

公共基础

第一章 数据结构与算法

§ 1.1 算法

1. **算法的定义**：是指解题方案的准确而完整的描述。（算法不等于程序，程序的设计不可能优于算法的设计）
2. **算法的基本特征**：可行性、确定性、有穷性、足够的情报。
3. **算法的基本要素**：
 - ① 对数据对象的运算和操作：
算术运算、逻辑运算、关系运算、数据传输。
 - ② 算法的控制结构：
 - a. 算法中各操作之间的执行顺序；
 - b. 描述算法的工具通常有传统流程图、N-S 结构化流程图、算法描述语言等；
 - c. 一个算法一般可以用顺序、选择（分支）、循环（重复）三种基本结构组合而成。
4. **算法的时间和空间复杂度**： 计算100次就是100次计算工作量
 - ① 时间复杂度：是指执行算法所需要的 计算工作量，可以用算法所执行的基本运算次数度量。
 - ② 空间复杂度：是指执行算法所需要的内存空间。包括算法程序、输入的初始数据以及算法执行过程中需要的额外空间。
 - ③ 算法的时间复杂度和算法的空间复杂度相互独立。

§ 1.2 数据结构的基本概念

1. **数据**：需要处理的数据元素的集合，一般来说，这些数据元素，具有某个共同的特征。
 - a. 数据元素是数据的基本单位，即数据集合中的个体。
 - b. 有时一个数据元素可有若干数据项组成。数据项是数据的最小单位。
2. **结构**：是集合中各个数据元素之间存在的某种关系（或联系）。
3. **数据结构**：是指相互有关联的数据元素的集合。
4. **数据结构的分类**：
 - ① 逻辑结构：线性结构（线性表、栈、队列）；非线性结构（树、图）。
 - ② 存储结构：顺序存储；链式存储。
 - ③ 运算：插入、删除、查找、排序。
5. **逻辑结构**：反应数据元素间的逻辑关系（即前后件关系）的数据结构。
 - ① 线性结构（线性表）：（举例：春→夏→秋→冬）
 - a. 有且只有一个根节点，它无前件；
 - b. 每一个节点最多有一个前件，也最多有一个后件。
 - ② 非线性结构：
 - a. 不满足以上两个条件的数据结构就称为非线性结构；
 - b. 非线性结构主要是指树形结构和网状结构。
6. **存储结构**：又称为数据的物理结构，是数据的逻辑结构在计算机存储空间中的存放方式
 - ① 顺序存储结构：主要用于线性的数据结构，它把逻辑上相邻的数据元素存储在物理上相邻的存储单元里。
 - ② 链式存储结构：每一个结点至少包含一个指针域，用指针的指向来体现数据元素之间在逻辑上的联系。



- a. 一种逻辑结构可以有多种存储结构
- b. 不同的存储结构其数据处理的效率不同

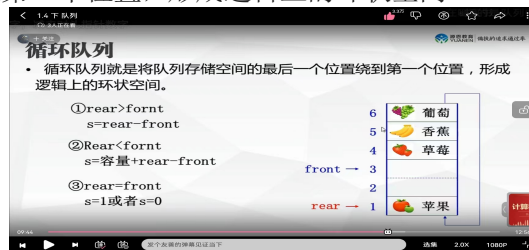
§ 1.3 线性表及其顺序存储结构

1. **线性表**: 数据的存储结构是指：数据的逻辑结构在计算机中的表示（因为计算机不知道逻辑结构，需要我们人来表示）
线性表是 n ($n \geq 0$) 个数据元素构成的有限序列，表中除第一个元素外的每一个元素，有且只有一个前件，除最后一个元素外，有且只有一个后件。
举例：英文字母表、地理学中的四向、表格
2. **线性表的顺序存储结构**:
 - ① 通常，线性表可以采用顺序存储和链式存储，但一般使用顺序存储结构。线性表的顺序存储又叫做顺序表（顺序分配）。
 - ② 特点：
 - a. 线性表中所有元素所占的存储空间是连续的；
 - b. 线性表中数据元素在存储空间中是按逻辑顺序依次存放的；
 - c. 可以随机访问数据元素；
 - d. 做插入、删除时需移动大量元素，因此线性表不便于插入和删除元素。

§ 1.4 栈和队列

1. **栈**: 栈是限定在一端进行插入和删除的线性表。
特点: ★
 - ① 栈是只能在栈顶进行插入和删除；
 - ② 栈的修改原则是“**先进后出**”或“**后进先出**”；
 - ③ 栈底指针 $bottom$ ，栈顶指针 top ，入栈，栈满，出栈；
 - ④ 栈底指针不变，栈中元素随**栈顶**指针的变化而动态变化；
 - ⑤ 栈具有记忆功能；
 - ⑥ 栈支持子程序调用。

子程序"和"函数"是编程中常用的两个术语，通常用来描述可被独立调用的代码块。在许多编程语言中，函数是子程序的一种实现方式。在我们的讨论中，"子程序"和"函数"可以互换使用。
2. **队列**: 队列是指允许在一端进行插入，而在另一端进行删除的线性表。
特点:
 - ① 队列只允许在队尾进行插入，而在队头进行删除；
 - ② 队列的修改原则是“**先进先出**”或“**后进后出**”；
 - ③ 队头指针 $front$ ，队尾指针 $rear$ ，入队，出队；
 - ④ 队列中元素随队头指针和队尾指针的变化而动态变化。
3. **循环队列**: 是讲队列存储空间的最后一个位置绕道第一个位置，形成逻辑上的环状空间
 $rear > front: s = rear - front$
 $rear < front: s = 容量 + rear - front$
 $rear = front: s = 1 \text{ 或者 } s = 0$



§ 1.5 线性链表

1. **线性链表**:
 - ① 线性表可以采用顺序存储和链式存储。线性表的顺序存储叫做顺序表，线性表的链

式存储结构叫做线性链表。

② 特点：

- a. 各数据结点的存储空间可以不连续；
- b. 各数据元素的存储顺序和逻辑循序可以不一致；
- c. 线性表的链式存储所占存储空间大于顺序存储结构；
- d. 查找结点时链式储存要比顺序存储慢；
- e. 链式存储插入删除元素比顺序存储灵活。

③ 线性链表的操作：在线性链表中插入与删除，不需要移动链表中的元素。

2. 线性表：

① 线性表顺序存储结构；

② 线性表链式存储结构（还包括双向链表、循环链表）。★

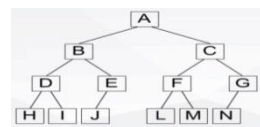
§ 1.6 树与二叉树 (★)

