

第一章 项目三

汉字编码

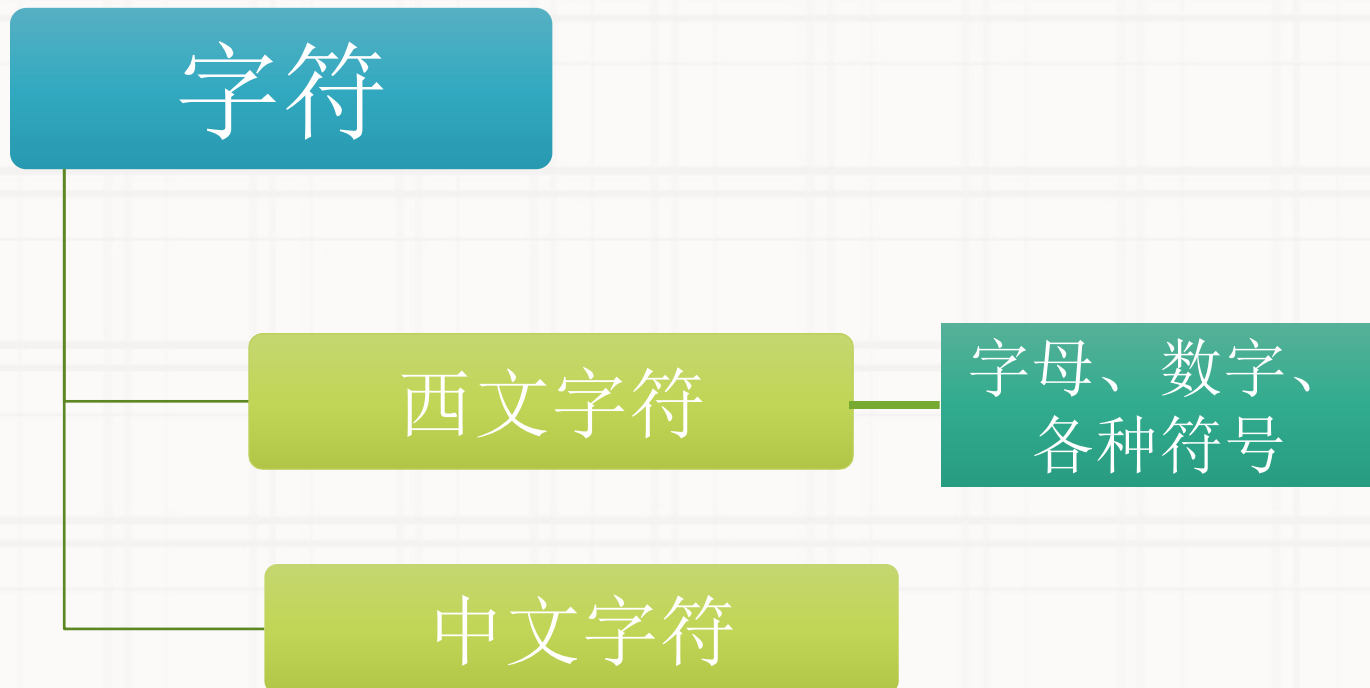


1.2.5 字符的编码

重点：汉字编码类型及编码方式

难点：ASCII码

1.2.5 字符的编码



1.2.5 字符的编码

❖ 1.西文字符的编码

- 计算机中的数据都是用二进制编码表示的，用以表示字符的二进制编码称为**字符编码**。
- 而，计算机最常用的就是ASCII。
- 美国信息交换**标准码ASCII**，7位码（128个字符）
扩展码ASCII，8位码（256个字符）

7位——国际通用

1.2.5 字符的编码

❖ 1.西文字符的编码

ASCII码对128个字符进行了编码，这其中包括大小写英文字母、阿拉伯数字、标点符号以及控制字符等特殊字符。

ASCII码中共有34个非图形字符（控制字符），94个可打印字符（图形字符）。

读表1-6可知，每个符号的ASCII编码为：高3位+低4位。

1.2.5 字符的编码

❖ 1.西文字符的编码

“a” 字符编码为1100001 (97) , bcde…依次增加

“A” 字符编码为1000001 (65) , BCDE…依次增加

“0” 数字字符编码为0110000 (48) , 1234依次增加

■计算机内部用一个字节（8个二进制位）存放一个7位ASCII码，最高位置为0。

重点掌握字符的推算！！！！

1.2.5 字符的编码

❖ 1.西文字符的编码

ASCII码中存在以下两个规律：

(1) 大小写英文字母和阿拉伯数字都是按其固有顺序排列在表中的。

(2) 在表中，阿拉伯数字排在大写英文字母前，而大写英文字母排在小写英文字母前。

1.2.5 字符的编码

❖ 1.西文字符的编码（非图形字符/控制字符）

SP（空格）的编码是 0100000

CR（回车）的编码是 0001101

DEL（删除）的编码是 1111111

BS（退格）的编码是 0001000

1.2.5 字符的编码

❖ 2.汉字的编码——国标码

■ 汉字分成两级：

一级汉字3755个，按汉语拼音字母的次序排列；

二级汉字3008个，按偏旁部首排列。

- 由于一个字节不足以表示6763个汉字，所以一个国标码用两个字节来表示一个汉字，每个字节最高位为0。



1位
7位

7位

1位

1.2.5 字符的编码

❖ 2.汉字的编码——区位码

- GB2312-80将汉字分为94行、94列，代码表分为94个区（行）、94个位（列）。
- 区位码是一个4位十进制数
- 国标码是一个4位十六进制数。

1.2.5 字符的编码

❖ 2.汉字的编码——区位码、国标码转换

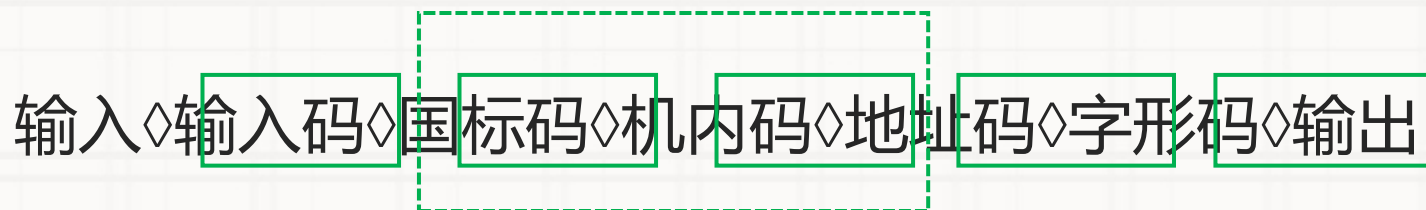
汉字国标码 = 区位码 + (2020)_H

区号和位号： 转换为16进制数，**分别**加上 (20)_H

1.2.5 字符的编码

❖ 3.汉字的处理过程

- 由于汉字的特殊性（象形文字），汉字的处理过程必须同时需要多种编码才能完成。



以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/558111031124006124>