

GB2312、GBK与UTF-8的区别

转载

于 2016-06-

☆

ZZ...

🕒 24

👁 1274 收藏

17:26:00 发布

7

首先，我们要明白，GB2312、GBK和UTF-8都是一种字符编码，除此之外，还有好多字符编码。只是对于我们中国人的网站来说，用这三种编码 比较多。简单的说一下，为什么要用编码，在计算机内，储存文本信息用ASCII码，每一个字符对应着唯一的ASCII码。最初计算机是由美国发明的，他们也用的是键盘和上面的字母，所以他们的字符ASCII好解决。但是我们中国的就不同了，每个汉字要对应唯一的ASCII码。这样，就出来了国家制定的字符编码标准：GB2312、GBK等。其他国家，其他语言也有他们对应的编码 标准。

GB 就是国标的意思，GB2312和GBK主要用于汉字的编码，而UTF-8是全世界通用的。意思就是说，如果你的网页主要面对使用汉语的中国人的话，使用 GB2312和GBK非常好，文字储存体积要小，有一些优点。如果你的网页要面向世界的话，你再用GB2312和GBK作为网页编码的话，有些电脑上的浏览器没有这种编码，你的网页汉字内容就会变成无法识别的乱码。

它们通常用在网页的meta标签内，例如：`<meta http-equiv="Content-Type" content="text/html; charset=gb2312" />`，表示这个页面使用的是GB2312编码。这个信息是给浏览器看的，浏览器会优先考虑使用从网页头部提取出来的编码信息对网页进行解码。当然，我们也可以强制浏览器使用某种编码解释网页，这样我们就看到了传说中的乱码。请看下图IE浏览器：

百度首页使用的是GB2312编码，我们可以看到现在是正常的。我们右击页面，选择“编码”->“其他”->“Unicode (UTF-8)”，意思就是强制浏览器使用UTF-8的编码方式解析页面，我们可以看到奇迹发生了：

百度页面上所有的汉字都变成了乱码。如果你的网页使用了GB2312编码，却被一台没有

GB2312编码的电脑访问了，里面所有的汉字都成了乱码。如果你使用UTF-8编码，在没有汉字的电脑里，仍然可以正常显示，因为UTF-8是通用的编码，所有电脑都有。

所以，在编写网页时，尽量使用UTF-8编码。

UTF-8：Unicode Transformation Format-8bit，允许含BOM，但通常不含BOM。是用以解决国际上字符的一种多字节编码，它对英文使用8位（即一个字节），中文使用24为（三个字节）来编码。UTF-8包含全世界所有国家需要用到的字符，是国际编码，通用性强。UTF-8编码的文字可以在各国支持UTF8字符集的浏览器上显示。如，如果是UTF8编码，则在外国人的英文IE上也能显示中文，他们无需下载IE的中文语言支持包。

GBK是国家标准GB2312基础上扩容后兼容GB2312的标准。GBK的文字编码是用双字节来表示的，即不论中、英文字符均使用双字节来表示，为了区分中文，将其最高位都设定成1。GBK包含全部中文字符，是国家编码，通用性比UTF8差，不过UTF8占用的数据库比GBD大。

GBK、GB2312等与UTF8之间都必须通过Unicode编码才能相互转换：

GBK、GB2312——Unicode——UTF8

UTF8——Unicode——GBK、GB2312

对于一个网站、论坛来说，如果英文字符较多，则建议使用UTF-8节省空间。不过现在很多论坛的插件一般只支持GBK。

GB2312是GBK的子集，GBK是GB18030的子集

GBK是包括中日韩字符的大字符集合

如果是中文的网站 推荐GB2312 GBK有时还是有点问题

为了避免所有乱码问题，应该采用UTF-8，将来要支持国际化也非常方便

UTF-8可以看作是大字符集，它包含了大部分文字的编码。

使用UTF-8的一个好处是其他地区的用户（如香港台湾）无需安装简体中文支持就能正常观看你的文字而不会出现乱码。

gb2312是简体中文的码

gbk支持简体中文及繁体中文

big5支持繁体中文

utf-8支持几乎所有字符

中国大陆最常用的就是GBK18030编码，除此之外还有GBK，GB2312，这几个编码的关系是这样的。

最早制定的汉字编码是GB2312，包括6763个汉字和682个其它符号

95年重新修订了编码，命名GBK1.0，共收录了21886个符号。

之后又推出了GBK18030编码，共收录了27484个汉字，同时还收录了藏文、蒙文、维吾尔文等主要的少数民族文字，现在WINDOWS平台必需要支持GBK18030编码。

按照GBK18030、GBK、GB2312的顺序，3种编码是向下兼容，同一个汉字在三个编码方案中是相同的编码。

GB2312编码大约包含6000多汉字（不包括特殊字符），编码范围为第一位b0-f7,第二位编码范围为a1-fe(第一位为cf时,第二位为a1-d3),计算一下汉字个数为6762个汉字。当然还有其他的字符。包括控制键和其他字符大约7573个字符编码

gbk编码是对G B2312编码的扩充，容纳的汉字更多，但仅仅是扩充，没有质的变化。保留了所有G B2312编码,在此基础上进行编码范围的扩充.容纳(包含特殊字符)共22014个字符编码.

gb18030编码是在gbk编码基础上的扩充，因为汉字更多，仅仅使用两位编码已经不能

容纳要求的汉字，所以采用了2\4位混和的办法，可以支持更多的汉字编码。并且保留了原有的gbk 2字节编码兼容G B2312和gbk编码的文件。大概容纳55657个编码(包含特殊字符)

unicode编码(也就是UTF编码)：俗称万国码，致力于使用统一的编码准则表达各国的文字。

为表达更多的文字，utf-8采用2/3混编的方式。目前容纳的汉字范围小于gbk编码。并且以

3字节的方式处理中文，带来了兼容性的问题，原有的gbk,G B2312,gb18030编码文件都不能正常的处理,还有很长的路要走。

GB2312转UTF-8（C代码）

12-04

本资源为GB2312转UTF-8的C语言代码，实现方式为...