1. 树：

- ① 是 $n (n > 0)$ 个元素的有限集合。它有且仅有一个称为根的元素；其余元素是互不相交的子树。
- ② 常用术语：
 - a. 父结点、子结点；
 - b. 根结点、叶子结点； 没有子女的
 - c. 结点的度、树的度（所有结点中最大的度称为树的度）； 度就是有多少后代
 - d. 树的深度；
 - e. 子树（以某个结点的一个子结点为根构成的树称为该结点的一颗子树）。

2. 二叉树：

- ① 是一个有限的结点集合，该集合或者为空，或者有一个根结点及其两颗互不相交的左右二叉子树所组成。



② 特点：

- a. 非空二叉树只有一个根结点；
- b. 每一个结点最多有两颗子树，且分别称为该结点的左子树与右子树。

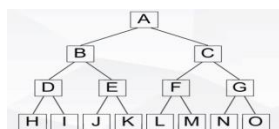
③ 五种基本形态：

- a. 空二叉树；
- b. 只有一个结点的二叉树；
- c. 只有左子树的二叉树；
- d. 只有右子树的二叉树；
- e. 左右子树双全的二叉树。

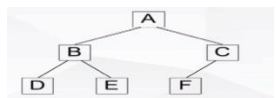


3. 特殊二叉树：

- ① 满二叉树：除最后一层外，每一层上的结点数均达到最大值。



- ② 完全二叉树：除最后一层外，每一层上的结点数均达到最大值，在最后一层上只缺右边的若干结点。



满二叉树是完全二叉树，但是完全二叉树不一定是满二叉树。

习题演练

[2005.4][2007.4][2009.3] 某二叉树有5个度为2的结点，则该二叉树中的叶子结点数为 9。

[2009.9][2010.9] 某二叉树有10个度为1的结点，7个度为2的结点，则该二叉树共有 25 个结点。

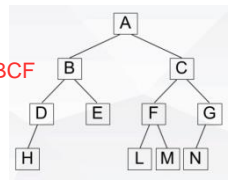
[2007.9] 某二叉树有70个叶子结点，80个度为1的结点，则该二叉树共有 219 个结点。

[2011.3] 某二叉树共有7个结点，其中叶子结点只有1个，则该二叉树的深度为(根结点在第1层) 7。

4. 二叉树特点: ★

非空二叉树只有一个根结点，每个结点最多有两颗子树，分别称为左子树和右子树

- ① 在二叉树的第K层上，最多有 2^{k-1} 个结点；**一行最多有多少个节点**
- ② 深度为 m 的二叉树最多有 $2^m - 1$ 个结点；**度为0: HELMN 度为2: ABCF**
- ③ 度为 0 的结点（叶子结点）总比度为 2 的结点多一个；
- ④ 有 n 个结点的二叉树深度至少为 $\lceil \log_2 n \rceil + 1$ 。

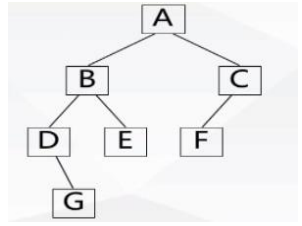


5. 二叉树的遍历:

$\lceil \log_2 11 \rceil + 1 = 3 + 1 = 4$

(按照一定的顺序访问二叉树中的结点，每个结点只被访问一次)

- ① 前序遍历: ABDGECF
访问根结点、前序遍历左子树、前序遍历右子树 (根左右)
- ② 中序遍历: DGBEAF C
中序遍历左子树、访问根结点、中序遍历右子树 (左根右)
- ③ 后序遍历: GDEBFCA
后序遍历左子树、后序遍历右子树、访问根结点 (左右根)



§ 1.7 查找技术

- 顺序查找:** 对于长度为 n 的线性表，平均要进行 $n/2$ 次比较，在最坏的情况下进行 **n** 次比较。
顺序查找适用于无序表或链式线性表 (不管无序还是有序) (适用于所有的线性表)
- 二分查找:** 适用于 **顺序存储的有序表**，对长度为 n 的线性表，在最坏的情况下进行 $\log_2 n$ 次比较。
注意: 即使是有序线性表，如果采用链式存储结构，也只能用顺序查找。

§ 1.8 排序技术

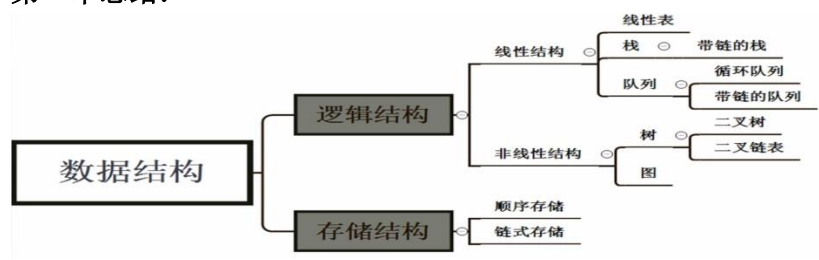
1. 排序:

排序		平均时间	最坏情况 (★)
交换类	冒泡排序	$n(n-1)/2$	$n(n-1)/2$
	快速排序	$n(n-1)/2$	$n(n-1)/2$
插入类	插入排序	$n(n-1)/2$	$n(n-1)/2$
	希尔排序	$n \log_2 n$	$n^{1.5}$
选择类	选择排序	$n(n-1)/2$	$n(n-1)/2$
	堆排序	$n \log_2 n$	$n \log_2 n$

2. 快速排序:

基本思想: 在要排序的序列中找一个数作为基准数 (通常为第一个数);
通过交换将这个序列中所有比基准数大的数放在右边, 比基准数小的数放在左边;
以基准数为分割线分为两个子表, 对两个子表重复上述步骤。

第一章总结:



第二章 程序设计基础

§ 2.1 程序设计方法与风格

1. 良好的程序设计风格：清晰第一，效率第二
2. 如何形成良好的程序设计风格：
 - ① 源程序内部文档化；
 - ② 数据说明；
 - ③ 语句的结构；
 - ④ 输入和输出
3. 源程序内部文档化：
 - ① 选择标识符的名字；
 - ② **注释**（序言性和功能性注释）：**序言性**注释一般位于模块的首部，用于说明模块的相关信息（标题、功能的说明、主要的算法、模块接口、开发历史、程序设计者、复审者和复审日期）；**功能性**注释位于源程序模块内部。
 - ③ 程序的视觉组织。

