


转载 shanluan_ 于 2011-12-20 13:29:26 发布 1939 收藏 1

分类专栏： 编程之路 文章标签： diacritics 扩展 forms microsoft shapes extension

 编程之路 专栏收录该内容

Unicode 到目前为止所定义五个平面中,第0平面(BMP)最为重要. 其编码分布如下.

注：中文范围 4E00-9FBF：CJK 统一表意符号 (CJK Unified Ideographs)

0000-007F：C0控制符及基本拉丁文 (C0 Control and Basic Latin)

0080-00FF：C1控制符及拉丁文补充-1 (C1 Control and Latin 1 Supplement)

0100-017F：拉丁文扩展-A (Latin Extended-A)

0180-024F：拉丁文扩展-B (Latin Extended-B)

0250-02AF：国际音标扩展 (IPA Extensions)

02B0-02FF：空白修饰字母 (Spacing Modifiers)

0300-036F：结合用读音符号 (Combining Diacritics Marks)

0370-03FF：希腊文及科普特文 (Greek and Coptic) 0400-04FF：西里尔字母 (Cyrillic)

0500-052F：西里尔字母补充 (Cyrillic Supplement)

0530-058F：亚美尼亚语 (Armenian)

0590-05FF：希伯来文 (Hebrew)

0600-06FF：阿拉伯文 (Arabic)

0700-074F：叙利亚文 (Syriac)

0750-077F：阿拉伯文补充 (Arabic Supplement)

0780-07BF：马尔代夫语 (Thaana)

07C0-077F：西非书面语言 (N'Ko)

0800-085F：阿维斯塔语及巴列维语 (Avestan and Pahlavi)

0860-087F：Mandaic

0880-08AF：撒马利亚语 (Samaritan)

0900-097F：天城文书 (Devanagari)

0980-09FF：孟加拉语 (Bengali)

0A00-0A7F：锡克教文 (Gurmukhi)

0A80-0AFF：古吉拉特文 (Gujarati)

0B00-0B7F：奥里亚文 (Oriya)

0B80-0BFF：泰米尔文 (Tamil)

0C00-0C7F：泰卢固文 (Telugu)

0C80-0CFF：卡纳达文 (Kannada)

0D00-0D7F：德拉维族语 (Malayalam)

0D80-0DFF：僧伽罗语 (Sinhala)

0E00-0E7F：泰文 (Thai)

0E80-0EFF：老挝文 (Lao)

0F00-0FFF：藏文 (Tibetan)

1000-109F：缅甸语 (Myanmar)

10A0-10FF：格鲁吉亚语 (Georgian)

1100-11FF：朝鲜文 (Hangul Jamo)

1200-137F：埃塞俄比亚语 (Ethiopic)

1380-139F：埃塞俄比亚语补充 (Ethiopic Supplement)

13A0-13FF：切罗基语 (Cherokee)

1400-167F：统一加拿大土著语音节 (Unified Canadian Aboriginal Syllabics)

1680-169F：欧甘字母 (Ogham)

16A0-16FF：如尼文 (Runic)

1700-171F：塔加拉语 (Tagalog)

1720-173F：Hanunóo

1740-175F：Buhid

1760-177F：Tagbanwa

1780-17FF：高棉语 (Khmer)

1800-18AF：蒙古文 (Mongolian)



