能力时的考点。

### 四、技术一面

技术面和机考差不多，栗栗姐都会提供相应的八股文+资料，然后自己去准备八股的同时也要刷刷题保持手感。

1.手撕：面试官上来就手撕一个简单算法题，是发送在聊天框的，具体的话是让写两个函数：一个用来向队列中添加元素；另一个获取当前队列中的中位数。

2.Java基本数据类型

3.包装类型缓存机制创建的缓存数据是在Java内存的哪个区域

4.成员变量和局部变量在哪个区域

5.创建线程的方式

6.线程之间怎么通信的

7.多线程编程中加锁的方式有哪些

8.垃圾收集算法

### 我的机考题目

第一个题目名是围棋的气，网上可以搜到，类比力扣，矩阵类题目，难点分别是二维数组到棋盘对象的映射，棋子去重（set存放，考数据结构），棋盘边界条件的判断。

第二题考勤信息，难点是规则映射为数据结构（list），滑动窗口进行规则判断。第三题是项目排期，用贪心算法+排序，下来看发现还可以用二分查找优化。

一面，上来先自我介绍，然后直接开始手撕算法，需共享屏幕。可能是我非科班无经验，给我的题目是一道简单题，是**栈的问题**。

**Leetcode.20（有效的括号）**。面试官发了题目和几个用例，**在本地直接写代码并输出给的样例就行**。由于有思路，几分钟就写完了。接着就开始提问Python的八股，主要问题有：**动态语言、数据类型、生成器、修饰器、队列的实现等**，由于我的简历上有mysql，也问了很多相关问题。考察的是基础，但面试官会在每个问题后都追问更深的知识，考察你的能力，被好好拷打了，面试官可能是见我答不上来后面也没深挖了，计算机基础好的同学可以好好准备一下，可能提高最后的定级。

# 为什么要学习数据结构？-计算机科学领域：图灵奖得主



Edsger W. Dijkstra  
1972, Netherlands  
ALGOL之父

提出单源最短路径Dijkstra算法



Tony Hoare  
1980, UK  
霍尔逻辑  
提出快速排序算法



Donald E. Knuth  
1974, USA  
算法分析之父



Michael O. Rabin  
1976, Israeli  
非确定自动机  
素数判定随机算法



Dana S. Scott  
1976, USA  
非确定自动机



Robert W. Floyd  
1978, USA  
最短路径Floyd算法



Stephen A. Cook  
1982, USA  
NP完全性



Richard M. Karp  
1985, USA  
NP完全性与  
网络流算法



John Hopcroft  
1986, USA  
最差情况分析  
数据结构与算法



Robert Tarjan  
1986, USA  
数据结构与图算法



Juris Hartmanis  
1993, Latvia  
计算复杂性理论



Richard E. Stearns  
1993, USA  
计算复杂性理论



Manuel Blum  
1995, Venezuela  
计算复杂性理论



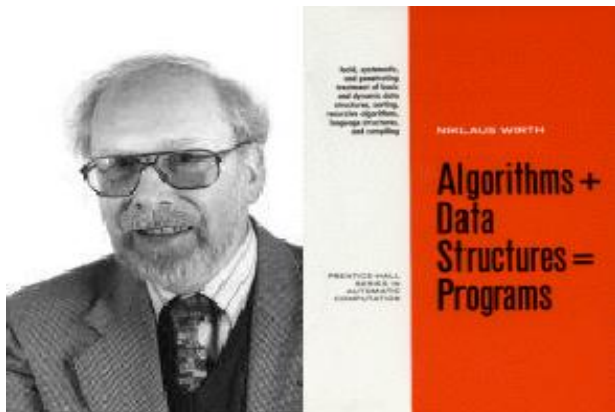
Andrew Yao  
2000, China  
伪随机数生成  
与通信复杂性



Leslie G. Valiant  
2010, Hungarian  
#P完全性与  
计算学习理论



# 程序 = 算法 + 数据结构



程序是对所要解决问题的各个对象和处理规则的描述，或者说是数据结构和算法的描述。

—Niklaus Wirth  
PASCAL语言之父  
结构化程序设计的首创者  
1984年图灵奖获得者

## 为什么要学习数据结构？ - 突破自己

- “精益求精，工匠精神”，设计实现性能更优的代码。
- “知其然知其所以然”，懂得各种封装接口、代码库背后的原理。
- “越过山丘，仁者见仁，智者见智”，看待问题的深度、理解问题的角度完全不一样。