§ 2.2 结构化程序设计

1. 结构化程序设计原则：（★）
 - ① 自顶向下：先考虑总体，后考虑细节；先考虑全局目标，后考虑局部目标
 - ② 逐步求精：对复杂问题，先设计一个目标作为过渡，然后逐步细化
 - ③ 模块化：把程序要解决的总目标分解为一个一个的模块
 - ④ 限用 goto：限制使用 goto 语句，程序的质量与 goto 语句数量成反比
2. 结构化程序的基本结构：
结构化程序设计常采用顺序、选择（分支）和循环三种基本结构
3. 程序设计语言的基本成分：数据成分、运算成分、控制成分和传输成分。

§ 2.3 面向对象的程序设计

1. 几个术语：
 - ① 对象：在现实世界中，每个实体都是对象（例如大学生、汽车、电视机、空调等）；
 - ② 属性：用于描述对象的状态；
 - ③ 方法：用于描述对象的行为；
 - ④ 类：类是一组具有相同属性和相同操作的对象的集合
（基于同一类产生的对象可以分别设置各自的属性）
2. 对象的基本特点：
 - ① 标识唯一性：对象可由内在本质来区分。而不是通过描述来区分。
 - ② 分类性：可以将具有相同属性和操作的对象抽象成类。
 - ③ 多态性：同一操作可以是不同对象的行为。
 - ④ 封装性：从外面看不到对象的内部，只能看到对象的外部特征。
 - ⑤ 模块独立性好：对象是面向对象的软件的基本模块，内聚性强。
3. 继承：使用已有的类建立新类的定义技术。能直接获得已有的性质，不必重复定义他们。
4. 消息：是一个实例与另一个实例之间传递的信息。对象间的通信靠消息传递。
组成：①接收消息的对象的名称；②消息标识符，也称消息名；③零个或多个参数。
5. 多态性：是指同样的消息被不同的对象接受时可导致完全不同的行动的现象。

第三章 软件工程基础

§ 3.1 软件工程基本概念

1. **软件**：是由程序、数据及其相关文档构成的集合。
2. **软件**的分类：
 - ① 系统软件：操作系统、编译程序、汇编程序、网络软件、数据库管理系统；
 - ② 应用软件：事务处理软件、工程与科学计算软件、实时处理软件、人工智能软件；
 - ③ 支撑软件（工具软件）：需求分析工具、编译工具软件、测试工具软件、维护工具软件。
3. **软件危机**：需求增长、开发难控、质量难保、难以维护、提高成本、生产率低。
4. **软件工程**：
 - ① 定义：应用于计算机软件的定义、开发和维护的一整套方法、工具、文档、实践标准和工序。
 - ② 其目的是提高软件生产率、提高软件质量、降低软件成本。
 - ③ 其核心思想是**把软件当作一个工程产品来处理**。
5. **软件工程三要素**：
 - ① 方法：是完成软件工程项目的技术手段；
 - ② 工具：支持软件的开发、管理和文档生成；
 - ③ 过程：支持软件开发的各环节的控制和管理。
6. **软件生命周期**：
 - ① 定义：将软件产品能够提出、实现、使用维护到停止使用退役的过程称为软件生命周期。
 - ② 分为**软件定义、软件开发及软件运行维护** 3 个时期。维护是持续时间最长，花费代价最大的一个时期。
 - ③ 软件工程学的一个目的就是提高软件的可维护性，降低维护代价。
7. **3 个时期共有 8 个阶段**：
 - ① 软件定义：问题定义可行性研究、需求分析；
 - ② 软件开发：概要设计、详细设计、实现、测试；
 - ③ 软件运行维护：使用、维护。
8. **需求分析**：（★）
 - ① 确定系统的逻辑模型。参加人员有用户、项目负责人和系统分析员。
 - ② 其工作：需求获取、需求分析、编写需求规格说明书、需求评审。
 - ③ 产生文档：为需求规格说明书。
 - ④ 需求规格说明书的作用：
 - a. 便于用户、开发人员进行理解交流；
 - b. 反映用户问题的结构，可以作为软件开发工作的继承和依据；
 - c. 作为确认测试和验收的依据。
9. **需求规格说明书**（SRS）：需求分析阶段产生的主要文档是“软件需求规格说明书”。
特点：
 - ① 正确性：体现待开发系统的真是要求；
 - ② 无歧义性：对每个需求只有一种解释；
 - ③ 完整性：包括全部有意义的需求；
 - ④ 可验证性：每个需求都是可验证的；
 - ⑤ 一致性：各个需求的描述不矛盾；

- ⑥ 可理解性：需求说明书必须简明易懂；
- ⑦ 可修改性：结构风格在改变时，是易于实现的；
- ⑧ 可追踪性：每个需求的来源和流向是清晰的。

§ 3.2 结构化分析方法

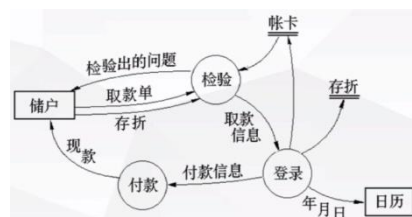
1. 结构化分析方法：

- ① 需求分析方法有：结构化需求分析方法和面向对象的需求方法
- ② 结构化分析方法：使用数据流图（DFD）、数据字典（DD）、判定表和判定树等工具，来建立系统的逻辑模型。

2. 数据流图（DFD）：

① 数据流图的图形元素

- 加工：输入数据经加工变换产生输出；
- **数据流**：沿箭头方向传递数据的通道；
- = 存储文件（数据源）：存放各种数据的文件；
- 源（潭）：系统和环境的接口。



- ② 数据字典（DD）是对数据流图中所有元素定义的集合，是结构化分析的核心。

§ 3.3 结构化设计方法

1. 软件分析是确定系统的逻辑模型，软件设计是确定系统的物理模型。

2. 软件设计的划分：

- ① 从工程管理角度划分：概要设计、详细设计。
- ② 按技术观点划分：结构设计、数据设计、接口设计、过程设计。

3. 软件设计基本原理：

- ① 抽象：在软件设计中，可以定出多个抽象级别，抽象层次从概要设计到详细设计逐步降低。
- ② 模块化：把一个待开发的软件分解成若干小的简单的部分，自顶向下逐层把软件换成若干模块。
- ③ 信息隐蔽：一个模块内的信息，对于不需要这些信息的其他模块来说不能访问。
- ④ 模块独立性：每个模块只完成独立的子功能，并且与其他模块的联系少且接口简单。模块的独立程度是评价设计好坏的重要度量标准。（**高内聚，低耦合**）

