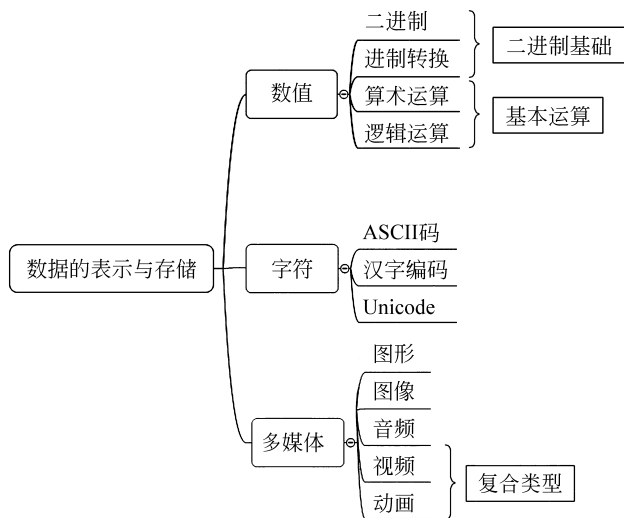


数据的表示与存储

3.1 数据的表示与存储简介



3.1.1 数据在计算机中的表示

1. 数制

1) 数制

数制也称“记数制”，是用一组固定的符号和统一的规则来表示数值的方法。

实际上，在讨论数制的时候都会涉及两个基本要素：基数和位权。

2) 基数

基数是指在某种进位记数制中，每个数位上能够使用数字的个数。例如，二进制的基数为 2，每个数位上能够使用的数字为 0、1；十进制的基数为 10，每个数位上能够使用的数字为 0~9。

不同进制数的表示方法：二进制用 B 表示，八进制用 O 表示，十进制用 D 表示，十六进制用 H 表示。

(1) 字母后缀。

例如,二进制数 1010 表示为 1010B,八进制数 23.45 表示为 23.45O,十进制数 123.45 表示为 123.45D,十六进制数 A1.23 表示为 A1.23H。一般来说,对于十进制数的后缀可以省略。

(2) 括号外面加下标。

例如,上述数值可表示为 $(1010)_2$ 、 $(23.45)_8$ 、 $(123.45)_{10}$ 、 $(A1.23)_{16}$ 。

注意: 以下内容中约定,十进制数按平时习惯均不作特别标注。

3) 位权

位权是指一个数字在某个固定位置上所代表的值,处在不同位置上的数字代表的值不同。例如,十进制数 123,1 的位权是 100,2 的位权是 10,3 的位权是 1。位权与基数的关系:各进制中位权的值是基数的对应位次幂。位幂次的排列方式以小数点为界,整数自右向左,最低位为基数的 0 次幂;小数自左向右,最高位为基数的 -1 次幂。任何一种数制表示的数都可以写成如下按位权展开的多项式之和。

例 3.1 把 123.45、1010.11B、23.45O、A1.23H 按位权展开。

$$123.45 = 1 \times 10^2 + 2 \times 10^1 + 3 \times 10^0 + 4 \times 10^{-1} + 5 \times 10^{-2}$$

$$(1010.11)_2 = 1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 0 \times 2^0 + 1 \times 2^{-1} + 1 \times 2^{-2}$$

$$(23.45)_8 = 2 \times 8^1 + 3 \times 8^0 + 4 \times 8^{-1} + 5 \times 8^{-2}$$

$$(A1.23)_{16} = 10 \times 16^1 + 1 \times 16^0 + 2 \times 16^{-1} + 3 \times 16^{-2}$$

2. 进制转换

1) 二进制数、八进制数、十六进制数转换成十进制数

二进制、八进制、十六进制数转换成十进制数采用上述介绍的按位权展开方法,把各项相加即可。

例 3.2 将 $(101.01)_2$ 、 $(24.4)_8$ 、 $(35.C)_{16}$ 转换成十进制。

$$(101.01)_2 = 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 + 0 \times 2^{-1} + 1 \times 2^{-2} = 4 + 0 + 1 + 0 + 0.25 = 5.25$$

$$(24.4)_8 = 2 \times 8^1 + 4 \times 8^0 + 4 \times 8^{-1} = 16 + 4 + 0.5 = 20.5$$

$$(35.C)_{16} = 3 \times 16^1 + 5 \times 16^0 + 12 \times 16^{-1} = 48 + 5 + 0.75 = 53.75$$

2) 十进制数转换成二进制数、八进制数、十六进制数

(1) 整数部分:除以基数取余数法。

(2) 小数部分:乘以基数取整数法。

注意: 当包含小数的数值要转换成不同进制数时,很多情况下是无法完全实现精确转换的。

3) 二进制数、八进制数、十六进制数之间的转换方法

二进制数转换成八进制数的方法:以小数点为界向左右两边进行分组,每 3 位为一组,不足 3 位就用 0 补足,每组用一个八进制数表示即可,简称“三合一”。同样,二进制数转换成十六进制数的方法就:以小数点为界向左右两边进行分组,每 4 位为一组,不足 4 位就用 0 补足,每组用一个十六进制数表示即可,简称“四合一”。

例 3.3 将二进制数 $(10110101.10101)_2$ 转换成八进制数和十六进制数。

$$\begin{array}{ccccccc} \underline{010} & \underline{110} & \underline{101} & \underline{101} & \underline{010} & & \\ 2 & 6 & 5 & 5 & 2 & & \end{array} (265.52)_8 \text{ (整数高位和小数低位分别补 0)}$$

$$\begin{array}{ccccccc} \underline{1011} & \underline{0101} & \underline{1010} & \underline{1000} & & & \\ B & 5 & A & 8 & & & \end{array} (B5.A8)_{16} \text{ (小数低位补 0)}$$

