

信息处理

## Internet 上的常用汉字编码及其显示和转换

朱江

(中国科学院成都文献情报中心, 成都610041)

**摘要** 本文对 Internet 上常用的汉字编码及其显示和转换进行了论述。Internet 上常用的汉字编码有国际码、HZ 码和大五码三种, 这三种汉字编码可利用 Internet 上的一些转换软件来相互转换。在使用 Internet 上不同编码的汉字信息资源时应启动不同的汉字操作系统, 双桥 UnionWay 等 Windows 中文平台则可用来对多种汉字编码进行显示。

**关键词** Internet 汉字编码 汉字操作系统 Windows 中文平台 汉字信息资源。

## Chinese Character Coding Schemes Frequently Used over Internet and Their Display and Conversion

Zhu Jiang

(Chengdu Documentation and Information Center, CAS, Chengdu 610041)

**Abstract** This paper discusses the Chinese character coding schemes frequently used over Internet as well as their display and conversion. They are GB, HZ and Big-5. The Chinese files in these three coding schemes can be converted with some converting software available on Internet. The different Chinese operating systems should be run to make use of the Chinese information resources with differently coded Chinese characters. The Chinese platforms of Windows such as TwinBridge and UnionWay, however, can be used to view several differently coded Chinese characters.

**Keywords** Internet, Chinese character coding, Chinese operating system, Chinese platform of Windows, Chinese information resources

Internet 上有着较为丰富的汉字信息资源, 但汉字编码的不统一却给广大用户带来了利用上的不便。目前, 国内很多 Internet 用户都是利用微机联入 Internet 的, 因此, 熟悉并掌握 Internet 上的汉字编码及其在微机上的使用和转换, 对提高广大用户的 Internet 应用水平, 促

收稿日期: 1996年5月27日

作者简介: 朱江, 男, 1968年生, 助研, 《中国专利数据库》编辑部副主编, 曾获国家科技进步三等奖, 中科院科技进步二等奖。

进 Internet 上汉字信息的使用和交流都有着非常积极的作用。

## 1 Internet 上常用的汉字编码

Internet 上常用的汉字编码有国标(GB)码、HZ 码和大五(Big-5)码三种。

### 1.1 国标码

国标码是“中华人民共和国国家标准信息交换用汉字编码”的简称。国标码是一种基于简体汉字的双字节汉字编码,在中国、新加坡等使用简体汉字的国家通行。国标字符集的基本集(GB 2312-80)共收录图形字符7445个(包括标点符号、数字、外文字母等,其中汉字6763个)。这7000余个图形字符被排列成一个 $94 \times 94$ 的方阵,每个汉字在这个方阵中的坐标被称为该汉字的区位码。每个汉字的区、位码各加上32,就变成了该汉字的国标码。由于国标码的码值与ASCII码中的94个图形符的码值完全重合,为了以示区别,人们将国标码两个字节的低位全部置1,这种编码就是大家习惯上所称为“国标码”。国标码高位置1后,采用国标码编码的汉字文件就不再是标准的ASCII文本文件了,这种文件在Internet上用简单邮政传输协议(SMTP)传输时有可能出现问题。人们在Internet上用SMTP协议传输国标码的汉字文件时,通常是用uuencode编码后再传输,到达后再用uudecode解码后阅读,我国出版的电子期刊《神州学人》目前就是用这种方法向订户“投递”刊物的。

### 1.2 HZ 码

为了解决国标码的汉字文件在Internet上传输不便的难题,海外的中国学者又设计了一种新的汉字编码——HZ码。HZ码属双字节7位汉字编码,它以“国标”为基础,用两个ASCII图形符来表示一个汉字。HZ码与国标码的主要区别是将国标码的高位还原为0,并通过在HZ码的前后分别加上“逃出码”(~{ })和“逃入码”(~{ })来将HZ码与ASCII码区分开。~{ }和~{ }里面为HZ码,外面即为正常的ASCII码。采用这种编码的汉字文件不仅解决了中西文的区分、混排问题,而且从表面上看是一个标准的ASCII文本文件,因而可以在Internet上自由传输<sup>[1]</sup>,这就是HZ码在Internet上比较流行的一个重要原因。Internet上的网络新闻组alt.chinese.text、“太阳升”中文Gopher服务器以及很多的中文电子期刊都使用HZ码。