4. 软件模块独立性：

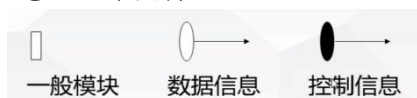
- ① 内聚性：指一个模块内部各个元素彼此结合的紧密程度（高内聚）。
- ② 耦合性：指模块间相互连接的紧密程度（低耦合）。非直接耦合、数据耦合、标记耦合、控制耦合、外部耦合、公共耦合、内容耦合。（★）

5. 概要设计：

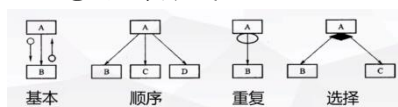
- ① 概要设计的任务：
 - a. 设计软件系统结构；
 - b. 数据结构及数据库设计；
 - c. 编写概要设计文档；
 - d. 概要设计文档评审。
- ② 概要设计的工具是程序结构图（SC）。

6. 程序结构图（SC）：

① 基本图符：



② 基本形式：



7. **详细设计：**

- ① 详细设计的任务：确立每个模块的实现算法和局部数据结构，用适当方法标识算法和数据结构的细节。
- ② 详细设计的常用工具：
 - a. 图形工具：程序流程图、N-S 图、PAD、HIPO
 - b. 表格工具：判定表
 - c. 语言工具：PDL（伪码）

8. **程序流程图：基本图符：**

- ① 控制流：表示为→或↓
- ② 加工步骤：表示为□
- ③ 逻辑条件：表示为◇

§ 3.4 软件测试

1. **软件测试：**

- ① 软件测试的目的是发现程序中的错误。
- ② 软件测试的准则：
 - a. 所有测试都应追溯到用户需求；
 - b. 在测试之前制定测试计划，并严格执行；
 - c. 充分注意测试中的群集现象；
 - d. 避免由程序的编写者测试自己的程序；
 - e. 不可能进行穷举测试；
 - f. 妥善保存测试分析报告，为维护提供方便。

2. **静态测试和动态测试：**

- ① 静态测试：不实际运行软件，通过人发挥思维优势发现程序的错误。
- ② 动态测试：基于计算机的测试，是为了发现错误而执行程序的过程。

3. **白盒测试和黑盒测试：**

- ① 白盒测试：把测试对象看作一个打开的盒子，利用程序内部的逻辑结构，对程序所有逻辑路径进行测试。（针对程序的内部逻辑结构）
 - a. 逻辑覆盖测试；
 - b. 基本路径测试。
- ② 黑盒测试：完全不考虑程序内部的逻辑结构，只检查程序是否能接收输入数据而产生正确的输出信息。（针对程序的外部功能）
 - a. 等价类划分法；
 - b. 边界值分析法；
 - c. 错误推测法。

4. **软件测试步骤：**（顺序不可换）

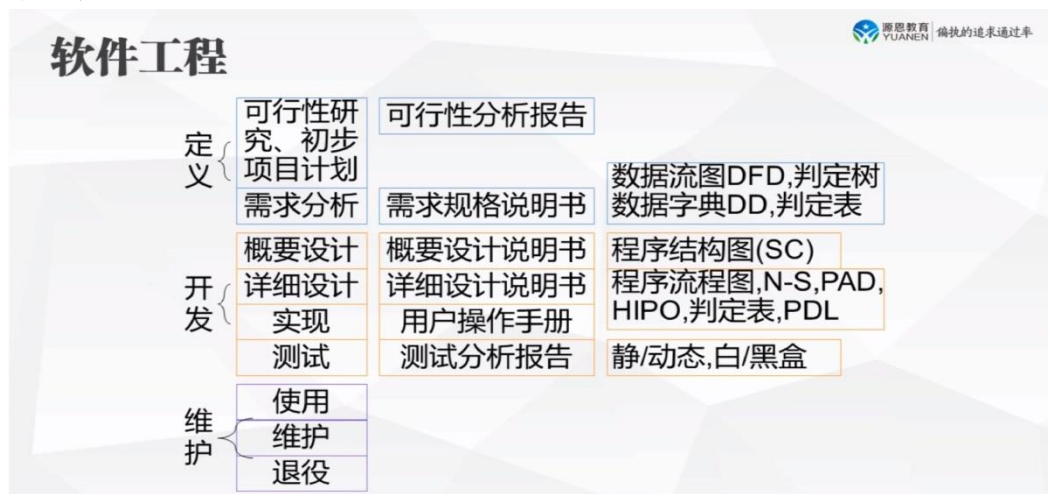
- ① 单元测试：是对软件设计的最小单位——模块进行测试，目的是发现各模块内部的错误。
- ② 集成测试：是把模块按照设计要求组装起来的同时进行测试，目的是发现与接口有关的错误。
- ③ 确认测试：是验证软件功能和性能是否满足各种要求，以及软件配置是否完全、正确。
- ④ 系统测试：是将软件作为一个元素，与计算机系统其他元素组合在一起，进行集成测试。

§ 3.5 程序的调试

1. 程序调试:

- ① 对程序进行了成功的测试之后将进入程序调试，通常称为 Debug（排错），主要在开发阶段进行。
- ② 程序调试的主要任务是**诊断和改正程序的错误**。
- ③ 基本步骤：
 - a. 错误定位；
 - b. 修改设计和代码，以排除错误；
 - c. 进行回归测试，防止引进新的错误。
- ④ 软件调试方法：强行排除法、回溯法、原因排除法。

第三章总结:



第四章 数据库设计基础

§ 4.1 数据库系统的基本概念

1. **数据** (data)：描述事物的符号纪录称为数据。
 - ① 实际上就是描述事物的符号纪录；
 - ② 软件中的数据一定是有结构的，有型和值两个概念。
2. **数据库** (DB)：是指长期存储在计算机内的，有组织的，可共享的数据集合。
数据库中的数据具有两大特点：“集成”“共享”。(★)
3. **数据库管理系统** (DBMS)：数据库管理系统是数据库系统的核心。(系统软件)
语言：
 - ① 数据定义语言 DDL：
 - a. 数据模式定义；
 - b. 数据存取的物理构建。
 - ② 数据操纵语言 DML：
 - a. 数据操纵，包括查询与增、删、改等操作。
 - ③ 数据控制语言 DCL：
 - a. 数据的安全性的定义域检查；
 - b. 并发控制与故障恢复；
 - c. 数据的完整性。
4. **数据库管理员** (DBA)：主要工作：
 - ① 数据库设计；
 - ② 数据库维护；
 - ③ 改善系统性能，提高系统效率。
5. **数据库系统** (DBS)：组成：
 - ① 数据库(数据)——集成，共享。
 - ② 数据库管理系统 DBMS(软件)——定义，构建，操纵，检查，控制，服务。DDL, DML, DCL.
 - ③ 数据库管理员 DBA(人员)——设计，维护，改善性能，提高效率。
 - ④ 软件平台——操作系统，开发工具，接口软件。
 - ⑤ 硬件平台——计算机，网络。
6. **数据库应用系统** (DBAS)：
数据库应用系统包括：数据库系统、应用软件以及应用界面。
7. **数据管理三个阶段**：人工管理、文件系统、数据库系统。
数据库技术的根本目标是：解决数据共享问题。
8. **数据库系统特点**：集成性，高共享低冗余，独立性，统一管理控制。
9. **三级模式和两级映射**：



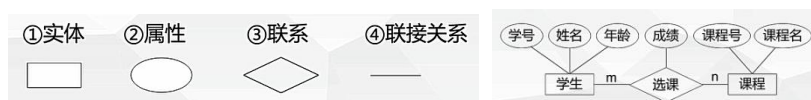
两级映射保证了数据库中数据具有较高的逻辑独立性和物理独立性。

§ 4.2 数据模型

1. 数据模型：

- ① 三要素：数据结构、数据操作和数据约束
- ② 分类：按不同的应用层次分为：
 - a. 概念数据模型（概念模型）：E-R 模型
 - b. 逻辑数据模型（数据模型）：层次模型、网状模型、关系模型、面向对象模型
 - c. 物理数据模型（物理模型）

2. E-R 模型：（实体联系模型）



联系：一对一，一对多（学生与宿舍），多对多（学生与课程）

3. 数据模型分类：

- ① 层次模型：树；
- ② 网状模型：无向图；
- ③ 关系模型：二维表。

4. 关系模型：采用二维表来表示，简称表，每一个二维表称为一个关系。

- ① 属性：二维表中的一列称为属性；
- ② 元组：二维表中的一行称为元组。（分量不可再分）
- ③ 关系操纵：查询、增加、删除和修改。
- ④ 关系中的数据约束：
 - a. 实体完整性约束；
 - b. 参照完整性约束；
 - c. 用户定义的完整性约束。

§ 4.3 关系代数

1. 关系代数：

- ① 关系模型的基本操作：插入、删除、修改、查询。

查询运算：

- a. 投影运算；（投影列）
- b. 选择运算；（选择行）
- c. 笛卡尔积运算（连接运算）（ $T=R \times S$ ）

- ② 关系代数中的扩充运算：

交运算、除运算、连接与自然连接运算。

并（ $T=R+S$ ）

差（ $T=R-S$ ）

交

除（ $T=R \div S$ ）

自然连接

R		S		T	
照片	位置	位置	特长	照片	位置
	前锋	前锋	灌篮		前锋
	前锋	中锋	篮板		前锋
	后卫				中锋
	中锋				篮板
	后卫				

自然连接

R		S		T	
学号	照片	球衣	位置	学号	照片
0101		10	前锋	0101	
0102		7	后卫	0101	
0201				0102	
				0102	
				0201	
				0201	

笛卡尔积

R		S		T	
学号	照片	球衣	位置	球衣	位置
0101		10	前锋	10	前锋
0102		7	后卫	7	后卫
0102		10	前锋		
0101		7	后卫		
0101		4	中锋		

除

小技巧：两个表为投影或选择，三个表为其他；属性列增加为自然连接，ST 属性列相加等于 R 为除。

§ 4.4 数据库设计与管理

1. **数据库设计概述**：设计一个能满足用户要求，性能良好的数据库。
 - ① 基本任务：
根据用户对象的信息需求、处理需求和数据库的支持环境设计出数据模式。
 - ② 两种方法：
 - a. 以信息需求为主，兼顾处理需求（面向数据的方法）。
 - b. 以处理需求为主，兼顾信息需求（面向过程的方法）。
 - c. 面向数据的设计方法已成为主流方法。
2. **数据库设计的步骤**：数据库设计目前一般采用生命周期法，分若干阶段：
 - ① **需求分析阶段**：建立数据字典；
 - ② **概念设计阶段**：设计 E-R 图；
 - ③ **逻辑设计阶段**：把 E-R 图转换为关系模式。实体与联系表示成关系，E-R 图中属性转换成关系的属性；
 - ④ **物理设计阶段**；
 - ⑤ 编码阶段；
 - ⑥ 测试阶段；
 - ⑦ 运行阶段；
 - ⑧ 进一步修改阶段。

在数据库设计中采用前四个阶段，并且重点以数据结构与模型的设计为主线。
3. **数据库管理**：
 - ① 数据库的建立；
 - ② 数据库的调整；
 - ③ 数据库的重组；
 - ④ 数据库安全性控制与完整性控制；
 - ⑤ 数据库的故障恢复；
 - ⑥ 数据库监控。

计算机基础

第一章 计算机的发展

§ 1.1 计算机的发展史

1. **ENIAC**（埃尼阿克）：战争催生了第一台电子计算机，30 多吨重，170 平方，速度却很低下，1946 年（20 世纪 40 年代）产于美国。
2. **冯·诺依曼**：现代计算机之父。
冯·诺依曼计算机原理：
 - ① 采用二进制：在计算机内部，程序和数据采用二进制形式进行存储。
 - ② 程序储存，自动执行：程序和数据存放在存储中，无需人工进行干涉。
3. **发展历程**：采用不同元件：
 - ① 电子管：稳定性好，功耗和体积大；
 - ② 晶体管：体积小；
 - ③ 中小规模集成电路：集成度越高，体积越小功耗越低；
 - ④ 大规模超大规模集成电路：集成度越高，体积越小功耗越低。
4. **电子计算机的发展过程**：（年份&器件考的最多，其他不考）

时代	年份	器件	软件	应用
一	46-57	电子管	机器语言 汇编语言	科学计算
二	58-64	晶体管	高级语言	数据处理 工业控制
三	65-70	中小集成电路	操作系统	文字、图形 处理
四	71年迄今	大规模 超大规模集成电路	数据库、网络 等	社会的 各个领域