## 课程基本信息

|      |                                  |      |                                    |
|------|----------------------------------|------|------------------------------------|
| 开课单位 | 信息科学与工程学院                        | 课程代码 | ZH108XK24                          |
| 课程名称 | 数据结构                             | 英文名称 | Data Structure                     |
| 课程性质 | 学类核心课                            | 学 分  | 4                                  |
| 总学时  | 72（课堂48+小班讨论8+实验16）              | 先修课程 | 高等程序设计                             |
| 任课教师 | 骆嘉伟、张羽丰、夏艳<br>蒋洪波、杨晓波<br>屈卫兰、谭光华 | 适应专业 | 计算机科学与技术<br>人工智能、信息安全<br>软件工程、通信工程 |
| 开课学期 | 2025春季                           |      |                                    |

# 教学目标

**(知识) CT1:** 掌握线性结构、树形结构和图结构等常用数据结构的抽象数据类型、物理实现及主要算法，形成程序员必备的基本数据结构工具箱。

**(知识) CT2:** 掌握排序和查找等重要问题类的有效算法，掌握基本的算法执行效率的度量分析方法，并具备基本的算法优化能力。

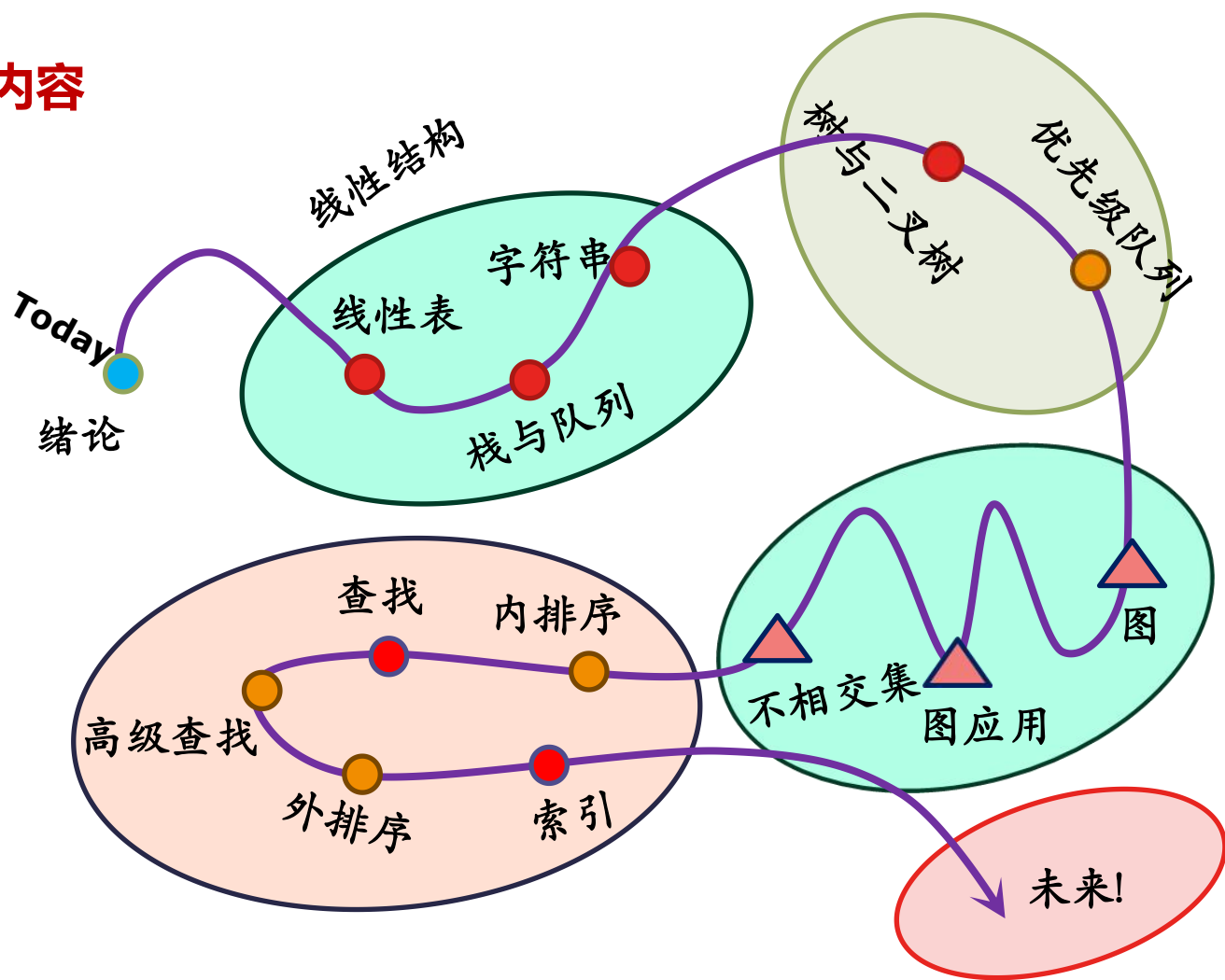
**(能力) CT3:** 能根据实际需求的需求，选择合适的数据结构进行表示、存储和处理。针对复杂工程问题所要求的资源限制，能够权衡时间与空间开销，设计更有效的解决方案。