同样,将八进制数转换成二进制数时,只需把每位八进制数拆分为 3 位二进制数,简称“一拆三”;将十六进制数转换成二进制数时,只需把每位十六进制数拆分为 4 位二进制数,简称“一拆四”。当需要将八进制数和十六进制数相互转换时,可以先将其转换为二进制数,再将这二进制数转换成对应的八(十六)进制数即可。

3. 数值在计算机中的表示与存储

1) 机器数和真值

在计算机中,对于数学上的正(+)、负(-)号也只能使用二进制中的 0 和 1 两个数码表示。规定用 0 表示正,1 表示负,因为计算机存储信息时是以字节为单位,符号就存放在该字节的最高位,称为符号位,也称数符。

例 3.4 二进制数 -110011 在计算机中表示为 8 位二进制数: 10110011。

这种把符号数值化了的数称为机器数,而它代表的数值称为该机器数的真值。在例 3.4 中,10110011 为机器数, -110011 为该机器数的真值。

(1) 原码。

原码的表示规则: 机器数的最高位表示符号位,正数的符号位为 0,负数的符号位为 1,其余各位是数值的绝对值。

(2) 反码。

反码的表示规则: 正数的反码就是其原码,负数的反码是符号位为 1,其余各位按位取反。

(3) 补码。

补码的表示规则: 正数的补码就是其原码,负数的补码是其反码在最低位加 1。

利用补码方法可以很方便地实现正负数的加法运算,规则简单。只要参与运算的数据是在数值表示范围内,符号位同数值位一样参与运算而不需要单独考虑。如果运算的结果为正数,结果的真值就是其本身;否则,就对结果再次求补即可得到其真值。且允许产生最高位的进位(丢弃)。

若一个 R 进制中的两个数 a 、 b 之和等于 R ,则称 a 和 b 互为补数(补码), R 称为模。模 R 的取值是指定数值范围中包含的整数个数。计算机中的数据都是采用定长方式存储,8 位二进制能表示的整数有 $2^8=256$, $R=256$,推广到 n 位二进制,即 $R=2^n$ 。以 1 字节为例,利用模 R 求补码的方法,能够在二进制和十进制中实现减法运算转换为加法运算。

例 3.5 计算 $105-83$ 的运算。

方法一,采用十进制方法, $2^8=256$,所以 $R=256$ 。按照公式 $R=a+b$,已知 $a=83$,求得 $b=R-83=256-83=173$ 。

所以 $105 - 83$ 可转换为 $105 + 173 = 278$, 丢弃模 256, $278 - 256 = 22$ 得出的结果正确。

方法二, 采用二进制补码方法。

$$\begin{array}{r}
 01101001 \quad \cdots \cdots \text{105 的补码} \\
 + \quad 10101101 \quad \cdots \cdots \text{-83 的补码} \\
 \hline
 \boxed{1} 00010110 \quad \cdots \cdots \text{运算结果}
 \end{array}$$

丢弃高位进位 1, 运算结果 $[0001 0110]_{\text{补}}$ 的符号位是 0, 说明结果是正数, 其真值就是 $(10110)_2$, 转换为十进制就是 22, 结果正确。

由此可见, 在指定范围内, 用补码方法实现减法运算转换成加法运算不但适用于二进制, 在其他进制中也同样适用。

2) 实数在计算机中的表示

(1) 定点整数。

定点整数规定小数点的位置在最低位的右边, 这种方法表示的数为纯整数。

(2) 定点小数。

定点小数规定小数点的位置在符号位的右边, 这种方法表示的数为纯小数。

(3) 浮点小数。

浮点数由阶码和尾数组成: 阶码用定点整数表示, 其所占的位数确定了数值的范围; 尾数用定点小数表示, 其所占的位数确定了数值精度, 即小数点后有效位数。因此, 实数就是用浮点数来表示和存储的。浮点数表示如图 3.1 所示。

阶符	阶码	尾符	尾数
----	----	----	----

图 3.1 浮点数表示

为了唯一地表示浮点数在计算机的存储, 对尾数采用了规格化的处理, 规定尾数的最高位为 1, 即所有规格化数必须转换成 $\pm 0.1 \times \times \times \cdots \times \times \times \times 2^{\pm p}$ 形式, 对于不符合要求的可以通过阶码 p 进行调整。

IEEE 在 1985 年制定了 IEEE 754 标准, 统一浮点数的存储格式。因此, 在程序设计语言中比较常见的有以下两种类型的浮点数。

① 单精度浮点数(Float 或 Single)占 4 字节, 其中阶符占 1 位, 阶码占 7 位, 尾符占 1 位, 尾数占 23 位。

② 双精度浮点数(Double)占 8 字节, 其中阶符占 1 位, 阶码占 10 位, 尾符占 1 位, 尾数占 52 位。

可见双精度浮点数比单精度浮点数占用的存储空间更大, 表示的数值范围更广, 且精度更高。

例 3.6 36.5 采用单精度浮点数在计算机中的存储形式如图 3.2 所示。

0	0000110	0	1001001000...0000000
---	---------	---	----------------------

图 3.2 单精度浮点数存储格式

规格化表示： $36.5=100100.1=0.1001001\times 2^6=0.1001001\times 2^{10}B$

3.1.2 字符的编码

1. ASCII 码

对于英文字符的编码,最常用的是美国信息交换标准代码字符编码(American Standard Code for Information Interchange, ASCII)。它是由美国国家标准学会(American National Standards Institute, ANSI)制定的,供不同计算机在相互通信时共同使用,后来它被国际标准化组织(International Organization for Standardization, ISO)定为国际标准,称为 ISO 646 标准。ASCII 码采用 7 位二进制编码,共有 $2^7=128$ 种不同的组合,可以表示 128 个字符,包括 10 个数字字符 0~9、26 个大小写的英文字母以及特殊字符和控制字符。