§ 1.2 计算机的应用、特点和分类

1. **计算机的特点**：（不常考）
 - ① 高速精确的运算能力；
 - ② 准确的逻辑判断能力；
 - ③ 强大的存储能力；
 - ④ 自动功能；
 - ⑤ 网络与通信功能。
2. **计算机的应用**：
 - ① 早期：主要用于大型计算；
 - ② 现在：应用广泛：
 - a. 科学计算：
 - 进行数值运算；推动科研技术的发展；
 - 应用领域：**基因测序/轨道计算/天气预报/云计算**。
 - b. 数据/信息处理：
 - 进行非数值运算；图像、文字、声音等信息处理；
 - 典型应用：**OA 办公系统**。
 - c. 过程控制：（实时控制）应用于工业制造；

d. 计算机辅助:

- 让计算机代替人的部分工作
- CAD 计算机辅助设计; CAM 计算机辅助制造; CAI 计算机辅助教育;
- CAT 计算机辅助技术; CIMSS 计算机集成制造系统。

e. 网络通信: 购物、聊天、搜索;

f. 人工智能: 游戏中的人机对战模式;

g. 多媒体应用: 文本、图像、声音、视频;

h. 嵌入式系统: mp3、相机、手机、电视。

§ 1.3 未来计算机发展趋势 (不考? 了解一下即可)

1. 发展方向:

- ① 巨型化: 计算速度更快、存储容量更大、功能更完善、可靠性更高、运算速度可达万万亿次/秒、存储容量超过几百 T 字节。
- ② 微型化: 微型计算机正在循序向便携机、掌上机发展的价格、方便的使用、丰富的软件, 使其受到用户的青睐。
- ③ 网络化: 利用技术和计算机技术, 把分布在不同地点的计算机互联起来, 按照网络协议互相通信, 以共享软件、硬件和数据资源。
- ④ 智能化: 模拟人的感觉和思维能力, 智能计算机具有解决问题和逻辑推理的功能, 以及处理知识和知识库管理的功能等。

2. 未来新一代的计算机: (了解五种计算机的名字即可不用展开)

- ① 模糊计算机: 基于模糊理论, 能够实现模糊的、不确切的判断进行工程处理的计算机。
- ② 生物计算机: 以生物元件构建的计算机。
- ③ 光子计算机: 一种用光子信号进行数字运算、信息存储和处理的计算机。
- ④ 超导计算机: 用超导材料替代半导体材料制造的计算机。能耗小、运算速度快。
- ⑤ 量子计算机: 基于量子动力学规律进行高速数学和逻辑运算、存储及处理量子信息的计算机。

3. 电子商务: 参加主体分: (了解)

- ① B2B: 企业与企业
- ② C2B: 企业与消费者
- ③ C2C: 消费者与消费者

4. 信息技术: 层次:

- ① 信息基础技术: 新材料、新能源、新器件的开发和制造技术;
- ② 信息系统技术: 感测技术、通信技术、计算机与智能技术和控制技术;
- ③ 信息应用技术: 信息管理、信息控制、信息决策而发展起来的具体技术群类。

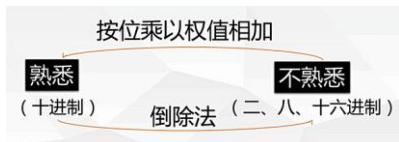
第二章 信息的表示和储存

§ 2.1 数据与信息

1. **数据**: 客观事物的符号表示, 如数值、文字、语言、图形、图像等都是不同形式的数据。
2. **信息**: 是事物变化的反映, 是事物联系作用的表征。既是对客观事物变化和特征的反映, 又是事物之间相互作用、相互联系的表征。
3. **信息与数据的区别**:
 - ① 数据是信息的载体, 信息是数据处理之后产生的结果;
 - ② 信息有意义, 而数据没有。
4. **计算机中数据的单位**:
 - ① 比特/位 (**bit**): 位是度量数据的**最小单位**。在计算机技术中用二进制表示数据, 1 位数据只能表示 0 和 1 两种代码。
 - ② 字节 (**Byte**): 字节是度量数据的**最小存储单位**。一个字节 (Byte) 由 8 位 (bit) 二进制数字组成。存储器容量通常以字节为单位 (Byte, B) 来描述。
 - 1 字节=8 位 (1B=8b)
 - 1kB=1024B; 1MB=1024kB; 1GB=1024MB; 1TB=1024GB……
 - ③ 字长:
 - a. 一次能处理的二进制数;
 - b. 字长越长处理速度越快;
 - c. 字长是字节的整数倍。如: 8 位、16 位、32 位、64 位。

§ 2.2 进制的转换

1. **进制的转换**: 人有十只手指, 因此选择了十进制。
熟悉: 十进制 (0-9 组成);
不熟悉: 二进制 (0 和 1 组成); 八进制 (0X, 0-7); 十六进制 (0X, 0-9 和 a-f)



2. **倒除法**: (商为 0 改转换结束) (余数需从下往上写) **相除、取余、倒着写**

要转换的几进制? **除数**

要转换的十进制数是几? **被除数**

商 **余数**

例: $(9)_{10} = (?)_2$

2	9	---- 余数为 1
2	4	---- 余数为 0
2	2	---- 余数为 0
2	1	---- 余数为 1
	0	

1001

例: $(15)_{10} = (?)_{16}$

16	15	---- 余数为 f
	0	

f

3. **位权法**: 按位乘以权值相加

$$\begin{aligned}(198)_{10} &= 1 \times 100 + 9 \times 10 + 8 \times 1 \\ &= 1 \times 10^2 + 9 \times 10^1 + 8 \times 10^0 \\ (1011)_2 &= (?)_{10} \\ &= 1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0 \\ &= 8 + 0 + 2 + 1 \\ &= (11)_{10}\end{aligned}$$

§ 2.3 字符的编码

1. 西文字符编码:

- ① 计算机中最常用的字符编码是 ASCII，美国信息交换标准交换代码。
- ② 微机中采用的是 ASCII，国际通用是 7 位 ASCII 码。
- ③ 计算机用一个字节（8 个二进制位）存放一个 7 位 ASCII 码，最高位置为 0。
- ④ 字符的编码：ASCII 码的排序，空格 < 数字 < 大写字母 < 小写字母

