

基础入门 - JAVA字符集详解（二）PDF转换可能丢失图片或格式，建议阅读原文

[https://www.100test.com/kao\\_ti2020/145/2021\\_2022\\_\\_E5\\_9F\\_BA\\_E7\\_A1\\_80\\_E5\\_85\\_A5\\_E9\\_c104\\_145066.htm](https://www.100test.com/kao_ti2020/145/2021_2022__E5_9F_BA_E7_A1_80_E5_85_A5_E9_c104_145066.htm) 3.4.2. 文件编译 假设文件是gbk编码保存的，而编译有两种编码选择：gbk或者iso8859-1，前者是中文windows的默认编码，后者是linux的默认编码，当然也可以在编译时指定编码。 `Jsp *(gbk:d6d0 cec4) java file *(gbk:d6d0 cec4) compiler read uicode(gbk: 4e2d 6587. iso8859-1: 00d6 00d 000ce 00c4) compiler write utf(gbk: e4b8ad e69687. iso8859-1: *) compiled file unicode(gbk: 4e2d 6587. iso8859-1: 00d6 00d 000ce 00c4) class`。 所以用gbk编码保存，而用iso8859-1编译的结果是不正确的。 `class unicode(4e2d 6587) system.out / jsp.out gbk(d6d0 cec4) os console / browser`。 | 文件可以以多种编码方式保存，中文windows下，默认为ansi/gbk。编译器读取文件时，需要得到文件的编码，如果未指定，则使用系统默认编码。一般class文件，是以系统默认编码保存的，所以编译不会出问题，但对于jsp文件，如果在中文windows下编辑保存，而部署在英文linux下运行/编译，则会出现问题。所以需要在jsp文件中用pageEncoding指定编码。Java编译的时候会转换成统一的unicode编码处理，最后保存的时候再转换为utf编码。当系统输出字符的时候，会按指定编码输出，对于中文windows下，System.out将使用gbk编码，而对于response（浏览器），则使用jsp文件头指定的contentType，或者直接为response指定编码。同时，会告诉browser网页的编码。如果未指定，则会使用iso8859-1编码。对于中文，应该为browser指定输出字符串的编码。

browser显示网页的时候，首先使用response中指定的编码（jsp文件头指定的contentType最终也反映在response上），如果未指定，则会使用网页中meta项指定中的contentType。

### 3.5. 几处设置

对于web应用程序，和编码有关的设置或者函数如下。

#### 3.5.1. jsp编译 指定文件的存储编码

很明显，该设置应该置于文件的开头。例如：`<%@page contentType="utf-8"%>`。另外，对于一般class文件，可以在编译的时候指定编码。

#### 3.5.2. jsp输出 指定文件输出到browser是使用的编码

该设置也应该置于文件的开头。例如：`<@output encoding="gbk"%>`。该设置和`response.setCharacterEncoding("GBK")`等效。

#### 3.5.3. meta设置 指定网页使用的编码

该设置对静态网页尤其有作用。因为静态网页无法采用jsp的设置，而且也无法执行`response.setCharacterEncoding()`。例如：如果同时采用了jsp输出和meta设置两种编码指定方式，则jsp指定的优先。因为jsp指定的直接体现在response中。需要注意的是，apache有一个设置可以给无编码指定的网页指定编码，该指定等同于jsp的编码指定方式，所以会覆盖静态网页中的meta指定。所以有人建议关闭该设置。

#### 3.5.4. form设置

当浏览器提交表单的时候，可以指定相应的编码。例如：`<input type="text" value="" encoding="utf-8"%>`。一般不必不使用该设置，浏览器会直接使用网页的编码。

## 4. 系统软件

下面讨论几个相关的系统软件。

### 4.1. mysql数据库

很明显，要支持多语言，应该将数据库的编码设置成utf或者unicode，而utf更适合与存储。但是，如果中文数据中包含的英文字母很少，其实unicode更为适合。数据库的编码可以通过mysql的配置文件设置，例如`default-character-set=utf8`。还可以在数据库链接URL中设置，例如：

```
useUnicode=true&characterEncoding=UTF-8。
```

注意这两者

应该保持一致，在新的sql版本里，在数据库链接URL里可以不进行设置，但也不能是错误的设置。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](http://www.100test.com)