字符编码详解：AS... m0_47584501的博客 568 无

C语言学习(四)ASCII、GB2312、GBK等...

8-15

C语言有时候使用ASCII码有时候不是,他有时候会使...

字符编码:ASCII,Unicode,UTF-8,GB2312...

8-10

从ASCII、GB2312、GBK到GB18030,这些编码方法...

常见中文字符编码... qq_43704127的博客

2357

中文字符编码是用来表示中文字符的一种方式，常...

计算机编码知识，ASCI... 梦悦foundation

3733

文章目录1. 编码的作用1. 方便和统一管理2. 常用的...

你真的了解计算机中文存储吗?_や空空`... 7-30
以GB2312字符集的第一个汉字“啊”字为例,它的区号...

计算机用几个字节储存,GB2312编码的... 8-10
每个汉字占两个字节。GB2312字符在计算机中存储...

GB2312、GBK、Unicode和U... dawuafang 84
汉字编码知识点 ASCII码是西欧编码的方式，采取7...

C语言入门教程，C语... digitalkee的博客 6801
C语言中文网推出辅导班啦，包括「C语言辅导班、...

彻底搞明白 GB2312、GBK 和 GB18030... 8-14
GB2312 兼容 ASCII 编码, GBK 兼容 GB2312 编码,G...

GB2312、GBK、Unicode和UTF-8编码介... 8-6
所以GB2312编码中汉字区码的十进制是从176到24...

C语言小... 热门推荐 qq_43160529的博客 1万+
第1章 编程基础 1.14 GB2312 编码和 GBK 编码，将...

计算机常用编码方式 Andy的博客 4021
计算机常用编码方式

网页编码为什么不用gbk或者gb2312... 09-22
GBK的文字编码是双字节来表示的，至于UTF-8编...

GB2312编码和UTF-8互转（c语言实... 01-01
GB2312编码与utf-8编码的字符串的转换，主要使用...

GBK、UTF-8编码转换工具 11-01
GBK、UTF-8批量文件3秒快速转码工具(支持GBK,UT...

GBK与UTF-8的区别 12-11
1.GBK是在国家标准 GB2312基础上扩容后兼容GB2...

计算机如何存储中文... li_ding_ding的博客 944
计算机如何存储中文字符： 1：计算机如何存储英文...

各种编码中字母... weixin_52293201的博客 696
1. ASCII码:上个世纪60年代,美国制定了一套字符编...

各种编码格式(GB231... guoweikd的博客 8543
汉字常用编码格式 为了在屏幕上显示字符。需要下...

gbk与gb2312的区别是什... 天涯路的专栏 5428
区别： GB2312，由中华人民共bai和国政府制定的d...

GBK与GB2312 weixin_43470000的博客 2008
#概述

java utf-8 gbk_Java 字符转... 最新发布 06-02
在 Java 中，可以使用 `String` 类的 `getBytes()` 方...

“相关推荐”对你有帮助么？



助

助

助

助

关于我们

招贤纳士

商务合作

寻求报道

400-660-0108

kefu@csdn.net

在线客服

工作时间 8:30-22:00

京公网安备11010502030143号

京网文〔2020〕1039-165号

经营性网站备案信息

北京互联网违法和不良信息举报中心

家长监护

网络110报警服务

中国互联网举报中心

Chrome商店下载

账号管理规范

版权与免责声明

版权申诉

出版物许可证

营业执照

©1999-2023北京创新乐知网络技术有限公司

ZZZZVSS

码龄7年

暂无认证

111

17万+

92万+

58万+

原创

周排名

总排名

访问

等级

6379

76

186

100

357

积分

粉丝

获赞

评论

收藏

私信

关注

搜博主文章

Q

热门文章

mybatis大于小于的转义 34284

Java 通过Socket监听指定服务器（IP）的指定端口，及向指定服务器的指定端口发送信息 27331

本地连接服务器mysql 20071

解决linux（Centos）没有 iptables 解决 17142

mySQL创建表时添加表和列的注释，以及查看表和列的注释 15537

分类专栏

国际化i18n

1篇

微信

25篇

java

49篇

Tomcat

11篇

mysql

17篇



最新评论

解决linux（Centos）没有 iptables 解决
qq_40113195: 查看状态 systemctl status iptables.service 这句最好

网页调用QQ聊天
2301_77076062: 你好，怎么联系客服？

通过建设银行外联平台进行转账/提现操作...
ZZZZVSS: 源码都在博文里了，这个好多年前了，早就找不到了。。。

通过建设银行外联平台进行转账/提现操作...
靠近亿点点: 楼主，能提供源代码嘛

解决linux（Centos）没有 iptables 解决
发量**: 执行到iptables -P INPUT DROP xshell断开连不上了

您愿意向朋友推荐“博客详情页”吗？



强烈不推荐



不推荐



一般般



推荐



强烈推荐

最新文章

获取某个时间的当月起止时间

tomcat配置去掉项目名称

MYSQL查询某字段中以逗号分隔的字符串

2020年 26篇	2019年 32篇
2018年 36篇	2017年 80篇
2016年 62篇	