字符	ASCII 码值
空格	32
'0' ~ '9'	48 ~ 57
'A' ~ 'Z'	65 ~ 90
'a' ~ 'z'	97 ~ 122

2. 汉字的编码:

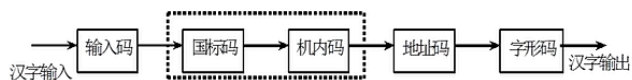
- ① 国标码（6763）：
 - 一级汉字：3755 个，按汉语拼音排列；
 - 二级汉字：3008 个，按偏旁部首排列。
- ② 区位码：
由 94 个区号和 94 个位号构成，每个汉字占两个字节。

3. 区位码转换为国标码:

区位码 \rightarrow 十六进制 + (20, 20)_H = 国标码

例 “火” 区位码 (27, 80)_D 转换为国标码
(27, 80)_D \rightarrow (1B, 50)_H + (20, 20)_H = (3B, 70)_H

4. 汉字的处理过程:



5. 汉字输入码:

- ① 音码类：全拼、双拼、微软拼音、自然码和智能 ABC 等；
- ② 形码类：五笔字型法、郑码输入法等；
- ③ 其他：语音、手写输入或扫描输入等。

6. 汉字内码:

2 个字节存储，每个字节最高位置“1”，一字一码无重码。

7. 国标码转换为内码:

国标码 + (80, 80)_H = 内码

8. 汉字字形码:

- ① 汉字字型点阵的代码。
- ② 用于汉字在显示屏或打印机输出。
- ③ 通常有两种表示方式：点阵和矢量表示方式。



9. 汉字地址码: 汉字库中存储汉字字形信息的逻辑地址码。

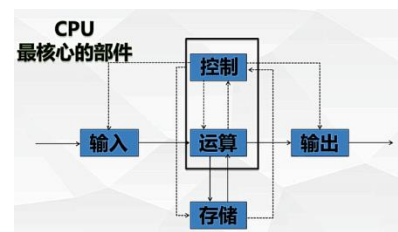
10. 其他汉字内码:

GBK——扩充汉字内码规范；
USC——通用多八位编码字符集；
Unicode——国际编码标准；
IG5——繁体汉字编码标准。

第三章 计算机硬件系统

§ 3.1 运算器

1. **运算器**（AU）：算术运算+逻辑运算。
 - ① 内部结构：ALU、寄存器、控制电路；
 - ② 执行操作：算术运算（ $+$ $-$ \times \div ）、逻辑运算（与或非）、移位操作（左移、右移）。
2. **中央处理器**（CPU）：运算器+控制器
 - ① 最主要的 3 个性能指标：字长、时钟主频、运算速度
 - ② 度量计算机运算速度常用的单位是 MIPS（百万次/秒）



§ 3.2 控制器

1. **控制器**（CU）：指令存储器+指令密码器+操作控制器+程序控制器
 - ① 作用：协调各部件有序工作
 - ② 组成：机器指令（操作码+操作数）+指令的执行过程

§ 3.3 存储器

1. 计算机的主机=CPU+内存
2. **存储器**：
 - ① 帮助计算机记忆信息，存储器是存取程序和数据的部件。
 - ② 类型划分：依据 CPU 是否可以直接访问：内存（容量小速度快）和外存（容量大速度慢）。

3. 内存：

RAM：随机存取存储器；

ROM：只读存储器。



4. 外存：

- ① CPU 不能直接访问外存；
- ② 硬盘、光盘、快闪存储器（flash）；
- ③ 硬盘：是主要的外部存储设备，内部结构：**磁头**（最贵最重要）、**磁道**（径向运动，同心圆由内向外逐渐减少）等。
- ④ 光盘：



种类	典型	原理
不可擦写	CD-ROM, DVD-ROM	结晶和非结晶间不能互换
可擦写	CD-R, CD-RW, DVD-R, DVD-RW	结晶和非结晶间可以互换

§ 3.4 输入输出设备

1. **输入设备**：键盘、鼠标、触摸屏、摄像头、扫描仪等；
2. **输出设备**：显示器、打印机。

第四章 计算机软件系统

§ 4.1 软件概念

1. **软件**: 软件是各种指令、程序和文档的总称。软件是计算机的灵魂, 是人机之间的接口。
2. **程序设计语言**:



§ 4.2 软件系统的组成

1. **系统软件**:
 - ① 为应用软件提供支持;
 - ② a. 操作系统 (OS): **最核心**; 五大功能: 处理器 (CPU) 管理、存储管理、文件管理、设备管理和作业管理。
b. 数据库管理系统;
c. 编译系统。
2. **应用软件**:
 - ① 可以直接给用户使用;
 - ② 办公软件、多媒体处理软件、Internet 工具软件。

第五章 多媒体技术的简介

§ 5.1 多媒体概念以及特征 (不考? 了解一下即可)

1. **媒体**: 图像、声音、视频、动画
2. **多媒体**: 能够对两种以上的媒体进行采集。
3. **多媒体的特征**: 交互性和集成性是核心 (了解四种性质名称即可不用展开)
 - ① 交互性: 具有人机交互功能;
 - ② 集成性: 集成多种媒体技术及获取、存储;
 - ③ 多样性: 媒体传播、展示手段等的多样化;
 - ④ 实时性: 声音和活动的视频图像等是强实时的。

§ 5.2 多媒体数字化之声音

1. **声音**: 模拟信号, 数字信号 (二者通过采样、量化、编码进行转变)。
2. **采样**:
 - ① 按照一定速率测得模拟信号的模拟值;
 - ② 采样率: 22khz、44khz、48khz、fm 广播音质、CD 音质、DVD 音质
3. **量化**:
 - ① 给模拟信号中取得指定量
 - ② 量化精度: 8 位、16 位、32 位
4. **文件大小**:

文件大小=采样时间 (s) × 采样率 (Hz) × 量化精度 (b) × 声道数/8

- 注：单声道代表乘以 1，双声道代表乘以 2；
- 注：量化精度的单位为位 (bit)，必须转化为字节 (Byte)。
- 注：在计算过程中一般认为 $1024 \approx 1000$ 。

§ 5.3 多媒体数字化之图像

1. 图像:

- ① 静态图像：矢量图像，点位图图像；
- ② 动态图像：视频，动画。

2. 常见图像文件格式:

BMP (Windows 采用的图像格式)

JPG (采用 jpeg 标准压缩过的图像格式)

PNG (图像背景可以是透明的支持矢量图格式)