Microsoft Word

分类专栏

0 订阅 15 篇文章

订阅专栏

18B0-18FF：Cham
1900-194F：Limbu
1950-197F：德宏泰语 (Tai Le)
1980-19DF：新傣仂语 (New Tai Lue)
19E0-19FF：高棉语记号 (Kmer Symbols)
1A00-1A1F：Buginese
1A20-1A5F：Batak
1A80-1AEF：Lanna
1B00-1B7F：巴厘语 (Balinese)
1B80-1BB0：巽他语 (Sundanese)
1BC0-1BFF：Pahawh Hmong
1C00-1C4F：雷布查语 (Lepcha)
1C50-1C7F：Ol Chiki
1C80-1CDF：曼尼普尔语 (Meithei/Manipuri)
1D00-1D7F：语音学扩展 (Phonetic Extensions)
1D80-1DBF：语音学扩展补充 (Phonetic Extensions Supplement)
1DC0-1DFF：结合用读音符号补充 (Combining Diacritics Marks Supplement)
1E00-1EFF：拉丁文扩充附加 (Latin Extended Additional)
1F00-1FFF：希腊语扩充 (Greek Extended)
2000-206F：常用标点 (General Punctuation)
2070-209F：上标及下标 (Superscripts and Subscripts)
20A0-20CF：货币符号 (Currency Symbols)
20D0-20FF：组合用记号 (Combining Diacritics Marks for Symbols)
2100-214F：字母式符号 (Letterlike Symbols)
2150-218F：数字形式 (Number Form)
2190-21FF：箭头 (Arrows)
2200-22FF：数学运算符 (Mathematical Operator)
2300-23FF：杂项工业符号 (Miscellaneous Technical)
2400-243F：控制图片 (Control Pictures)
2440-245F：光学识别符 (Optical Character Recognition)
2460-24FF：封闭式字母数字 (Enclosed Alphanumerics)
2500-257F：制表符 (Box Drawing)
2580-259F：方块元素 (Block Element)
25A0-25FF：几何图形 (Geometric Shapes)
2600-26FF：杂项符号 (Miscellaneous Symbols)
2700-27BF：印刷符号 (Dingbats)
27C0-27EF：杂项数学符号-A (Miscellaneous Mathematical Symbols-A)
27F0-27FF：追加箭头-A (Supplemental Arrows-A)
2800-28FF：盲文点字模型 (Braille Patterns)
2900-297F：追加箭头-B (Supplemental Arrows-B)
2980-29FF：杂项数学符号-B (Miscellaneous Mathematical Symbols-B)
2A00-2AFF：追加数学运算符 (Supplemental Mathematical Operator)
2B00-2BFF：杂项符号和箭头 (Miscellaneous Symbols and Arrows)
2C00-2C5F：格拉哥里字母 (Glagolitic)
2C60-2C7F：拉丁文扩展-C (Latin Extended-C)
2C80-2CFF：古埃及语 (Coptic)
2D00-2D2F：格鲁吉亚语补充 (Georgian Supplement)
2D30-2D7F：提非纳文 (Tifinagh)
2D80-2DDF：埃塞俄比亚语扩展 (Ethiopic Extended)
2E00-2E7F：追加标点 (Supplemental Punctuation)
2E80-2EFF：CJK 部首补充 (CJK Radicals Supplement)
2F00-2FDF：康熙字典部首 (Kangxi Radicals)
2FF0-2FFF：表意文字描述符 (Ideographic Description Characters)
3000-303F：CJK 符号和标点 (CJK Symbols and Punctuation)
3040-309F：日文平假名 (Hiragana)
30A0-30FF：日文片假名 (Katakana)
3100-312F：注音字母 (Bopomofo)
3130-318F：朝鲜文兼容字母 (Hangul Compatibility Jamo)
3190-319F：象形字注释标志 (Kanbun)



unicode

31A0-31BF：注音字母扩展 (Bopomofo Extended)
31C0-31EF：CJK 笔画 (CJK Strokes)
31F0-31FF：日文片假名语音扩展 (Katakana Phonetic Extensions)
3200-32FF：封闭式 CJK 文字和月份 (Enclosed CJK Letters and Months)
3300-33FF：CJK 兼容 (CJK Compatibility)
3400-4DBF：CJK 统一表意符号扩展 A (CJK Unified Ideographs Extension A)
4DC0-4DFF：易经六十四卦符号 (Yijing Hexagrams Symbols)
4E00-9FBF：CJK 统一表意符号 (CJK Unified Ideographs)
A000-A48F：彝文音节 (Yi Syllables)
A490-A4CF：彝文字根 (Yi Radicals)
A500-A61F：Vai
A660-A6FF：统一加拿大土著语音节补充 (Unified Canadian Aboriginal Syllabics Supplement)
A700-A71F：声调修饰字母 (Modifier Tone Letters)
A720-A7FF：拉丁文扩展-D (Latin Extended-D)
A800-A82F：Syloti Nagri
A840-A87F：八思巴字 (Phags-pa)
A880-A8DF：Saurashtra
A900-A97F：爪哇语 (Javanese)
A980-A9DF：Chakma
AA00-AA3F：Varang Kshiti
AA40-AA6F：Sorang Sompeng
AA80-AADF：Newari
AB00-AB5F：越南傣语 (Víṭt̚ Thái)
AB80-ABAO：Kayah Li
AC00-D7AF：朝鲜文音节 (Hangul Syllables)
D800-DBFF：High-half zone of UTF-16
DC00-DFFF：Low-half zone of UTF-16
E000-F8FF：自行使用区域 (Private Use Zone)
F900-FAFF：CJK 兼容象形文字 (CJK Compatibility Ideographs)
FB00-FB4F：字母表达形式 (Alphabetic Presentation Form)
FB50-FDFF：阿拉伯表达形式A (Arabic Presentation Form-A)
FE00-FE0F：变量选择符 (Variation Selector)
FE10-FE1F：竖排形式 (Vertical Forms)
FE20-FE2F：组合用半符号 (Combining Half Marks)
FE30-FE4F：CJK 兼容形式 (CJK Compatibility Forms)
FE50-FE6F：小型变体形式 (Small Form Variants)
FE70-FEFF：阿拉伯表达形式B (Arabic Presentation Form-B)
FF00-FFEF：半型及全型形式 (Halfwidth and Fullwidth Form)
FFF0-FFFF：特殊 (Specials)

