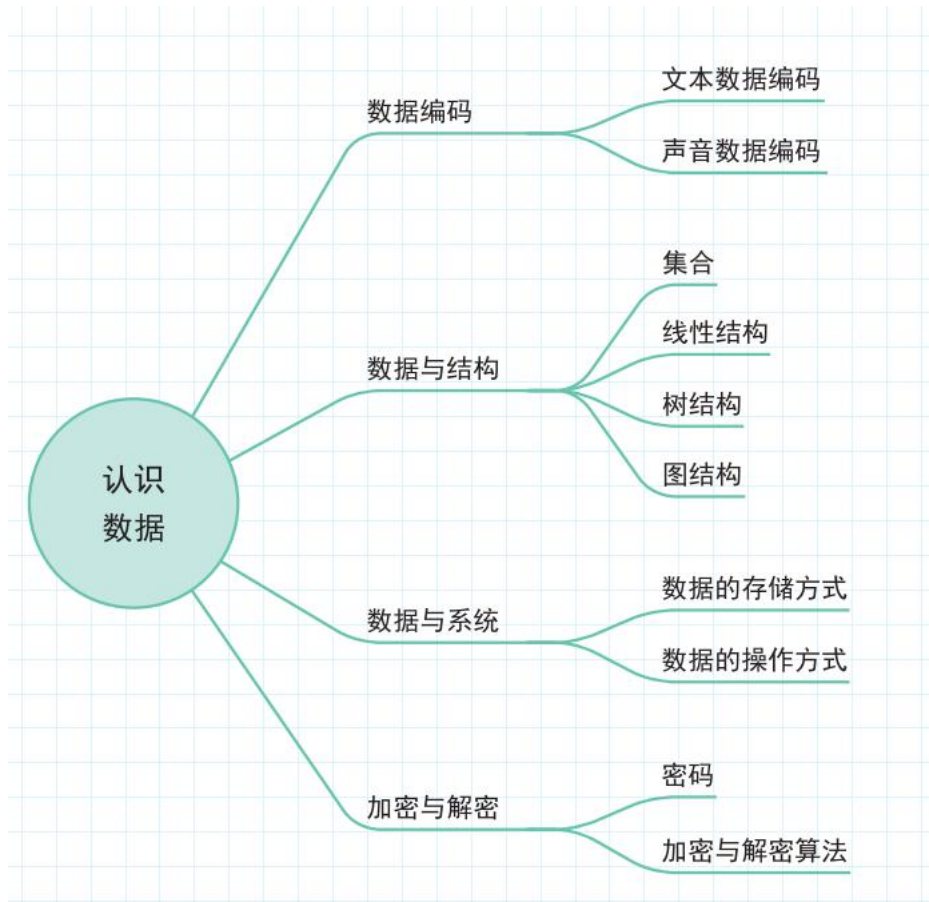


必修1 第三单元：认识数据（知识梳理）



►►2. 1 数据采集与编码 ◀◀◀

【知识链接】

1.采样和量化

采样：根据输入的模拟信号按适当的时间间隔得到各个时刻的样本值。

量化：把经采样得到的各个时刻的样本值用二进制码来表示。

2.模拟信号和数字信号

按照取值特征，可以将信号分为模拟信号和数字信号。

模拟信号是指用连续变化的物理量所表达的信息，如温度、湿

度、压等强。我们通常又把模拟信号称为连续信号，它在一定的时间范围内可以有多个不同的取值。模拟信号的值是随时间连续变化的，波形光滑。

数字信号随时间的变化是非连续的，即取值上是离散的、不连续的信号。数字信号可以由模拟信号转换得来，即在模拟信号的基础上经过采样、量化和编码而形成。

3.声音的数字化

声音信号是通过采样和量化实现模拟信号的数字化的，这个过程称为模数转换。

音频采样：按一定的时间间隔，测得模拟信号的模拟值。每秒采样的次数称为采样频率，单位是赫兹(Hz)。如 CD 的采样频率为 44.1 kHz，即每秒采样 44100 次。