BMP (.bmp, 标准Windows图像格式)
GIF (.gif, 使用LZW压缩算法, 支持多画面循环显示)
TIFF (.tiff, 位图图像格式)
PNG (.png, 保留GIF文件的一些特性如: 流式读/写性能、透明性、无损压缩等, 同时增加了一些新特性)
WMF (.wmf, 剪贴画)

3. 点位图:

- ① 用图片中的每个点的颜色来描述图片。
- ② 拍摄的照片一定是点位图。
- ③ 放大后变模糊。

4. 矢量图:

- ① 放大后不会变模糊。
- ② 用绘图软件制作的才有可能是矢量图。

5. 像素: 图片中的每一个点。

第六章 计算机病毒及其防治

§ 6.1 计算机病毒的特征和分类

1. 计算机病毒的含义: 人为编写的一段程序代码或是指令集合，能够通过复制不断感染其他程序、文件等。

2. 病毒:

- ① 本质是一段程序代码；
- ② 普通程序通常有独立的程序文件，病毒程序寄生在其他文件上；
- ③ 特性：
 - a. 寄生性：寄生在可执行程序中，享有该程序的一切权利。
 - b. 传染性：病毒可以自己扩散到其他文件；
 - c. 破坏性：破坏系统功能、删除重要文件；
 - d. 传播途径：互联网、U 盘；
 - e. 潜伏性：病毒程序往往不会马上发作；
 - f. 隐蔽性：有些病毒即使发作，你也注意不到它的存在。
- ④ 防治方法：安装杀毒软件（及时更新和升级）。

第七章 Internet 基础及其应用

§ 7.1 计算机网络的基本概念

1. 网络的本质:

- ① 资源共享: 静 (访问彼此资源的能力) 共享文件、共享打印机、访问站点。
- ② 快速通信: 动 (秒杀以往一切通信手段) 极大地提高了人们沟通的便利性。

2. 数据通信: (弹幕说书上没有?)

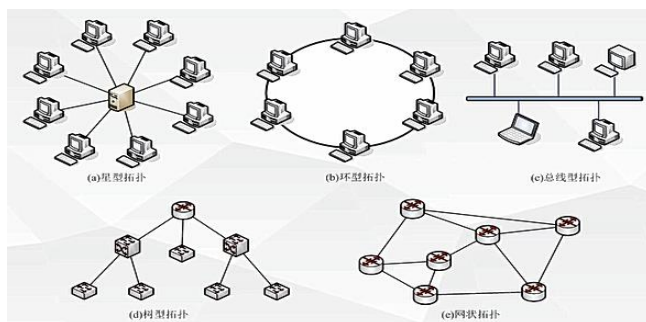
- ① 信道: 信息传输的媒介或渠道。有线、无线。
- ② 调制与解调: 数字脉冲信号转换成模拟信号的过程称为调制 (Modulation); 将接收端模拟信号还原成数字脉冲信号的过程称为解调。将调制和解调两种功能结合在一起的设备称为调制解调器。
- ③ 带宽与传输速率: 信号的最高频率和最低频率之差表示, 即频率的范围。单位: **bps/s** 即二进制位/秒。
- ④ 误码率: 二进制比特在数据传输系统中被传错的概率, 是通信系统的可靠性指标。

3. 计算机网络的分类:

- ① 局域网 (Local Area Network, **LAN**): 是一种在有限区域内使用的网络, 其传送距离一般在几公里之内, 因此适用于一个部门或一个单位组建的网络。
- ② 城域网 (Metropolitan Area Network, **MAN**): 局域网的传输可靠、误码率低, 结构简单, 容易实现。
- ③ 广域网 (Wide Area Network, **WAN**): 又称远程网, 覆盖的地理范围要比局域网大的多, 从几十公里到几千公里。

4. 网络拓扑结构:

- ① 星型拓扑: 网络风险集中在中心节点上;
- ② 环型拓扑: 每一个节点都存在稳定风险;
- ③ 总线型拓扑: 搭建容易, 成本最低;
- ④ 网状拓扑: 部署成本高, 网络稳定性好;
- ⑤ 树型拓扑: 可满足汇集信息的应用要求, 成本低, 易扩建推广。



• 总结: 星型、总线型拓扑适用于小型局域网; 网络拓扑适用于大型广域网、因特网。

5. 网络硬件:

- ① 传输介质: 同轴电缆、双绞线和光缆;
- ② **网络接口卡: 网卡;**
- ③ 交换机: 支持多端口并发连接的集线设备无线;
- ④ AP: 无线访问点或网络桥接器;
- ⑤ **路由器:** 实现局域网与广域网互联的主要设备。

6. 以太网: 一种可用于搭建局域网的技术标准。网线的水晶头就属于以太网标准的一种。

7. 网卡: 网络适配器、网络接口卡、以太网卡。网卡是我们联网最基本的硬件设备。

8. **拓扑和网速:**
 - ① 拓扑: 以太网内部采用什么结构;
 - ② 网速: 以太网可以达到多快的速度。

§ 7.2 Internet 基础

1. **因特网:** 核心问题是如何整合全球的局域网。
2. **TCP/IP 协议:** 因特网的技术基础。
3. **IP 协议:** IP 地址、路由选择
IP 地址格式: 一共四个字节, 每一个字节可以表示的数据范围 0-255
4. **域名:**
 - ① 主机名. 二级域名. 一级域名。
 - ② 比 IP 地址更加易于记忆。

举例: 百度——域名: www.baidu.com; IP 地址: 119.75.218.77
www 主机名, baidu 二级域名, com 一级域名

主要用途	一级域名	中国域名
商业公司	.com	.com.cn
教育机构	.edu	.edu.cn
政府机关	.gov	.gov.cn
军事部门	.mil	.mil.cn
其他机构	.org	.org.cn (★)

5. **TCP 协议:** 传输层协议
6. **其他协议:** 基于 TCP 产生的应用协议。
远程登录 Telnet; 邮件传输 SMTP; 文件传输 FTP; 超文本传输 HTTP。

§ 7.3 Internet 应用

1. **网上漫游:**
 - ① 关于 Internet 的几个概念:
 - a. 万维网 (World Wide Web);
 - b. 超文本和超链接;
 - c. 统一资源定位器 (Uniform Resource Locator, URL);
 - d. 浏览器;
 - e. **FTP 文件传输协议**, 位于应用层, 因特网提供的最基本的服务。
 - ② 浏览网页
2. **电子邮件**