GB13000 UCS-2格式转UTF-8 10-17
身份证读卡器生成的基本信息TEXT文件，由于是采用 GB 13000 的 UCS-2 编码格式，java读取出来是乱码，这个段代码就是解决转码问题的，亲测可用



请发表有价值的评论， 博客评论不欢迎灌水，良好的社区氛围需大家一起维护。

抢沙发



评论

【编码】unicode、ucs-2、ucs-4、utf-16、utf-32、utf-8 介绍 2-3
其实BOM是个小聪明的想法。由于UCS-2没有定义U+FFFE,因此只要出现 FF FE 或者 FE FF 这样的字节序列,就可以认为它是U+FEFF,并且可以判断出是...

编码方式部分信息整合:Unicode、UCS-2/4、UTF-8/16/32、GB2312、GBK 1-27
UCS-2和UTF-16也可以理解为和ASCII以及ISO-8859-1兼容,在ASCII编码或者ISO-8859-1编码的每个字节前加上0x00,就得到相应字符的UCS-2编码。UCS-...

UCS-2编码与UTF-8编码 热门推荐 BlackStone的博客 1万+
我们都知道对于UNICODE来说，UCS-2是内码，而UTF-8则是它的实现方式。每一个字节都有8个位，而对于UTF-8来说，每一个字节的前两位尤为重要，...

UCS-2转换为utf-8代码 05-21
将ucs-2转换为utf-8代码集完整版

java ucs 2,【字符编码系列】JavaScript使用的编码-UCS-2 1-27
本文属于 字符编码系列文章之一,更多请前往 字符编码系列。 在JavaScrip中,进行一些GBK或者UTF-8编码的字符操作时,打印出来的经常是乱码,其原因就...

shanluan_

码龄10年 暂无认证

12
原创

911
积分

私信

关注

搜博主文章

热门文章

Unicode 汉字内码表 81521

如何在HTML中引用jQuery函数库

如何在代码中为组件设置dip, sp值

用NDOC自动生成.NET文档 2476

网络舆情监控系统的爬虫子系统开发总结

最新评论

lucene入门简单实现

www-文子: 学习了

如何在HTML中引用jQuery函数库

_codedreamer: 你好！可以帮我看一下我下面的代码为何使用不了Jquery库中的函数...

如何在HTML中引用jQuery函数库

Kingsly39664: 其实我是想问的是，html页面怎么利用request对象吗？可以吗？比如...

