**2-3-1　文字データ**

在计算机内部，文字数据是由 0 和 1 的组合来表示的。计算机开发初期，使用 8 位（1 字节）的比特模式来对应 1 个字符。另外，与字符对应的比特模式称为**文字コード**，其代表性的种类主要有以下六种。

需要注意的是，如果在使用不同字符编码的计算机之间交换数据，可能会发生**文字化け(乱码)** 现象，即显示出的文字与原始数据不同。

1. **ASCIIコード**

**ASCIIコード**是美国标准协会 **ANSI**于 1962 年制定的文字コード。

它由表示アルファベット(字母)和数字等的**符号ビット**(符号位)**（7 位）**以及用于检错的**パリティビット**(奇偶校验位)**（1 位）**组成，**共 8 位**。

虽然在 PC 等设备中使用，但并不包含有关日文字符（如汉字、假名）的规定。

1. **ISOコード**

**ISOコード**是国际标准化组织 **ISO**于 1967 年在 ASCII 编码的基础上制定的 **7 位**字符编码。这是一种构成全球各国通用字符编码基础的编码方式。

(3) **JISコード**

JIS码是由 **JISC**（日本産業標準調査会）基于ISO码，按照《工业标准化法》审议制定的，用于表示日本特有字符的**日本工业标准**的文字编码。

① **JIS 7単位符号／JIS 8単位符号（JIS X 0201）**

这是能够表示半角片假名的字符编码。使用方法是：从编码表中找到想要表示的字符，取其**列**（高4位）和**行**（低4位），按顺序组合成字符编码。

② **JIS漢字コード（JIS X 0208）**

为了表示平假名和汉字而制定的文字编码，每个字符使用 **2字节（16位）** 表示。

③ **シフトJISコード（Shift\_JIS）**

是扩展自 **JIS汉字码** 的编码体系，它不使用特殊的**切替えコード**(切换码)，就能在同一系统中混合使用1字节字符与2字节字符

考试中常涉及**JIS 8**的看图答文字编码问题

例. “JIS”的文字编码:

J : 0100 1010

I: 0100 1001

S: 0101 0011

一些文字和图案

AI 生成的内容可能不正确。

1. EBCDIC

EBCDIC 是由美国 IBM 公司开发的8ビット文字コード。最初是 IBM 为自家计算机设计的专用编码，但到了第 3 代（20 世纪 60 年代），由于 IBM 占据了计算机市场的大部分份额，这种编码就成为大型计算机领域的**業界標準（デファクトスタンダード）**。

1. **Unicode**

**Unicode** 是由美国的 Apple 公司、IBM 公司、Microsoft 公司等为使 PC 的数据交换顺利化而作为**2 字节系的万国统一代码**来构想／提倡的字符编码。支持的文字有字母、汉字与平假名／片假名、韩文字、阿拉伯文字等，支持了许多国家的文字。对此，ISO 将其作为国际标准进行规范化，目前定义了 **UCS-2（2 字节）** 与 **UCS-4（4 字节）**。

另外，还有称为 **UTF-8、UTF-16、UTF-32** 的**规格（编码方法）**。

* **UTF-8** : 与ASCII兼容,

在 **UCS-2** 中属于 ASCII 代码所包含的文字为 **1 字节长度**，

其余的文字以 **2 字节、3 字节，…** 按字节为单位表示的**可変長方式**

* **UTF-16:** 一个字符用 2 或 4 字节表示，**可変長方式**
* **UTF-32** : 一个字符4 字节表示, **固定長方式**。

另外，在很多场合，**Unicode** 一词在与 **UTF-8** 同时出现时，它实际上**不是真的泛指 Unicode 标准本身**，而是**特指 UTF-16 编码方式**。

“这个文件是 Unicode 编码的。”  
实际上就是 **UTF-16**。

1. **EUC（Extended Unix Code；拡張UNIXコード）**

EUC（扩展 UNIX 码）是 AT&T 贝尔实验室为 UNIX 操作系统的国际化支持而开发的字符编码。半角英数字：1 字节**;** 汉字/假名：2 字节

汉字编码是在 JIS 汉字码的基础上加上十六进制 “80”

因此在汉字的字节中最高位总是 “1”

这就可以区分 1 字节的英数字和 2 字节的汉字。

除上述6种之外，有一种叫 **ゾーン10進数（Zone Decimal）** 的数字表示方式，它**不是字符编码**（不用于表示任意文字），但有时候也会被归类到“表示数字字符的方法”里。

* Zone Decimal 主要用在老式的大型机和某些金融系统中，用“区位”+“数字”组合的方式来表示十进制数。它更多是**数据存储格式**而不是通用的字符编码。
* 我们会在下一节[数値データ]种详细介绍