

编码知识学习笔记之一

原创

littletigerat



于 2011-08-04 10:49:57 发布



1688



收藏

版权

分类专栏:

协议分析

Java

Android

文章标签:

transformation

character

internet

文档

basic

工作



协议分析

同时被 3 个专栏收录 ▾

0 订阅

7 篇文章

订阅专栏

编码知识学习笔记之一

一． 有哪些编码

1. ANSI
2. Unicode 🔍
3. Unicode big Endian
4. Unicode - ASCII Escaped
5. UTF-7
6. UTF-8
7. UTF-16
8. UTF-8 - 无 BOM
9. UTF-16 - 无 BOM
10. UTF-16 - Big Endian
11. UTF-16 - Big Endian - 无 BOM
12. GBK

二． Unicode是什么的缩写

Universal Multiple-Octet Coded Character Set

三． Unicode是什么

Unicode 是一种字符编码方法，它是由国际组织设计，可以容纳全世界所有语言文字的编码方案。Unicode的全称是"Universal Multiple-Octet Coded Character Set"，简称为UCS。UCS可以看作是"Unicode Character Set"的缩写。

四． Unicode的历史

目前存在两个试图独立设计Unicode的组织：

1. 国际标准化组织（ISO）
2. 一个软件制造商的协会（unicode.org）

各自研发的项目

ISO开发了ISO 10646项目

Unicode协会开发了Unicode项

在1991年前后，双方都认识到世界不需要两个不兼容的字符集。于是它们开始合并双方的工作成果，并为创建一个单一编码表而协同工作。从Unicode2.0开始，Unicode项目采用了与ISO 10646-1相同的字库和字码。

五．Unicode的最新标准

目前两个项目仍都存在，并独立地公布各自的标准。

1.Unicode协会现在的最新版本是2005年的Unicode 4.1.0。

2.ISO的最新标准是10646-3:2003。

六．UCS是什么的缩写

Unicode Character Set

七．UTF是什么的缩写

UCS Transformation Format

八．UCS与UTF的关系是怎样的？

UCS规定了怎么用多个字节表示各种文字。

怎样传输这些编码，是由UTF(UCS Transformation Format)规范规定的

小结：

UCS：表示各种文字

UTF：传输这些文字

九．UTF有哪些常见规范（也就是编码格式）

1.UTF-7

2.UTF-8

3.UTF-16

十．UTF的描述UTF-8和UTF-16编码的规范文档是什么？

IETF 的RFC2781和RFC3629以RFC的一贯风格，清晰、明快又不失严谨地描述了UTF-16和UTF-8的编码方法。

RFC2781描述UTF-16编码规范

RFC3629描述UTF-8编码规范

十一．IETF是什么缩写

Internet Engineering Task Force

十二. IETF组织的作用是什么

IETF负责维护的RFC

十三. RFC文档的作用是什么

RFC是Internet上一切规范的基础

十四. Unicode与ASCII, 以及GB的兼容性关系

Unicode只与ASCII兼容 (更准确地说, 是与ISO-8859-1兼容), 与GB码不兼容

十五. UCS两种格式

UCS-2和UCS-4

十六. UCS-2和UCS-4的区别

1.UCS-2就是用两个字节编码

2.UCS-4就是用4个字节 (实际上只用了31位, 最高位必须为0) 编码

十七. UCS-2和UCS-4各自有多少个码位

UCS-2有 $2^{16}=65536$ 个码位, UCS-4有 $2^{31}=2147483648$ 个码位。

十八. 什么叫UCS-4 BMP

UCS-4根据最高位为0的最高字节分成 $2^7=128$ 个group。每个group再根据次高字节分为256个plane。每个plane根据第3个字节分为256行(rows), 每行包含256个cells。当然同一行的cells只是最后一个字节不同, 其余都相同。

也就是说:

UCS-4是4个字节, 从高到低, 从第1个字节到第4个字节,

第1个字节是高字节, 第4个字节是低字节

第1个字节分成128个group

第2个字节分成256个plane

第3个字节分成256个rows

第4个字节分成256个cells

group 0的plane 0被称作Basic Multilingual Plane, 即BMP。或者说UCS-4中, 高两个字节为0的码位被称作BMP。

将UCS-4的BMP去掉前面的两个零字节就得到了UCS-2。在UCS-2的两个字节前加上两个零字节, 就得到了UCS-4的BMP。而目前的UCS-4规范中还没有任何字符被分配在BMP之外。

html基础[知识学习](#)笔记分享

04-09

对html的一个基础[知识](#)的了解, 相比后端中包含大量的算法之类的[编码](#)。前端的编写明显的简单了很...

emule中的字符串转换

云淡风轻 815

emule中的字符串转换主要在stringconversion.h和stringconversion.cpp中实现, 其实掌握了utf-8和utf-...

UTF-8 8-bit Unicode Transformation Format 万国码_一秒变桌子的博客... 8-12

UTF-8(8-bitUnicodeTransformation Format)是一种针对Unicode的可变长度字符**编码**,又称万国码,由Ken...

C#高性能大容量SOCKET并发(七):协议字符集_SQLDebug_Fan的博客 7-16

UTF-8是UNICODE的一种变长字符**编码** 又称万国码,由Ken Thompson于1992年创建。现在已经标准化为...

字符**编码** UTF-8 **学习笔记** kfpiza的博客 147

百度百科说UTF-8最大支持4字节,但看其特点, 支持6字节也是可以的一字节时和ascii , iso8859-1 , gbk 是...

RFC3629 lyclolevel的专栏 2244

Network Working Group F. Yergeau Request for Comments: 3629 Alis Technol

精述字符**编码**(读这篇就够了)_latin_1恋喵大鲤鱼的博客 8-11

UTF-8(8-bit Unicode Transformation Format)是一种针对Unicode的可变长度字符**编码**,也是一种前缀码,...

IOT-MQTT协议-简介_leeahuamsng的博客 7-30

Bradner,S。“用于RFC指示需求水平的关键词”,BCP 14,RFC 2119,1997年3月.http://www.ietf.org/rfc/rfc...

UCS-2、UCS-4 gao_zhennan的博客 2193

汉字常用**编码**格式 为了在屏幕上显示字符。需要下面几个步骤: 制作所有字符对应的字模。比如大写字...

谈谈Unicode**编码**(2) 青云直上不为仙 902

3、UCS-2、UCS-4、BMP UCS有两种格式: UCS-2和UCS-4。顾名思义, UCS-2就是用两个字节**编**...

物联网传输协议MQTT研究_草根大哥的博客 8-4

在一个UTF-8字符数据**编码**的字符串必须是格式良好的UTF-8的Unicode规范[UNICODE]定义和重述的RF...

编码相关的问题_zhengudaoer的博客 7-30

IETF的RFC2781和RFC3629以RFC的一贯风格,清晰、明快又不失严谨地描述了UTF-16和UTF-8的**编码**方...

字符**编码**(ucs2 ucs4 utf) dolphin98629的专栏 1859

源地址: http://blog.charlee.li/unicode-intro/ 最近一直在忙点私活, 又好久没写blog了, 再不写点的话...

下载的附件名总乱码? 你该去读一下 RFC 文档了! Java笔记虾 643

纸上得来终觉浅, 绝知此事要躬行Web 开发过程中, 相信大家都遇到过附件下载的场景, 其中, 各浏...

安全: Web安全**学习笔记** 02-26

一直没有真正系统的**学习**web安全**知识** (认证和授权除外), 这个月看了一本《Web安全设计之道》, ...

HTML5**学习笔记**之html5与传统html区别 12-14

一. HTML5语法的改变 该**知识**点所说变化指的是基于HTML4基础上所定义的改变, 主要有如下: 1.HT...

leetcode部分排序类-coding-interview-notes::pencil:我的**编码**面试**学习笔记** 06-30

编码面试可能是程序员压力最大的面试之一。这个 repo 的灵感来自于面试准备。凯文的建议有四个步...

lrucacheleetcode-Notes:有关**编码**、数据库、架构、DevOps等的动手**学习笔记** 06-29

算法是一种解决问题的思路和方法, 从长期来看, 大脑思考能力是个人最重要的核心竞争力, 而算法是...