英文字符除了常用的 ASCII 编码外,还有 BCD 码(Binary-Coded Decimal),它将十进制数的每位分别用 4 位二进制数表示,又称二-十进制编码;还有另一种 EBCDIC 码(Extended Binary Coded Decimal Interchange Code),即扩展的二-十进制交换码,这种编码主要在大型机器中使用。

2. 汉字编码

GB 2312—1980《信息交换用汉字编码字符集 基本集》收集和定义了 6763 个汉字及 682 个拉丁字母、俄文字母、汉语拼音字母、数字和常用符号等,共 7445 个汉字和字符。其中使用频度较高的 3755 个汉字为一级汉字,按汉字拼音字母顺序排列,使用频度较低的 3008 个汉字为二级汉字,按部首排列。

计算机汉字处理流程如下(见图 3.3)。

- (1) 将汉字以输入码方式输入计算机中。
- (2) 将输入码转换成计算机能够识别的汉字机内码进行处理、存储。
- (3) 将机内码转换成汉字字形码输出。

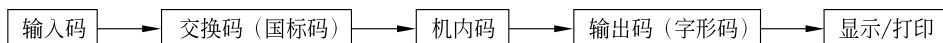


图 3.3 汉字处理流程

1) 汉字输入码

汉字输入码是指从键盘上输入汉字时采用的编码,又称外码。

2) 汉字交换码(国标码)和机内码

1980 年,我国颁布了汉字编码的国家标准 GB 2312—1980,这个字符集是我国中文信息处理技术的发展基础,也是目前国内所有汉字系统的统一标准,GB 2312—1980 称为国标码。GB 2312—1980 规定每个汉字用 2 字节的二进制编码,每字节的最高位为 0,其余 7 位用来表示汉字信息。为保证中、英文字符兼容,在计算机内部能区分 ASCII 字符和汉字,将汉字国标码的 2 字节二进制编码的最高位置 1,从而得到对应的汉字机内码。汉字机内码是汉字在计算机内部被存储、处理和传输时使用的编码,简称内码。

3) 汉字字形码

汉字字形码又称汉字字模,用于汉字在显示屏或打印机输出。汉字字形码通常有两种表示方式:点阵方式和矢量方式。用点阵表示汉字时,汉字字形码指的是这个汉字字形点阵的代码。矢量方式存储的是描述汉字字形的轮廓特征,当要输出汉字时,通过计算机的计算,由汉字字形描述生成所需大小和形状的汉字。

常用的点阵有 16×16 、 24×24 、 32×32 、 64×64 或者更高。以 16×16 点阵为例,每个汉字字形就要占用 32 字节($16 \times 16 / 8 = 32$ 字节),两级汉字共 7400 多汉字和字符大约占用 256KB。

3. Unicode(统一码、万国码)

Unicode 在 1992 年被国际标准化组织确定为国际标准 ISO 10646,成为可以用于表示世界上所有文字和符号的字符编码方案。目前,所有的计算机都支持 Unicode。Unicode 用一些基本的保留字符制定了 3 套编码方式,分别是 UTF-8、UTF-16 和 UTF-32,UTF 是 Unicode Transformation Format 的缩写。在 UTF-8 中,字符是以 8 位二进制即 1 字节来编码的。用 1 字节或几字节来表示 1 个字符,这种方式的最大好处是保留了 ASCII 字符的编码作为它的一部分。而其他字符,例如中文、日文、韩文等大部分的常用字,使用 3 字节编码;UTF-16 和 UTF-32 分别是 Unicode 的前 16 位和前 32 位编码方式。

3.1.3 多媒体数据表示

1. 图形与图像

图形一般指用计算机软件绘制的由直线、圆、圆弧、任意曲线等图元组成的画面,以矢量图形文件形式存储。矢量文件中存储的是一组描述各个图元的大小、位置、形状、颜色、维数等属性的指令集合,通过相应的绘图软件读取这些指令可将其转换为输出设备上显示的图形。因此,矢量图文件的最大优点是对图形中的各个图元进行缩放、移动、旋转而不失真,且占用的存储空间小。

图像则是指由输入设备(如扫描仪、数码相机等)捕捉的实际场景画面,经数字化后以位图形式存储的画面。位图文件中存储的是构成图像的每个像素点的亮度、颜色,位图文件的大小与分辨率和色彩的颜色种类有关,放大、缩小会失真,占用的空间比矢量文件大。

1) 数字化

(1) 采样。

采样就是将二维空间上连续的图像转换成离散点的过程,采样的实质就是用多少个像素(Pixels)点来描述这一幅图像,称为图像的分辨率,用“列数 \times 行数”表示,分辨率越高,图像越清晰,存储量也就越大。

(2) 量化。

量化是在图像离散化后,将表示图像色彩浓淡的连续变化值离散化为整数值的過程。把量化时所确定的整数取值个数称为量化级数,表示量化的色彩值(或亮度)所需的二进制位数称为量化字长。一般可用 8 位、16 位、24 位或以上来表示图像的颜色,24 位可

以表示 $2^{24}=16\ 777\ 216$ 种颜色,称为真彩色。量化字长越大,越能真实地反映原有图像的颜色,但得到的数字图像的容量也越大。

在多媒体计算机中,图像的色彩值称为图像的颜色深度,有3种表示方式。

① 黑白图:图像的颜色深度用一个二进制位1和0分别表示纯白、纯黑两种状态。

② 灰度图:图像的颜色深度用1字节表示,灰度级别就有 $256(2^8=256)$ 级,通过调整黑白两色的程度(称为灰度)来显示单色图像。

③ RGB(Red、Green、Blue,三原色):24位真彩色显示时,由红、绿、蓝三原色通过不同的强度混合而成。强度分为256级,占24位(3字节),就构成了 $2^{24}=16\ 777\ 216$ 种“真彩色”图像。