### 1.3 大五码

大五码是一种基于繁体汉字的双字节8位汉字编码,在台湾、香港等使用繁体汉字的地区通行。大五码收录汉字约13,000个(包括标点符号、数字、外文字母等),比国标码多出近一倍。大五码的首字节高位为1(第二字节的高位既可为1,也可为0),因而在Internet上传输时存在着同国标码一样的问题。目前,台湾、香港地区的Internet主机使用的汉字编码大多是大五码。

## 2 Internet 上常用汉字编码的使用

使用Internet上不同编码的汉字信息资源时必须启动相应的汉字操作系统或Windows中文平台,这些系统软件有一部分可从Internet上通过匿名FTP的方式获取。

## 2.1 处理国标码的汉字操作系统

我国目前通行的各种汉字操作系统,如UCDOS等都用来输入和显示国标码。

## 2.2 处理HZ的汉字操作系统

HZ码的输入和显示可利用Internet上的共享软件(shareware)——ZWDO S。ZWDO S 2.0的存取地址是:

```
ftp://ftp.bd748.pku.edu.cn/pub/chinese/zw dos220.zip
```

ZWDO S是内存驻留程序,启动后就一直驻留在内存中监视着屏幕上的内容,一旦发现用~{和~}括起来的ASCII码,就将~{和~}之间的ASCII码转换为汉字,清屏后再重新显示一次。如果屏幕上不再有HZ码,就自动退出汉字显示状态,回到DOS状态。ZWDO S对内存的需求很小,只占用26K系统内存<sup>[2]</sup>,可支持CGA,EGA,VGA和单显多种显示器,使用起来非常灵活。ZWDO S的优势还在于它能同UCDOS等汉字操作系统一起使用,互不干扰,可以用来对国标码和HZ码编码的汉字文件进行交替显示。

ZWDO S的主要缺陷是显示速度较慢,特别是微机主频较低的时候,这一缺陷会更加明显。

最近,对ZWDO S 2.0作了全面改进的ZWDO S 3.0已在Internet上正式推出,其存取地址是:

```
ftp://ftp.edu.tw/Chinese/ifcss/software/dos/ZWDO S/zw *. *
```

## 2.3 处理大五码的汉字操作系统

用来处理大五码的汉字操作系统也可从Internet上通过匿名FTP的方式获取,KC是其中的一种,其存取地址是:

```
ftp://moers2.edu.tw/chinese- pub/chinese- sys/kc/kc.zip
```

kc.zip的字节数近4M,感觉不便的用户可分次“抓取”kc1.zip—kc5.zip共五个文件,效果也是一样的。具体的安装使用说明请参考该子目录下的readme文件。

## 2.4 Windows中文平台

对习惯使用MS-Windows的用户来说,也有一些相应的Windows中文平台可供选择,双桥(TwinB ridge)是其中一种。双桥是美国PC Express公司开发的多语言平台系统,除中文版外,还有相应的日文版、韩文版、欧洲版等。中文版的双桥3.2可在Windows 3.x和中文Windows 3.x上安装,其演示版的存取地址是:

```
ftp://cnd.org/pub/software/ms-w in/c- sys/tbcdemo.zip
```

```
ftp://ftp.ifcss.org/pub/software/ms-w in/c- sys/tbcdemo.zip
```

双桥3.2拥有国标码和大五码两套内码,可十分方便地显示国标码、HZ码和大五码的汉字文件。当双桥的内码置为国标码时,可自动识别国标码和HZ码并予以正确地显示,置为大五码时则只能正确显示大五码,但不管是国标码还是以大五码为内码,双桥均可对汉字文件用简、繁两种字体进行显示。双桥的两套内码可通过一个按键随意切换,使用起来十分方便。此外,双桥还拥有拼音、注音、仓颉、区位、英文、GB码、Big-5码等多种汉字输入法,全角标点、日文假名、俄文、希腊字母也可直接输入,是一种比较有效的Windows中文平台。由于双桥3.2已