**(能力) CT4:** 通过课程实验以及课程设计报告的撰写，训练学生计算思维能力、算法设计和分析能力以及程序设计与实现能力，进而提高学生在软件系统的认知、设计、开发、应用等方面的能力。

**(能力) CT5:** 通过课前预习、小班讨论、合作完成项目、文献的搜索和整理，培养学生自主学习、团队沟通和协作能力；培养对他人工作和自我表现的客观评价能力。

**(素养) CT6:** 知行合一提升学生核心素养，引导学生从理论、抽象和设计的角度解决问题，树立严谨的工程态度、科学的学术态度、精益求精的大国工匠精神。

# 教学内容





## 课程资料

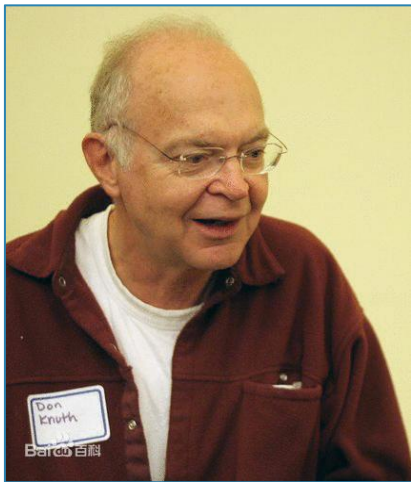
- 教材

- 数据结构. “101计划”核心教材,

俞勇、张铭、陈越、韩文弢主编. 高等教育出版社, 2024.7



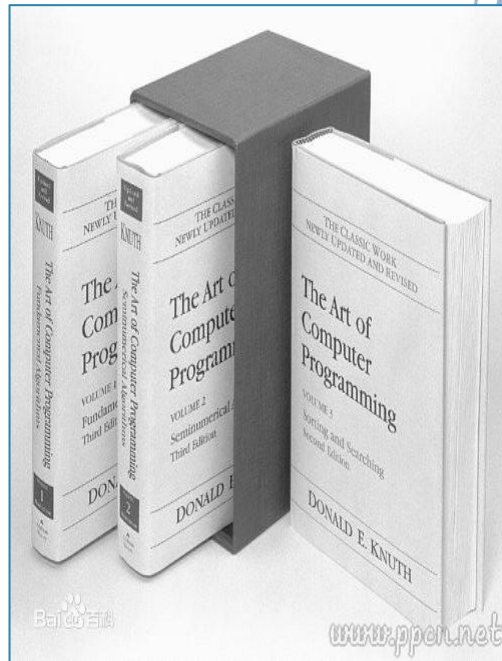
## Donald Ervin Knuth



### 《计算机程序设计的艺术》系列 “三部曲”

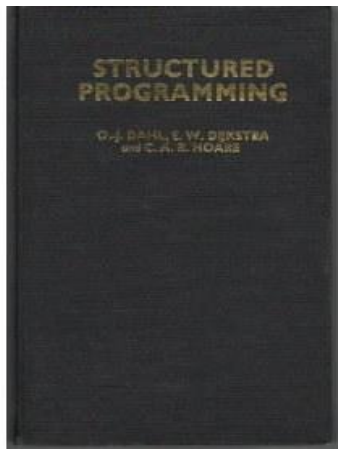
1. 第一卷《基本算法》，1968
2. 第二卷《半数字化算法》，1969
3. 第三卷《排序与搜索》，1973