(3) 编码。

将采样和量化的数据转换成为二进制数的表示形式。

图像的分辨率和像素位颜色深度决定图像文件的大小,即图像的质量,其计算公式为

$$\text{列数} \times \text{行数} \times \text{颜色深度} / 8 = \text{图像字节数}$$

例 3.7 要表示一个分辨率为 1024×1024 的24位“真彩色”图像,则图像大小为

$$1024 \times 1024 \times 24 / 8 = 3\text{MB}$$

这仅仅是一张数字化后的图像存储容量,可见其数据量巨大,无论是存储还是传输都不方便,必须采用压缩技术来解决。

2) 常见图形和图像文件格式

(1) BMP文件: BMP(Bitmap)是一种与设备无关的图像文件格式,是Windows环境中经常使用的一种位图格式。其特点是包含的图像信息较丰富,几乎不进行压缩,故文件占用空间较大。

(2) JPG文件: JPEG是由联合照片专家组(Joint Photographic Experts Group)开发的。它既是一种文件格式,又是一种压缩技术。它用有损压缩方式去除冗余的图像和彩色数据,在获得极高压缩率的同时也能展现十分丰富生动的图像。JPEG 2000作为JPEG的升级版,其压缩率比JPEG高约30%,同时支持有损和无损压缩,且向下兼容,因此可取代传统的JPEG格式。

(3) GIF文件: GIF(Graphic Interchange Format)采用了压缩存储技术,最多支持256种色彩的图像。其特点是压缩比高,磁盘空间占用较少、下载速度快、可以存储简单的动画,被广泛用于Internet中。

(4) PNG文件: PNG(Portable Network Graphics)的压缩比高,并且是无损压缩,适合在网络中传播。但是它不支持动画功能。

(5) WMF文件: WMF(Windows Metafile Format)是Windows中常见的一种图元文件格式,它具有文件短小、图案造型化的特点,整个图形常由各个独立的组成部分拼接而成,但其图形往往较粗糙。Windows中许多剪贴画图像是以该格式存储,广泛应用于桌面出版印刷领域。

(6) SVG文件: SVG(Scalable Vector Graphics)是一种基于XML、由World Wide Web Consortium(W3C)联盟开发的,开放、标准的矢量图形文件。它可以使图像在改变

尺寸的情况下,图形质量不会有损失,与 JPEG 和 GIF 图像比起来,其尺寸更小,可压缩性更强,方便下载,是目前比较流行的图像文件格式。

2. 音频

音频是声音的信息表示,通常指在 20~20 000Hz 频率范围的声音信号,是连续变化的模拟信号,而计算机只能处理数字信号,必须把它转换成数字信号计算机才能处理,这就是音频的数字化。

1) 音频的数字化

音频的数字化过程要经过采样、量化和编码。采样和量化的过程可由模数转换器(Analog-to-Digital Converter, ADC)实现。ADC 以固定的频率去采样、量化,经采样和量化的声音信号再经编码后就成为数字音频信号,以数字声波文件形式保存在计算机存储介质中。若要将数字声音输出,则通过数模转换器(Digital-to-Analog Converter, DAC)将数字信号转换成原始的模拟信号即可。

模拟波形声音经数字化后,其音频文件的存储容量(未经压缩)计算公式为

$$\text{采样频率(Hz)} \times \text{量化位数(b)} / 8 \times \text{声道数} \times \text{时间(s)} = \text{存储容量}$$

例 3.8 用 44.1kHz 的采样频率进行采样,量化位数为 16 位,则在录制 1s 的立体声节目时,其 WAV 文件所需的存储容量。

根据上述计算公式: $44.1 \times 1000 \times 16 / 8 \times 2 \times 1 \text{B/s} = 176\ 400 \text{B/s}$

2) 音频文件格式

在多媒体音频技术中存储声音信息的文件格式有多种,常见的有以下 3 种。

(1) WAV 文件: WAV(Waveform Extension, 波形扩展)是微软公司开发的一种音频文件格式,用于保存 Windows 平台的音频信息资源。WAV 文件直接记录了真实声音的二进制采样数据,被称为无损的音乐,但通常文件较大,多用于存储简短的声音。

(2) MIDI 文件: MIDI(Musical Instrument Digital Interface, 乐器数字化接口)是为了把电子乐器与计算机相连而制定的规范,是数字音乐的国际标准。MIDI 标准规定了各种音调的混合及发音,通过输出装置可以将这些数字重新合成为音乐。近年来,国外流行的声卡普遍采用波表法进行音乐合成,使 MIDI 的音乐质量大大提高。

(3) MP3 文件: MP3(Moving Picture Experts Group Audio Layer III, 动态影像专家压缩标准音频层面 3)是一种音频压缩技术。利用该技术,将音乐以 1:10 甚至 1:12 的压缩比,压缩成存储容量较小的文件,而对于大多数用户,重放的音质与最初的不压缩音频相比没有明显的下降,非常适合网上传播,是当前使用最多的音频格式文件。

上述的 WAV 和 MIDI 文件格式均可以压缩成为 MP3 文件格式。

3. 视频

视频是将一幅幅独立图像组成的序列按一定的速率连续播放,利用人的视觉暂留特征形成连续运动的画面。模拟视频的数字化过程需要先采样,将模拟视频的内容进行分解,得到每个像素点的色彩组成,然后采用固定采样率进行采样、量化、编码,生成数字化视频并以文件形式存储在磁盘上,这一过程一般通过视频采集卡完成。

1) 视频信息常用的参数

(1) 帧速: 每秒播放的静止画面数, 用帧/秒(f/s)表示。PAL 制式为 25f/s, NTSC 制式为 30f/s。

(2) 数据量: 未压缩的每帧图像数据量乘以帧速。

(3) 画面质量: 与原始图像和视频数据压缩比有关, 压缩比越高, 数据量则越小, 图像质量就越差。

2) 常见的视频文件格式

(1) AVI 文件: AVI(Audio Video Interleaved)文件允许视频和音频交错在一起同步播放, 是较为常见的视频文件格式, 但数据量较大。