经推出,双桥3.2演示版最近已被其他软件取代。

另一种类似的Windows中文平台是UnionWay 4.0。UnionWay 4.0是美国ComStar公司开发的通用双字节窗口平台,具备处理中、日、韩三种文字的能力,并可在Windows 3.x和Windows 95上运行,同时支持相应的中、日、韩三种文字的Windows系统,其标准演示版的存取地址是:

ftp://www.unionway.com/unionway/uwCJKstd/uwCJKstd.exe

ftp://ftp.ainet.com/pub/users/chinabus/uwCJKstd.exe

用来处理中文的标准演示版的存取地址是:

ftp://www.unionway.com/unionway/uwCstd/uwCstd.exe

UnionWay 4.0有拼音、注音、仓颉等多种汉字输入法,并可在国标码和大五码之间方便地切换,同时支持HZ码的显示,使用效果比较理想。

此外,Internet上的南极星,汉文以及四通利方公司的RichWin,方正新天地公司的中文之星(Chinese Star)等软件也是同一类型的软件,本文不再赘述。需要指出的是,双桥UnionWay等Windows中文平台可支持多种汉字编码的显示,而且切换方便,使用效果应该说比处理单一汉字编码的汉字操作系统要好些,用户在选择相应的中文系统时应对此予以充分考虑。

### 3 Internet 上常用汉字编码的转换

利用上述汉字操作系统和Windows中文平台就可以比较方便地使用和阅读Internet上的各类汉字信息资源了,但在下传Internet上各种编码的汉字文件时,国内很多用户都希望能用国标码进行存贮,以方便日后的交流和使用的。这种不同汉字编码之间的转换问题可利用Internet上的一些转换软件来实现。这类转换软件的存取地址是:

ftp://ftp.ifcss.org/pub/software/dos/convert/hz-20.zip

ftp://ftp.ifcss.org/pub/software/dos/convert/hc-30.zip

hz-20.zip解压缩后可生成GB2HZ和HZ2GB两个可执行文件,可用来实现国标码到HZ码、HZ码到国标码的转换,转换命令格式分别为:

GB2HZ            国标码源文件        HZ码目标文件

HZ2GB            HZ码源文件            国标码目标文件

hc-30.zip解压缩后,须运行一次hcsetup安装程序,以便正确生成G2B和B2G两个批处理文件,然后键入以下两个命令即可实现国标码与大五码间的相互转换:

G2B        国标码源文件        大五码目标文件

B2G        大五码源文件        国标码目标文件

此外,国标码和大五码之间的相互转换也可利用新天地公司的中文之星等软件来实现,感兴趣的读者可以自己动手试一试。

### 4 结 束 语

我国联入Internet的时间还不长,很多用户对Internet上的汉字编码和中文软件还不十分熟悉,因此,应加强这方面的理论研究和实践探索。

本文所介绍的部分中文软件虽然有这样或那样的不足之处,但这些软件却可帮助国内广

大 Internet 用户即刻用上 Internet 上的各类汉字信息资源,应该说,这些软件还是非常有价值的,值得广大用户学习、使用。相信随着各种中文软件的进一步完善,Internet 上汉字信息资源的使用将变得更加简便易行。

### 参 考 文 献

- [1] 魏亚桂:漫谈中文编码,New sgroup 论文  
[2] 张 钧:如何在网络上直接阅读中文 E-MAIL? New sgroup 论文

(责任编辑 孙月湘)

## 新信息环境下学科建设研讨会征文通知

由中国科技情报学会和中国国防科技信息学会主办,中国科技情报学会理论方法委员会和中国国防科技信息学会学术委员会承办的新信息环境下学科(情报、信息)建设研讨会,拟于1997年10月在浙江省富阳市召开。现将有关征文事项通知如下:

**主题:** 新信息环境下学科(情报、信息)的建设与发展

- 内容:**
- 1 新的信息环境对学科建设与发展的影响;
  - 2 学科建设如何促进我国信息服务业改革与发展;
  - 3 国内外学科建设最新进展评述;
  - 4 发展学科基础理论的基本思路;
  - 5 与主题有关的其它内容。

请论文作者于1997年8月31日前将论文寄会议筹备组。若论文入选,将函请作者与会。

**联系人:** 陈昭楠、庞景安、霍忠文、时文生

**单 位:** 中国科技信息研究所

**地 址:** 北京市复兴路15号 邮 编: 100038

**电 话:** (010)68515544- 2307

**传 真:** 68514025

**电子邮箱:** huozw @N S DST N. CANET. CN

中国科技情报学会理论方法委员会  
中国国防科技信息学会学术委员会