1974年图灵奖的获得者



开创了数据结构与算法  
课程体系

# 经典



*O.-J. Dahl, E. W. Dijkstra and C. A. R. Hoare (1972). Structured Programming. Academic Press.*



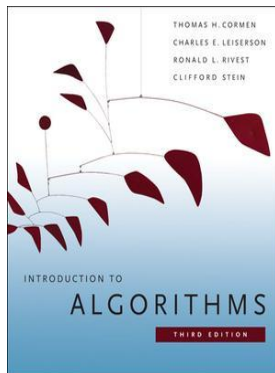
**Ole-Johan Dahl**  
**挪威**  
**2001年图灵奖**



**Edsger Dijkstra**  
**荷兰**  
**1972年图灵奖**



**C. A. R. Hoare**  
**英国**  
**1980年图灵奖**



***Introduction to Algorithms***  
**(1990) by Cormen** (2009杰出  
教员奖) , **Leiserson, Rivest**  
(2002年图灵奖) , **and Stein.**

# 课程资源

## • 课程网站

- 数据结构SPOC 网页登录湖南大学网络教学平台<https://mooc1.chaoxing.com/>，  
登录账号：学号，初始密码：s654321s
- 实验系统CG平台<https://cg.hnu.edu.cn/>，登录账号：学号，密码：原有密码，  
如果忘记请找助教重置

## • 超星学习通

- 学习通APP登录，首页选择其它登录方式
- 进入机构登录
  - 学校：湖南大学
  - 登录账号：学号
  - 初始密码：s654321s（可通过绑定手机号修改）
- 备注：
  - APP登录是跳过手机验证
  - 如果已经是超星的湖大学生用户，密码是个人设置的。



机构登录

学校/单位

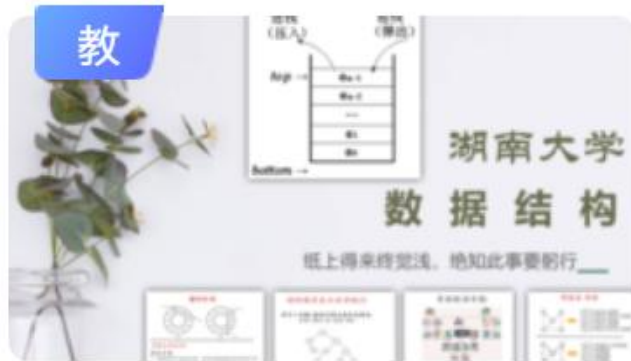
学号/工号

密码

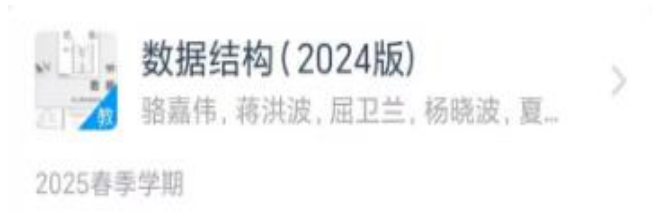
登录

## 课程资源

- 登录成功后，请确认班级是否正确，在课程模块，选择课程：数据结构（2024版）
- 在章节/课程介绍/阅读课程管理文档中的课程相关制度
- 在班级QQ群/下载电子课件
- 观看章节模块/第一章中的视频，解读课程相关制度，安排和要求。



数据结构（2024版）





## 课程考核

- 平时成绩(50%)
  - 线上学习(5%): 线上学习情况统计。
  - 线上测试(10%): 每次45-90分钟。
  - 课后作业(7%): 所有作业在SPOC中布置, 作业完成后拍照上传。
  - 小班讨论(8%): 根据学生课前资料准备(生生互评)、课堂小组报告和小班讨论课上表现等进行评价。
  - 实验(20%): 基础实验(测试结果评定)+综合实验(源码测试+实验报告评定)。
- 期中考试(10%), 闭卷, 90分钟。
- 期末考试(40%), 闭卷, 120分钟。