(2) MPEG 文件: MPEG(Moving Pictures Experts Group)格式是 PC 上全屏幕活动视频的标准文件格式, 使用 MPEG 技术进行压缩的全运动视频图像, 数据量较小。MPEG 的平均压缩比为 50 : 1, 最高可达 200 : 1。

(3) ASF 文件: ASF(Advanced Streaming Format)是一种高级流媒体格式, 以网络数据包的形式传输, 可以在 Internet 上实现实时播放。它使用 MPEG-4 压缩算法, 压缩比很高, 且图像质量很好。其特点是数据量小, 本地或网络回放、邮件下载都可以。

(4) QuickTime 文件: QuickTime 是苹果公司采用的面向桌面系统用户的低成本、全运动视频的格式, 现在软件压缩和解压中也都使用这种格式。向量化是 QuickTime 软件的压缩技术之一, 它在最高为 30f/s 下提供的视频分辨率是 320×240 , 其压缩比为 1 : 4 ~ 1 : 32。

4. 动画

动画是活动的画面, 实质上是一幅幅静态图像连续播放。这种连续画面在时间和内容上都是连续。组成动画的每个静态画面称为帧(Frame), 动画的播放速度称为帧速率, 以每秒播放的帧数描述, 用帧/秒(f/s)表示。一般情况下, 动画每秒播放 12 帧画面, 而视频每秒播放 25 帧或以上画面, 人眼睛看到的就是连续的画面。

1) 动画的表现形式

动画有两种表现形式。

一种是帧动画, 由一幅幅图像组成的连续画面, 它的运动只能是平移; 另一种是造型动画, 是对每个运动物体分别进行设计, 赋予它们各自的特征, 如物体的大小、形状、颜色、位置等。

2) 常见的动画文件格式

(1) FLI 格式: Autodesk 公司开发的较低分辨率的文件格式, 具有固定的画面尺寸 (320×200) 及 256 色的颜色分辨率。计算机可用 320×200 或 640×400 的分辨率播放。

(2) FLC 格式: Autodesk 公司开发的较高分辨率的文件格式。FLC 格式改进了 FLI 格式尺寸固定与颜色分辨率低的不足, 是一种可使用各种画面尺寸及颜色分辨率的动画格式, 可适应各种动画的需要。

(3) SWF 格式: Flash 支持的矢量动画格式。这种格式的动画在缩放时不会失真, 文件的存储容量很小, 还可以带有声音, 因此被广泛应用。

奏的音节和音符。多音轨同步播放合成音乐。MID 文件用于存储来自 MIDI 的数据,也可以存储作曲软件创作的乐谱。

(2) WAV 是一种波形扩展文件,存储真实声音的二进制采样数据,无压缩。

(3) AVI 是音频、视频交错格式,同步存储、播放音频信号和视频信号。

(4) MP4 是一种封装格式、包括视频编码和音频编码的组合。

5. 如果某种进位记数制被称为 r 进制,则 r 称为该进位记数制的()。

- A. 数制 B. 位权 C. 基数 D. 数符

【答案】 C

【解析】 进位记数制简称进制,用有限个符号(称为数符或数码)和确定的规则表示数值的方法。进制包含 3 个要素和一个规则。

(1) 数符。表示不同的数、记数顺序及大小关系。例如,八进制的数符从小到大有 0、1、2、3、4、5、6、7。

(2) 基数。确定数符的个数。如果基数为 r ,则称为 r 进制。例如,八进制就需要且仅需要 8 个数符。

(3) 位权。数值中出现的每位数符的权重,权重值和基数及这个数符的位置有关。位置值以小数点为分界,向左数位置值为 0、1、2……,向右数位置值为 -1、-2、-3……。设位置为 i ,基数为 r ,位权值等于 r^i 。如,八进制数 23.4,2 的位权 $(10)_8$ 、3 的位权 $(1)_8$ 、4 的位权 $(0.1)_8$,表示的数值大小为 $(2 \times 10 + 3 \times 1 + 4 \times 0.1)_8$ 。等价于十进制位权表示 $2 \times 8^1 + 3 \times 8^0 + 4 \times 8^{-1}$ 。

(4) 记数规则。每个数符从小到大记数一遍后就进位,即逢 r 进一。例如,八进制数 27 加 1 后,个位数就逢 8,需要进位,数值记为 30。

6. 计算机中最通用的信息交换标准编码是()。

- A. Unicode B. ASCII
C. BCD D. GB 2312—1980

【答案】 B

【解析】 无论是什么地方语言,微型计算机中都会有 ASCII 码。例如,使用简体汉字的地区,计算机系统包含 ASCII 码、GB 2312—1980 字符编码;而使用繁体汉字的地区,计算机系统包含 ASCII 码、BIG5 字符编码。

7. 在数字音频信号获取与处理过程中,下述顺序中正确的是()。

- A. 采样、D/A 转换、压缩、存储、解压缩、A/D 转换
B. A/D 转换、采样、压缩、存储、解压缩、D/A 转换
C. 采样、A/D 转换、压缩、存储、解压缩、D/A 转换
D. 采样、压缩、A/D 转换、存储、解压缩、D/A 转换

【答案】 C

【解析】 由于音频信号源是模拟信号,计算机存储、处理的是数字信号,而播放的音频信号又是模拟信号,因此计算机处理音频信号需经历从模拟音频信号转换为数字音频信号、再从数字音频信号转换为模拟音频信号的过程。数字音频信号获取与处理就要经过采样、模数(A/D)转换、压缩、存储,再到解压缩、数模(D/A)转换、播放的过程。