用NDOC自动生成.NET文档

paulwinflo: 这个还能用么 我在开发Windows 8 使用工具为VS2012 语言为C#

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

你必须知道的10个提高Canvas性能技巧

Android中dip、dp、sp、pt和px的区别

如何在代码中为组件设置dip, sp值

2014年 8篇	2013年 5篇
2012年 14篇	2011年 5篇

万字字符编码之UCS-2与Utf-8_imxiangzi的专栏_ucs2和utf8		19万+	1-14
周排名 当前,Unicode深入人心,且UTF-8大行其道,Unicode基本被等同于UTF-16,UTF-32,所以目前UCS基本淡出人们的视野中。(Windows NT用的就是U...		总排名	1-14
8汉字UCS-2编码表.doc 最新发布	6 获赞	4 评论	5 收藏 09-15
java 文字信息采用GB 13000的UCS-2进行存储(身份证读卡器)			09-09
文字信息采用GB 13000的UCS-2进行存储,各项目分配如下: 项目长度(字节)说明 姓名30汉字 性别2代码 民族4代码 出生16年月日:YYMMDD 住...			
Linux 和 Windows 平台不同的 UCS-2 编码_路漫漫其修远...			2-8
以下是有关两个平台 UCS-2 编码的潜规则:1,UCS-2 不等于 UTF-16。UTF-16 每个字节使用 ASCII 字符范围编码,而 UCS-2 对每个字节的编码可以超出 A...			
ASCII、Base64、Unicode(UCS-2、UTF-8、Little endian...			1-14
Unicode编码指的是UCS-2编码方式,直接用两个字节存入字符的Unicode码。这个选项用的little endian格式(这种情况下如果一个字符的unicode超过2个字...			
Unicode Basic			weixin_30349597 的博客 79
转载: Unicode详解 2008-08-30 13:41 Unicode详解 版权声明: 可以任意转载,但转载时必须标明原作者charlee、原始链接http://tech.idv2.com/2008/02/...			
GB13000 中的UCS-2 怎么转换?			03-23
一段经过GB13000 转意的字符存进txt文本中,然后再取出来时怎么给转换回去,我是用vs2008 winform下开发的,顺便再问一下怎么从txt文本中取出字段...			
对UCS-2进行解码(笔记)			天高任鸟飞 3178
1、UCS-2是16位的,进行解码时,UTF-16BE 和 UTF-16LE charset 忽略字节顺序标记;进行编码时,不写入字节顺序标记。2、解码时将字节流转换为...			
字符编码之UCS-2与Utf-8			weixin_30279671 的博客 279
很多操作系统都直接支持utf-8字符串操作,只有MS这个异类用的Unicode,就是所谓的ucs-2 如果写关于跨平台的代码,那么避免不了要做编码转化 这里...			
ucs2编码 c语言,科普unicode与ascii			weixin_28839629 的博客 37
该楼层疑似违规已被系统折叠隐藏此楼查看此楼1.多语言字符集和编码相关的参考规范之中最重要的有两个,一个是ISO/IEC 10646,一个是Unicode。...			
转帖——再谈编码			weixin_34237596 的博客 128
以下是本人从网上找的,旨在让程序员能真正掌握编码以及解决相关的问题谈谈Unicode编码,简要解释UCS、UTF、BMP、BOM等名词 这是一篇程序员写...			
UCS2编码与解码			weixin_30794851 的博客 233
/*-----"/// UCS2编码// 输入: pSrc - 源字符串指针// nSrcLength - 源字符串长度// 输出: pDst - 目标编码串指...			
Linux 和 Windows 平台不同的 UCS-2 编码			weixin_34355715 的博客 144
2019独角兽企业重金招聘Python工程师标准>>> ...			
c语言 ucs2编码,科普unicode与ascii			weixin_42398130 的博客 169
该楼层疑似违规已被系统折叠隐藏此楼查看此楼1.多语言字符集和编码相关的参考规范之中最重要的有两个,一个是ISO/IEC 10646,一个是Unicode。...			
计算机ucs汉字编码,字符编码基础知识 - Unicode,UCS,GBK,GB2312,UTF-8			weixin_32506761 的博客 926
字符编码基础知识- Unicode、UCS、GBK、GB2312、UTF-8最近遇到一个Unicode和UTF-8关系比较的问题,之前在处理中文显示时也遇到过类似的问题,于是...			
java gb13000 ucs2_采用GB 13000的UCS-2进行存储的文件怎么转换			weixin_30124601 的博客 155
不是sql中的数据吧你那个是单个文件吗?假设文件头采用标准UCS2格式的两个字节,每个字段的数据是通过\分隔的,每行文字是一条记录,如果有不同...			
字符串UCS-2格式转码(C#)			fzhptr 的专栏 2146
 实现功能: (1) 把个人信息写到智能卡中。(2) 从智能卡读出个人信息。 实现方法: 将界面上的个人信息数据提取出来,然后转换为一个由...			
一个txt文件GB 13000的UCS-2进行存储,怎么读取出来转成string型的呢???			03-30
谢谢了!! 比较的着急!!			
浅谈 GB13000			weixin_34279579 的博客 629
引言 我前几天写了一篇随笔"随机生成常用汉字(再次改进版)".园友"菜菜灰"评论说"看着太累,不认识的字还是多,代码也很多,几段代码显示2万多汉字...			
©2022 CSDN 皮肤主题: 大白 设计师: CSDN官方博客 返回首页			

shanluan_ 关注

0 0 1



专栏目录

务 中国互联网举报中心 Chrome商店下载 ©1999-2022北京创新乐知网络技术有限公司 版权与免责声明 版权申诉 出版物

