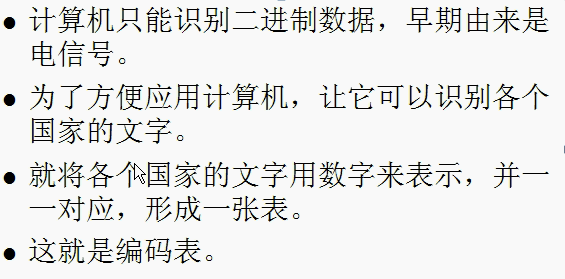
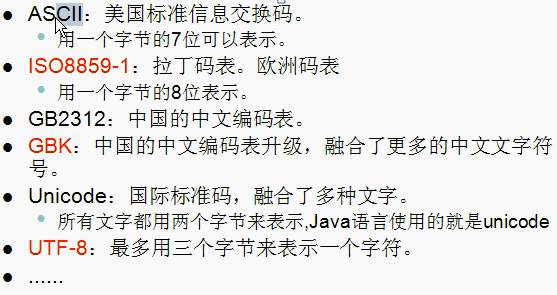
1. 编码表的由来



1. 常见编码表



常用字符集分类

* 1. ASCII及其扩展字符集

作用：表语英语及西欧语言。

位数：ASCII是用7位表示的，能表示128个字符；其扩展使用8位表示，表示256个字符。

范围：ASCII从00到7F，扩展从00到FF。

* 1. ISO-8859-1字符集

作用：扩展ASCII，表示西欧、希腊语等。

位数：8位，

范围：从00到FF，兼容ASCII字符集。

* 1. GB2312字符集 (中文两个字节表示，中文编码两个字节都是负数)

作用：国家简体中文字符集，兼容ASCII。

位数：使用2个字节表示，能表示7445个符号，包括6763个汉字，几乎覆盖所有高频率汉字。

范围：高字节从A1到F7, 低字节从A1到FE。将高字节和低字节分别加上0XA0即可得到编码。

* 1. BIG5字符集

作用：统一繁体字编码。

位数：使用2个字节表示，表示13053个汉字。

范围：高字节从A1到F9，低字节从40到7E，A1到FE。

* 1. GBK字符集 (中文两个字节表示，中文编码第一个字节都是负数，第二个字节不一定是负数，例如” 琲”，编码为-84 105)

作用：它是GB2312的扩展，加入对繁体字的支持，兼容GB2312。

位数：使用2个字节表示，可表示21886个字符。

范围：高字节从81到FE，低字节从40到FE。

* 1. GB18030字符集

作用：它解决了中文、日文、朝鲜语等的编码，兼容GBK。

位数：它采用变字节表示(1 ASCII，2，4字节)。可表示27484个文字。

范围：1字节从00到7F; 2字节高字节从81到FE，低字节从40到7E和80到FE；4字节第一三字节从81到FE，第二四字节从30到39。

* 1. UCS字符集

作用：国际标准 ISO 10646 定义了通用字符集 (Universal Character Set)。它是与UNICODE同类的组织，UCS-2和UNICODE兼容。

位数：它有UCS-2和UCS-4两种格式，分别是2字节和4字节。

范围：目前，UCS-4只是在UCS-2前面加了0x0000。

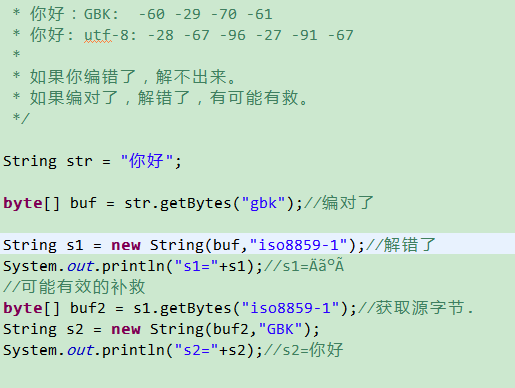
* 1. UNICODE字符集

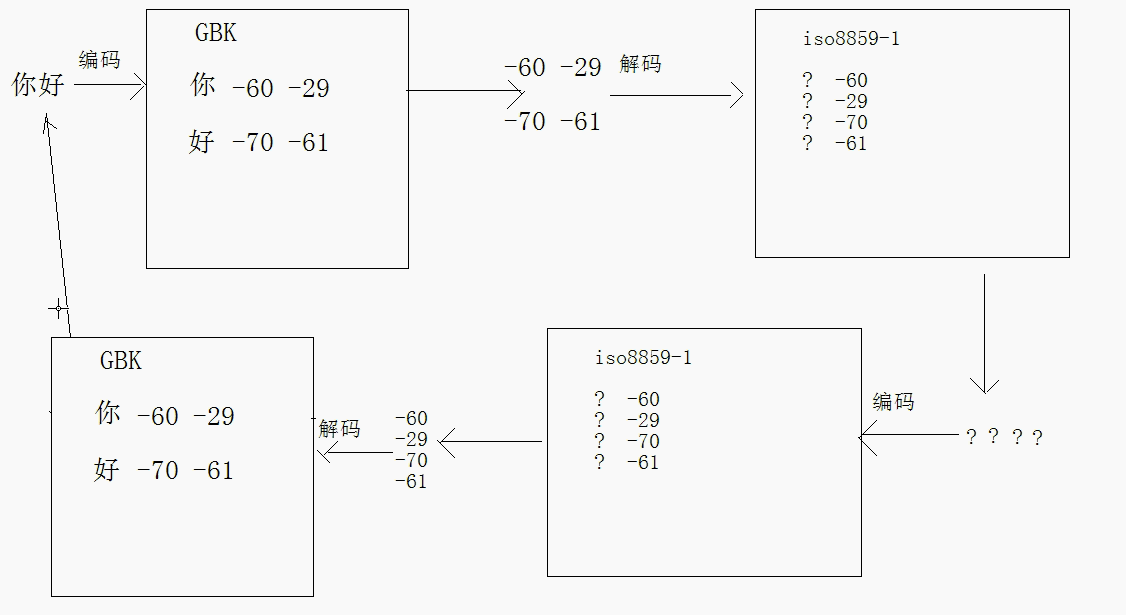
作用：为世界650种语言进行统一编码，兼容ISO-8859-1。

位数：UNICODE字符集有多个编码方式，分别是UTF-8，UTF-16和UTF-32。

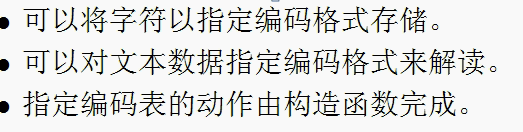
UTF-8一个汉字采用3个字节表示，并且全是负数

1. 编码解码问题





1. 转换流编码应用



1. 按字节截取字符串