8. 用 48kHz 频率作音频采样,每个采样点用 32 位的精度储存,则录制 1 秒的立体声(双声道)音频,其 WAV 文件所需的储存量为()。

- A. 375KB B. 96KB C. 187.5KB D. 750KB

【答案】 A

【解析】 WAV 文件存储的是模拟音频信号经采样、量化、无压缩编码的结果。根据题意,存储量计算式为

$$(2(\text{双声道}) \times 1(\text{时长}) \times 48\,000(\text{采样频率}) \times 32(\text{采样点精度}) / 8 / 1024) \text{KB} \approx 375 \text{KB}$$

9. () 不可能是七进制数。

- A. 10000 B. 1234 C. 657 D. 1004

【答案】 C

【解析】 依据进制的定义,七进制数的基数为 7,使用 7 个数符表示数值。数符从小到大分别是 0,1,2,3,4,5,6。657 中的 7 不属于七进制的数符。该数不符合七进制数的定义。

10. () 是关于二进制错误的论述。

- A. 二进制加法运算是逢二进一
B. 二进制数只有 0 和 1 两个数码
C. 二进制数只有两位数组成
D. 二进制数整数部分各个位上的位权对应十进制数的值分别为 1, 2, 4, 8, 16, …

【答案】 C

【解析】

(1) 二进制数的基数是 2,数符只有两个,即 0、1。

(2) 虽然二进制数的基数是所有进制中最小的,但是其表达能力和其他进制是一样的。

(3) 二进制整数部分的各个位置的位权分别是 1,10,100,1000,10000,⋯,对应十进制值是 1,2,4,8,16,⋯

(4) 根据记数规则,二进制数自然是逢二进一。

11. 一般来说,要求声音的质量越高,则()。

- A. 量化级数越高和采样频率越低 B. 量化级数越低和采样频率越高
C. 量化级数越高和采样频率越高 D. 量化级数越低和采样频率越低

【答案】 C

【解析】 音频信号经采样、量化转换为数字化信号。数字化信号的保真度取决于采样频率和量化级数。采样频率越高、量化级数越高,数字音频信号越逼真。

12. 数据在计算机存储器中的表示称为(),也是指数据的逻辑结构在计算机中的表示。

- A. 数据的关系结构 B. 数据结构
C. 数据的逻辑结构 D. 数据的存储结构

【答案】 D

【解析】 由于存储器上的数据都是以字节为单位的二进制数或二进制码。因此,只有先确定数据分类(如数据类型)和存储结构(如编码格式)再存入,读取时才能依照这个存储结构和数据分类正确解读该数据项。

13. 多媒体信息从时效上可分为静态媒体和动态媒体两大类,动态媒体包括()。

- A. 音频、文本、图形和图像
- B. 文本、图形和图像
- C. 音频、视频和动画
- D. 音频、图形和图像

【答案】 C

【解析】 在多媒体信息中,将那些和时间参数无关的信息称为静态媒体,如文本、图形、图像等;将那些和时间参数有关的信息称为动态媒体,如音频、视频、动画等。

14. 对波形声音采样频率越高,存储该声音文件所需的存储空间()。

- A. 不变
- B. 越大
- C. 越小
- D. 不能确定

【答案】 B

【解析】 音频数字化结果的数据量计算公式为

$$\text{数据量} = \text{声道数} \times \text{时长}(s) \times \text{采样频率} \times \text{量化精度}$$

可见,采样频率越高,数据量越大。

15. 多媒体数据具有的特点是()。

- A. 数据量小、输入和输出复杂
- B. 数据量大、数据类型多
- C. 数据类型间区别大、数据类型少
- D. 数据类型多、数据类型间区别小

【答案】 B

【解析】 多媒体数据包括文本、图形、图像、音频、视频、动画、3D等多种数据类型,有些数据类型的数字化结果的数据量非常大,需要使用压缩、解压缩技术进行处理。

16. 模数转换器(ADC)的功能是将()。

- A. 数字量和模拟量混合处理
- B. 模拟量转换为数字量
- C. 数字转换为二进制
- D. 声音转换为声波

【答案】 B

【解析】 ADC是将模拟信号转换为数字信号的方法。在音频信号数字化的过程中,ADC就是将采样点的音强进行数字量化的过程,量化精度越高,音质越逼真。

17. 已知英文大写字母B的ASCII码是42H,那么英文大写字母G的ASCII码为十进制数()。

- A. 15
- B. 46
- C. 68
- D. 71

【答案】 D

【解析】 ASCII码中,数字、小写字母、大写字母等拥有连续的ASCII编码值。用十进制表示,数字的ASCII码值分别是48,49,⋯,57,小写字母的ASCII码值分别是97,98,⋯,122,大写字母的ASCII码值分别是65,66,⋯,90。

字母B的十六进制ASCII码值是42H,对应十进制值为66。G排在B之后第五位,

对应的十进制 ASCII 码值是 71。

18. 以下不同进制的 4 个数中,最大的一个数是()。

A. $(19)_D$

B. $(11100011)_B$

C. $(67)_O$

D. $(CA)_H$

【答案】 B

【解析】 $(19)_D$ 、 $(67)_O$ 和 $(CA)_H$ 都是两位数,其中 $(CA)_H$ 的高位位权是 16,比另外两个数的位权大,而且 $(CA)_H$ 中的高位数码是 C,等于十进制的 12,也比另外两个数的数码值大。在这三个数中, $(CA)_H$ 的值最大。

将 $(CA)_H$ 转换为二进制数得 $(11001010)_B$,比 $(11100011)_B$ 小。

综上所述,四个数中 $(11100011)_B$ 最大。

19. 用键盘输入汉字时,在键盘上输入的是汉字的()。

A. 机内码

B. 交换码

C. 输入码

D. 字型码

【答案】 C

【解析】 在输入、存储、显示的不同阶段,汉字都有不同的码与之对应。使用输入码(即外码)输入汉字,使用机内码存储汉字,使用字形码显示汉字。1 个汉字只有 1 个机内码,但可以有多多个输入码和字形码。