音频量化：将信号波形的纵轴划分为若干个区间，把落到某个区间的采样点按四舍五入的规则取整数值。

4.容量单位

计算机中的信息是使用二进制数（即 0 和 1）来表示的；计算机中存储数据的最小单位是二进制位，用比特(bit)表示；计算机存

储数据的基本单位为字节(Byte)， 8 比特构成 1 个字节。

存储容址单位还有 KB、MB、GB、TB 等，它们之间的转换关系如下：

1 Byte = 8 bits

1 KB = 1024 Bytes

1 MB = 1024 KB

1 GB = 1024 MB

1 TB = 1024 GB

5.数制与数制转换

数制是各种计数方法或计数表示方法的总称，包括非进位数制和进位数制两种。

相邻位之间存在 " 逢十进一，借一当十 " 的关系，称为十进制数制。计算机中的数制均为进位数制。常见的有二进制、八进制，十六进制等。

常用数制的基本特征如下表所示

数制	基数	进位规则	位权	数码	表示
十进制	10	逢十进一	10'	0,1, 2, 3, 4 , 5,6,7,8,9	D
二进制	2	逢二进一	2'	0,1	B
八进制	8	逢八进一	8'	0,1, 2, 3, 4 , 5, 6, 7	O
十六进制	16	逢十六进一	16'	0,1, 2;… , 8,9, A, B,C, D, E, F	H

数制转换：

- (1) 把十进制的非负整数转换成 R 进制数 (R 大于等于 2)
采用“除 R 取余法”,即用整数部分不断去除以 R,并记下每次的余数,直到商为 0 为止,反向取余即为最终的转换结果;
- (2) 把 R 进制非负整数转换成十进制数可用“按权展开求和法”。

6.文本数据的编码

计算机要处理文本,就必须将其转换为计算机能识别的二进制数才能处理,该过程称为文本数据编码。常见的文本数据的编码方案有 ASCII 码和 Unicode 码。

ASCII (American Standard Code fo r Information Interchange) ,即美国信息交换标准码。用 7 位 (标准 ASCII 码) 或 8 位 (扩展 ASCII 码) 的二进制数表示一个字符,共有 128 个字符,每个字符用一个字节来存储,具体如下表所示。

ASCTI 码 (十进制)	表示的字符
0~ 31	控制字符或通信专用字符 (不可见)
32	空 格
48 ~ 57	数字 0~ 9
65~ 90	大写英文字母 A~ Z
97~ 122	小 写英文字母 a~ z
其 他	英文标点符号、运算符号、  等

Unicode 是为了解决传统的字符编码方案的局限而产生的,它为每种语言中的每个字符设定了统一并且唯一的二进制编码,以满足跨

语言、跨平台进行文本转换、处理的需求。Unicode 通常用两个字节表示一个字符，原有的英文编码要从单字节变成双字节，只需把高位字节全部填 0 即可。

▶▶▶ 2.2 数据的组织 ◀◀◀

【知识链接】

1. 数据结构

数据结构是存在特定关系的数据元素的集合。常见的数据结构有集合结构、线性结构、树结构和图结构等。

2. 数据结构类型

数据结构有线性结构与非线性结构之分。

在线性数据结构中，除首元素没有前趋元素、尾元素没有后继元素外，其他元素都只有一个前趋元素和一个后继元素，如队列。非线性数据结构有树和图等。

3. 队列

队列是一种先进先出(FIFO)的线性表，它的数据元素只能在队列的一端依次添加（入队），在另一端依次删除（出队）。

4. 树结构

树结构是一种具有层次关系的非线性结构。在树结构中，每一个节点只有一个前趋节点，称为父节点，没有前趋的节点只有一个，称为树的根节点。

每一个节点可以有多个后继节点，它们称为该节点的子节点。没有后继的节点称为叶子节点。在树结构中，数据元素之间是一对多的关系。

5. 图结构

图结构是由一组节点（称为顶点）和一组节点间的连线（称为边或弧）构成的一种非线性数据结构。图结构中的每个顶点都可以与其他顶点有边相连，图结构中数据元素之间是多对多的关系。