# 课程实验

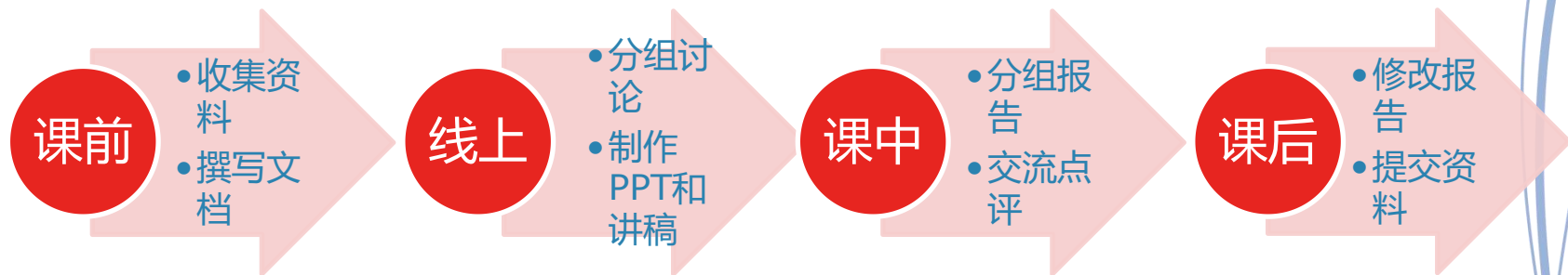
- 基础实验

- 基础实验是综合实验的基础，依据教学进程表，在规定时间内完成。
- 基础实验成绩依据线上测试结果评定。

- 综合实验

- 实验当天CG平台提交源代码，周六23:00之前在课程网站用附件的形式提交实验报告和源码。
- 综合实验成绩=源码测试(50%)+ 实验报告(50%)
  - 源码测试
    - CG自动评测系统评测。
    - 源码测试，采用白盒测试，检查代码编写是否符合课程要求。
  - 实验报告：依据内容从知识水平和表达水平两个维度评判。

## 小班讨论



讨论课成绩= 个人资料(20%)+小组报告(80%)

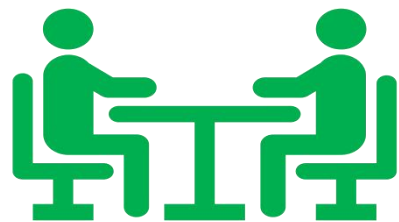
- 个人资料：采用生生互评方式，用讨论课资料生生互评的打分和评语生成小程序。
- 小组报告：采用随机分组方式，课前分组讨论，课堂上进行小组报告，老师和其他同学可以提出疑问，指出报告中的不足，给出修改建议。老师根据小组报告情况给出小组报告分数。

## 小班讨论: 个人资料

- 讨论课课前，每位同学独立收集资料，整理并撰写报告，在数据结构课程网络平台上提交报告文档。
- 在报告文档中引用了从各种渠道（书籍，论文，网络等）收集的资料，要给出**参考资料的引用说明**，而且一般应对资料进行整理，不能用直接复制的方式。（这是知识产权和学术诚信的要求，不遵守会扣分），参考文献的格式说明文档，在网课资料模块中下载，
- 资料内容采用电子格式完成，不能用拍照，截屏等方式。
- 涉及算法时一般要用自然语言、算法流程图或伪代码说明算法思想，尽量不使用源代码说明。根据题目要求设计合理的案例，尽量多用图表，文字应尽量凝练。
- **课前资料用附件的形式提交。讨论课提交的文件必须转换为PDF格式。文件名和文档中都不能出现本人信息。如果未按照要求提交，个人资料成绩记0分。**

## 小班讨论: 生生互评

- 提交课前资料文档截止后，启动生生互评。
- 生生互评时，可以多次评分，互评时间截止后不能补做。
- 生生互评要认真负责，公平公正，实事求是，必须用生生互评的评分量表生成评分和评语。
- 如果没有参与生生互评，或者评阅人工作不认真负责，被投诉，核实后取消评阅人本次讨论课的成绩。





我的作业 待批作业

| 作业标题 | 标题 | 互评开始日期           | 互评截止日期           | 操作 |
|------|----|------------------|------------------|----|
| 互评测试 | 1  | 2021-02-24 21:15 | 2021-02-24 22:15 | 查看 |

课程评价

互评测试

返回

| 姓名  | 班级  | 批阅时间             | 分数 | 状态 | 操作 |
|-----|-----|------------------|----|----|----|
| *** | *** | 2021-02-24 21:45 |    | 待批 | 批阅 |

1 互评测试

正确答案:

