

WEB开发中的JAVA字符编码经验总结 PDF转换可能丢失图片或格式，建议阅读原文

https://www.100test.com/kao_ti2020/492/2021_2022_WEB_E5_BC_80_E5_8F_91_E4_c67_492419.htm 一、概要

在JAVA应用程序特别是基于WEB的程序中，经常遇到字符的编码问题。为了防止出现乱码，首先需要了解JAVA是如何处理字符的，这样就可以有目的地在输入/输出环节中增加必要的转码。其次，由于各种服务器有不同的处理方式，还需要多做试验，确保使用中不出现乱码。

二、基本概念 2.1 JAVA中字符的表达

JAVA中有char、byte、String这几个概念。char指的是一个UNICODE字符，为16位的整数。byte是字节，字符串在网络传输或存储前需要转换为byte数组。在从网络接收或从存储设备读取后需要将byte数组转换成String。String是字符串，可以看成是由char组成的数组。String和char为内存形式

，byte是网络传输或存储的序列化形式。举例：英 String ying = “英”. char ying = ying.charAt(0). String yingHex =

Integer.toHexString(ying). 82 F1 byte yingGBBytes = ying.getBytes(“GBK”). GB编码的字节数值 D3 A2 2.2 编码方式的简介

String序列化成byte数组或反序列化时需要选择正确的编码方式。如果编码方式不正确，就会得到一些0x3F的值。常用的字符编码方式有ISO8859_1、GB2312、GBK

、UTF-8/UTF-16/UTF-32。ISO8859_1用来编码拉丁文，它由单字节（0 - 255）组成。GB2312、GBK用来编码简体中文，它有单字节和双字节混合组成。最高位为1的字节和下一个字节构成一个汉字，最高位为0的字节是ASCII码。

UTF-8/UTF-16/UTF-32是国际标准UNICODE的编码方式。用

得最多的是UTF-8，主要是因为它在对拉丁文编码时节约空间。UNICODE值 UTF-8编码 U-00000000 - U-0000007F:
0xxxxxxx U-00000080 - U-000007FF: 110xxxxx 10xxxxxx
U-00000800 - U-0000FFFF: 1110xxxx 10xxxxxx 10xxxxxx
U-00010000 - U-001FFFFFF: 11110xxx 10xxxxxx 10xxxxxx 10xxxxxx
U-00200000 - U-03FFFFFF: 111110xx 10xxxxxx 10xxxxxx
10xxxxxx 10xxxxxx U-04000000 - U-7FFFFFFF: 1111110x
10xxxxxx 10xxxxxx 10xxxxxx 10xxxxxx 10xxxxxx 三、J2SE中相关的函数

```
String str = "英" .//取得GB2312编码的字节 byte[]
bytesGB2312 = str.getBytes( " GB2312 " ).//取得平台缺省编码的
字节(solaris为ISO8859_1,windows为GB2312) byte[] bytesDefault
= str.getBytes(). //用指定的编码将字节转换成字符串 String
newStrGB = new String(bytesGB2312, " GB2312 " ).//用平台缺省的
编码将字节转换成字符串(solaris为ISO8859_1,windows
为GB2312) String newStrDefault = new String(bytesDefault). //用指
定的编码从字节流里面读取字符 InputStream in = xxx.
InputStreamReader reader = InputStreamReader( in, " GB2312 " ).
char aChar = reader.read().
```

100Test 下载频道开通，各类考试题目直接下载。详细请访问 www.100test.com