队列、树结构和图结构的关系如下表所示。

结构类型	数据（节点）之间的关系	生活中相应结构应用举例
队列（线性）	一对一	上车、过马路、付款时排的队、医院就诊时电子牌上的就诊队列
树结构	一对多	行政区划、书的目录结构、磁盘文件存储结构、注册表结构
图结构	多对多	全国航运图、铁路运输图、高速公路网

▶▶▶ 2.3 数据处理 ◀◀◀

【知识链接】

1、图形用户界面（GUI）

图形用户界面（GUI）又称图形用户接口，是指采用图形方式显示的计算机操作用户界面。在 Python 语言中，tkinter 是实现图形用

户界面的模块,可以用来快速创建 GUI 应用程序。例如,导入 tkinter 模块后,就可以运用代码创建矩形容器(Frame)、文本框(Entry) 命令按钮(Button)、文本标签(Label) 等交互对象。

2. 文件: 存储在外存储设备中的相关数据的集合。

3. 在 Python 语言中,用 open 函数把外存储器里的文件打开 ; 用 write 方法将内存中的数据存储到外存储器的文件中; 文件操作结束时,用 close 方法关闭文件。

4.数据库是以一定的组织方式存储在计算机中的相互关联的数据集合。

在数据库中,数据是以二维表的形式组织存储的,称为表。表中的一列称为一个字段,每个字段都有一个名字,称为字段名;表中的一行称为一条记录。

5.数据库管理系统(DBMS)是定义、创建、维护数据库的一种工具,它允许用户管理和控制数据库中的数据。常见的数据库管理系统有 Sybase、ORACLE、MySQL、Access 等。

6.数据库系统是由数据库、数据库管理系统(及其应用开发工具)、数据库应用系统、数据库管理员和用户组成的存储、管理、处理和维
护数据的系统。

7.利用 Python 访问 Access 数据库一般可分为以下 5 个步骤:

- (1) 建立与数据库的连接;
- (2) 根据需要产生记录集;
- (3) 访问记录集处理数据;
- (4) 根据需要把处理好的数据更新到数据库中;
- (5) 断开与数据库的连接。

▶▶▶ 2.4 数据的保护 ◀◀◀

【知识链接】

1.数据安全的威胁有人为因素,也有非人为因素。主要的威胁有计算机病毒、黑客攻击、数据存储介质损坏和个人失误等。

2.数据安全有数据备份和数据加密两方面的含义。

3.数据备份是将需要备份的数据从应用主机的硬盘或磁盘阵列复制到其他的存储介质或不同位置存储空间的过程。其目的是在设备发生故障或发生其他威胁数据安全的灾害后,利用备份进行恢复,从而达到保护数据的目的。常见的数据备份方法有可移动存储设备备份和网络备份等。

4.云备份的优点是成本低、管理方便和可扩展性高,缺点是安全性不足等。

5.密码是指用来核对用户 ID 以验证用户是否是本人的一组字符。

6.一般来说,创建安全密码有以下几个技巧:

- (1) 使用长度不少于 8 个字符的密码,密码越长越不容易被破

解。

(2) 在可能的情况下，尽量使用字母、数字和特殊符号相结合的密码。

(3) 不要使用电话号码、身份证号或生日作为密码。

(4) 不要使用整个用户 1D 或用户 ID 的一部分作为密码。

(5) 不要使用字典中能找到的词语作为密码，即使是字母顺序颠倒过来的常用词语也不可以。

(6) 加密就是将原始信息（数据）隐匿起，来使之在缺少特殊信息（数据）时不可读。原始 信息（数据）称为明，文加密后的信息（数据）称为密文。将密文还原成明文的过程称为解密（或解码。）

(7) **加法密码**又称为**移位密码**。在加法密码算法中，明文中的所有字母都在字母表上向后或向前按照一个固定数目进行偏移后被替换成密文。

恺撒密码只是简单地将明文中的每一个字母用字母表中该字母后的第 3 个字母替换。例如，将明文中的 A 用 D 替换，B 用 E 替换，……，以此类推，X 用 A 替换，Y 用 B 替换，Z 用 C 替换。像恺撒密码这样，明文中的所有字母都在字母表上向后（或向前）按照一个固定数目进行偏移后被替换成密文，这种加密方法使用的密码就是加法密码。