20. 设汉字点阵为 32×32 ,那么 1 个汉字的字形码信息所占用的字节数是()。

A. 1280

B. 128

C. 1024

D. 32

【答案】 B

【解析】 点阵汉字存储本质是位图格式存储。考虑汉字颜色为黑白色, 32×32 点阵的汉字对应 32 行、32 列,色彩为黑白的一幅位图,1 个点的色彩用 1 位二进制表示。1 个汉字量化为字形码的字节数为 $32 \times 32 / 8$,等于 128B,1 个汉字字形码所占用的空间为 128B。

21. 多媒体技术是指利用计算机技术对()等多种存储在不同介质上的信息综合一体化,使它们建立起逻辑联系,并能进行加工处理的技术。

A. 拼音码和五笔字型

B. 硬件和软件

C. 中文、英文、日文和其他文字

D. 文本、声音、图形、图像、视频和动画

【答案】 D

【解析】 不同种类的信息,其输入介质、存储介质、处理工具、传输载体、展示媒体都有可能不同,呈现多媒体特征。多媒体技术就是研究利用计算机处理不同种类信息的方法并将其综合起来形成逻辑关联的技术。信息的种类包含文本、声音、图形、图像、视频和动画等。

22. 1 字节由 8 个二进制位组成,它所能表示的最大无符号八进制数为()。

A. 377

B. 777

C. FF

D. 255

【答案】 A

【解析】 1 字节含 8 个二进制位,能表示的最大无符号数为 11111111。根据八进制数和二进制数的转换关系,1 位八进制数对应 3 位二进制数,8 位无符号二进制数 11111111 转换为无符号八进制数的值为 377。

23. 24 位真彩色用 #RRGGBB 格式表示,其中,RR、GG、BB 分别为()位十六进制数所表示的红色、绿色、蓝色成分的强度值。

- A. 2 B. 8 C. 1 D. 16

【答案】 A

【解析】 24 位真彩色是分别用 8 位(1 字节)二进制来表示红(R)、绿(G)、蓝(B)三色强度值构成的颜色编码。其中,0 代表最弱强度值,255 代表最强强度值。#RRGGBB 格式表示是十六进制。

24. ()为非压缩格式位图文件的扩展名。

- A. gif B. jpg C. bmp D. png

【答案】 C

【解析】

(1) jpg。标识以 JPEG 组织制定的图像压缩格式存储的文件,压缩率高。

(2) bmp。Windows 操作系统中的标准图像格式文件类型,属于位图类型,支持 24 位真彩色,无压缩。

(3) png。一种与平台无关的光栅图像,颜色深度可定义。

(4) gif。只有少量颜色种类(256 种颜色)的位图图像。一个文件可存储多幅图像,以实现动画。

25. 以下选项中与四进制数 123 相等的二进制数是()。

- A. 10101 B. 11011 C. 10111 D. 11101

【答案】 B

【解析】 四进制数的基数是二进制数基数的二次幂,1 位四进制数对应 2 位二进制数。四进制数 123 转换为二进制数为 011011,即 11011。

26. 用 1 字节可以表示最大的无符号十进制整数是()。

- A. 255 B. 8 C. 256 D. 127

【答案】 A

【解析】 1 字节含 8 个二进制位,能表示的最大无符号数为 11111111,依据逢二进一的记数规则,等价于 $100000000 - 1$,对应的十进制值等于 $256 - 1$,即 255。