\*\*\*的答案:  
asdf

得分: 0 分

使用评分量表打分

上传附件

客观题得分: 0.0 分

总分: 0.0

提交

取消

评分量表模板详情

指标项名称: 提交规范和内容正确性与完整性

☐ 指标1 (0.0分):

提交方式不正确 (没有用附件形式提交), 得分0分

☐ 指标2 (0.0分):

文件格式和名字不正确 (没有提交pdf文件或文件名不规范或泄露个人信息), 得分0分

☒ 指标3 (40.0分):

按照要求提交课前收集准备的资料, 报告格式规范和内容完整, 得分40分

取消

确认

1 互评测试

正确答案:

\*\*\*的答案:  
asdf

得分: 92.0 分

使用评分量表打分

上传附件

按照要求提交课前收集准备的资料, 报告格式规范和内容完整, 得分40分

排版格式规范, 得分10分。

参考文献引用规范, 得分10分。

客观题得分: 0.0 分

总分: 92.0

提交

取消

## 小班讨论: 小组报告

- 报告PPT封面要求有讨论课周次、小组主题、班级、报告人姓名，小组成员的姓名。
- 讨论课程结束前，由报告人以附件形式提交小组报告PPT文件（有录音，录音可嵌入小组报告PPT文件或者单独以音频/视频文件形式提交）。文件名：第几周讨论课第几组报告讲稿.PPT，以及在作业提交框中，附上小组成员姓名和学号。
- 课程结束后，小组可以依据老师和同学提出的意见和建议，修改报告，重新录制报告，并由同一人再次提交修改报告和报告讲稿。

## 诚信、守时

- 端正学习态度、调动学习兴趣

- 提倡讨论，但严禁抄袭

- 可以讨论思路，请同学看算法的逻辑问题和效率问题。

- 但要亲自动手实现。

- 发现抄袭，则抄袭者和被抄袭者本次作业、讨论、实验记0分。

- 按时完成各项任务

- 所有视频学习、课后作业、实验（源代码和实验报告）、小班讨论（课前资料、讨论课件及录音）等都必须指定的期限内完成并在数据结构SPOC课程网站提交。

# 课程收获三层次



知识



方法（能力）



素养（习惯）

# 具体安排

|         |  |      |            |      |     |
|---------|--|------|------------|------|-----|
| 课程代码    | ZH108XK24  | 课程名称 | 数据结构       |      |     |
| 授课学期    | 2024-2025 学年第二学期   | 授课班级 | 计科 2401-02 | 授课教师 | 骆嘉伟 |
| 选用教材    | 《数据结构》，俞勇、张铭、陈越、韩文弢著. 高等教育出版社.   |      |            |      |     |
| 授课时间，地点 |  |      |            |      |     |
| 大班授课    | 2025 年 2 月 16 日—2025 年 6 月 7 日（1-16 周）   |      |            |      |     |
|         | 1. 星期二第 3、4 节，研 A302；                      2. 星期四第 1、2 节（单周），研 A404  |      |            |      |     |
| 小班讨论    | 2025 年 2 月 16 日—2025 年 6 月 7 日（4、8、12、16 周）<br>计科 2301 班：周四第 1、2 节， 研 B205<br>计科 2302 班：周四第 9、10 节， 研 B304 |      |            |      |     |
| 实验课     | 2025 年 2 月 16 日—2025 年 6 月 7 日<br>线上实验：CG 平台（4，8，12，14 周）<br>课堂综合实验：星期二第 5-8 节（5，9，13，15 周），院机房            |      |            |      |     |
| 答疑时间、地点 | 2025 年 2 月 16 日—2025 年 6 月 7 日，星期二 15: 30-17: 00，院 443   |      |            |      |     |
| 网络答疑    | 学习通 APP，QQ   |      |            |      |     |

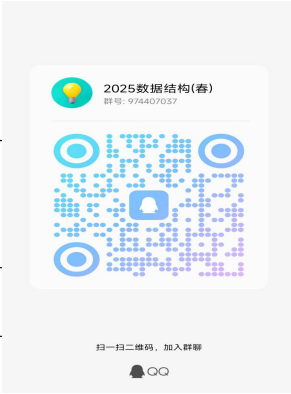


2025数据结构(春)  
群号: 974407037



扫一扫二维码，加入群聊







征途漫漫，  
惟有奋斗。

**成功=勤奋 and 勤奋!**