27. 以下选项中()不是图形图像文件的扩展名。

- A. png B. rm C. gif D. bmp

【答案】 B

【解析】

(1) rm。一种流媒体视频文件格式,可以根据网络数据传输速率制定相应的压缩比率,实现在低速网上实时传送和播放视频文件。

(2) bmp。Windows 操作系统中的标准图像格式文件类型,属于位图类型,支持 24 位真彩色,无压缩。

(3) gif。一种在网络上广泛使用的位图文件,支持 256 种颜色,可有多重图像。

(4) png。一种与平台无关的光栅图像,颜色深度可定义。

28. 以下关于字符 ASCII 码值大小关系的表示中,正确的是()。

【答案】 C

【解析】

(1) 根据负数的补码和原码的数值关系求原码: 补码 = $2^8 - \text{原码} + 2^7$, 即

$$2^8 - 10011011 + 2^7 = 11100101$$

(2) 根据负数的补码、反码、原码之间的关系求原码。

负数原码减符号位得正数原码: $10011011 - 10000000 = 00011011$ 。

正数原码求反得负数反码: $11111111 - 00011011 = 11100100$ 。

反码加 1 得补码: $11100100 + 000000001 = 11100101$ 。

38. 用 1 字节表示数据, 十进制数 75 的二进制原码是()。

- A. 10110101 B. 01001011 C. 11001011 D. 00110100

【答案】 B

【解析】

(1) 根据进制转换关系将十进制数 75 转换为二进制数为 1001011。

(2) 二进制数 1001011 转换为 1 字节长原码为 01001011。

39. 用 1 字节表示数据, 十进制数 -85 的二进制原码是()。

- A. 00101010 B. 01010101 C. 11010101 D. 10101010

【答案】 C

【解析】

(1) 根据进制转换关系将十进制数 -85 转换为二进制数为 -1010101。

(2) 二进制数 1010101 转换为 1 字节长原码为 01010101。

(3) 二进制原码 01010101 转换为对应负数的原码为 11010101。

40. 用 1 字节表示数据, 十进制数 116 的二进制反码是()。

- A. 10000100 B. 11110100 C. 01110100 D. 00001011

【答案】 C

【解析】

(1) 根据进制转换关系将十进制数 116 转换为二进制数为 1110100。

(2) 二进制数 1110100 转换为 1 字节长反码为 01110100。

41. 用 1 字节表示数据, 十进制数 -112 的二进制反码是()。

- A. 10010000 B. 01110000 C. 11110000 D. 10001111

【答案】 D

【解析】

(1) 根据进制转换关系将十进制数 -112 转换为二进制数为 -1110000。

(2) 二进制数 1110000 转换为 1 字节长原码为 01110000。

(3) 二进制原码 01110000 求反转换为对应负数的反码为 10001111。

3.2.2 多选题

1. 当前多媒体技术中主要有三大编码及压缩标准, () 和 () 不属于压缩标准。

A. H.261 B. ASCII C. EBCDIC D. JPEG E. MPEG

【答案】 B,C

【解析】

(1) H.261。数字视频编解码标准,主要用于网络实时视频通信。

(2) JPEG。JPEG 组织制定的图像压缩系列标准。

(3) MPEG。MPEG 组织制定的音频、视频压缩系列标准。

2. 下列选项中()和()属于静态媒体。

A. 声音 B. 图形 C. 动画 D. 图像 E. 视频

【答案】 B,D

【解析】 在多媒体技术中,将与时间相关的各类信息称为动态媒体,而与时间无关的各类信息称为静态媒体。例如,声音、动画、视频为动态媒体,文本、图形、图像为静态媒体。

3. 模拟音频的数字化过程包括音频的采样、()和()。

A. 录制 B. 量化 C. 模拟 D. 编码 E. 压缩

【答案】 B,D

【解析】 音频信号数字化包括采样、量化(A/D 转换)和编码等步骤。

4. 下列选项中正确的是()。

- A. 西文字符在计算机中以 ASCII 码表示
- B. 存储在计算机中的信息以十进制编码表示
- C. 中文的区位码与机内码相同
- D. 汉字字型码也称汉字输出码
- E. ASCII 码采用 7 字节表示一个西文字符的编码

【答案】 A,D

【解析】

(1) 依据现代计算机体系结构,所有代码、数据都以二进制形式存储。

(2) ASCII 码是基于西文字符的编码。标准 ASCII 码采用 7 位二进制编码,共定义了 128 个字符。

(3) 汉字字符编码分为输入码(外码)、机内码、输出码(字形码)。一个汉字可以有多个输入码,也可有多个输出码,但只有一个机内码。

5. 下列选项中()不能作为数据库中的数据进行存储。

A. 人员 B. 图形 C. 电流 D. 文字 E. 声音

【答案】 A,C

【解析】 计算机内存储的是信息的载体——数据,信息是实体的静态特征、动态特征、实体间联系的描述。例如,图形、文字、声音等各表示一种信息类型,描述这些类型信息的就是具体数据,可以存储在数据库中。而人员、电流表达的是一类实体,属于这些类的是实体,不是实体的描述,数据库无法存储。

6. 多媒体信息类型主要有()。

A. 音箱、摄像头 B. 文本、图形

- C. 图像、音频
E. 视频播放器
- D. 软盘、硬盘、光盘

【答案】 B,C

【解析】 多媒体信息种类主要有文本、图形、图像、音频、视频、动画等。

7. 以下选项中与二进制数 11010110.01011 相等的数是()和()。

- A. $(326.26)_O$ B. $(326.13)_O$ C. $(D6.58)_H$
D. $(D6.51)_H$ E. $(214.23)_D$

【答案】 A,C

【解析】 题中 O、H、D 分别代表八进制、十六进制、十进制。如下是二进制数 11010110.01011 转换为其他进制值。

(1) 八进制,每 3 位二进制对应 1 位八进制。整数部分从右往左每次取 3 位,得 326;小数部分从左往右每次取 3 位,不足 3 位右侧补 0,得 26。八进制值为 326.26。

(2) 十六进制,每 4 位二进制对应 1 位十六进制。整数部分从右往左每次取 4 位,得 D6;小数部分从左往右每次取 4 位,不足 4 位右侧补 0,得 58。十六进制值为 D6.58。

(3) 十进制, $1 \times 128 + 1 \times 64 + 1 \times 16 + 1 \times 4 + 1 \times 2 + 1/4 + 1/16 + 1/32$, 得 214.343 75。

8. 下列选项中关于 ASCII 码的描述正确的有()和()。

- A. A 的 ASCII 码值比 B 的 ASCII 码值大
B. ASCII 码可以对 128 个符号进行编码
C. 字符 0 至字符 9 的 ASCII 码值递减
D. 7 位的 ASCII 码用 1 字节表示,其最高位为 0
E. 7 位的 ASCII 码用 1 字节表示,其最高位为 1

【答案】 B,D

【解析】 标准 ASCII 码使用 7 位二进制数组合来表示 128 个字符,包括所有的大小写西文字母、数字、标点符号、运算符号,以及一些特殊控制字符。数字、小写字母、大写字母等拥有连续的 ASCII 编码值。用十进制表示,数字 0~9 的 ASCII 码值分别是 48, 49, ..., 57, 小写字母 a~z 的 ASCII 码值分别是 97, 98, ..., 122, 大写字母 A~Z 的 ASCII 码值分别是 65, 66, ..., 90。在计算机内,使用 1 字节表示 ASCII 码,其最高位为 0。

9. 十进制数转换成 r 进制数的算法为()和()。

- A. 小数部分采用乘 r 取整法
B. 整数部分采用除 r 取余法
C. 小数部分采用除 r 取余法
D. 采用乘 r 取整法或者除 r 取余法
E. 整数部分采用乘 r 取整法

【答案】 A,B

【解析】 在 r 进制的基数、数符都转换为对应的十进制值的前提下,可以依据十进制运算规则将十进制数转换为 r 进制数。使用的算法类似于剥离法,即整数部分用除 r 取余剥离出各个整数位权上的数符,小数部分用乘 r 取整剥离出各个小数位权上的数符。