MQTT V3.1.1 协议 规范 Xanthium 3242

目录 1.简介 1.1术语 1.2 数据表示 1.2.1 位 1.2.2整数数据值 1.2.3 UTF-8**编码**的字符串 2 MQTT控制包格...

UTF-8 IT老兵的驿站 1366

摘自: https://zh.wikipedia.org/wiki/UTF-8, 个别地方有几个特殊符号没被处理, 显示成层叠状态。维...

Unicode Basic weixin_30349597的博客 153

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...

utf-8编码的字符串转成unicode(ucs-4)编码的字符串 papaya的博客 1250
本人第一篇原创博客，尽力把代码注释清楚 typedef unsigned int mchar_t; // 无符号的4字节，用于表...

Unicode, UTF-8, UTF-16, UTF-32, UCS-2, UCS-4是什么? flagyili的博客 618
Unicode与ISO 10646 全世界很多个国家都在为自己的文字编码，并且互不想通，不同的语言字符编码...

ASCII, Unicode, UTF-8, UCS-2, UCS-4等等编码知识集结 皮皮君的博客 610
一篇简洁很好的文章：阮一峰 字符编码笔记：ASCII, Unicode 和 UTF-8 一篇讲unicode的文章: charle...

python ucs-2与ucs-4编码方式判断 小白成神路 3065
一、判断当前python的编码方式：以 -enable-unicode=ucs4 编译则：>>> import sys >>> print sys.ma...

zynq-linux移植学习笔记之 最新发布 06-28
回答1: "Zynq-Linux移植学习笔记"是一篇关于在Zynq SoC平台上移植Linux操作系统的学习笔记。...

“相关推荐”对你有帮助么？

非常没帮助 没帮助 一般 有帮助 非常有帮助

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
公安备案号11010502030143 京ICP备19004658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



littletigerat

专家

码龄14年

暂无认证

562	1万+	38万+	297万+	等级
原创	周排名	总排名	访问	
2万+	1449	842	964	318
积分	粉丝	获赞	评论	收藏



私信

关注

搜博主文章



热门文章

24段魔尺，可以折出哪些精美图案 170538

2014年24段魔尺变三叶花视频教程 150099

为什么要报考系统架构设计师考试 85210

程序员很少上《非诚勿扰》电视节目相亲之分析 59252

2018年CSDN博客排名第一名，花落谁家？ 58890

分类专栏

	松滋历史	2篇
	解析项目管理师试题	91篇
	解析系统分析师试题	29篇
	系统架构师	94篇
	软考天下	46篇
	通过率	1篇

最新评论

巨波公第3子登国公后裔在荆州(巨波公6子...
m0_56898650: 三：登国公 字陶发，号桂用，妣郭氏。进士出身，官四川重庆知府...
软考证到底有多大个鸟用？
Aylan_: 并没有啥用
明朝开国第一重臣李善长长子李祺之长子...
littletigerat: 为什么取“兰” 取自：义结金兰
朱元璋与李善长关系：义结金兰。
为什么要报考系统架构设计师考试
辣椒油li: 工作一年做的是硬件工程师，我的
项目经验在系统架构设计师考试里是不是...
明朝开国第一重臣李善长长子李祺之长子...
Lanian: 但为什么改姓取个“兰”字呢？ 我们
家就是松滋兰氏，但家里老人都不清楚祖...

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

松滋历史（2）-高成县为什么这么高光？
松滋历史（1）-松滋县建县历史考证
这么学，好玩，有趣，逗死了！（2017年系统规划与管理师上午综合知识试题试题解析

(六))

2021年 2篇	2020年 17篇
2019年 7篇	2018年 51篇
2017年 147篇	2016年 3篇
2015年 9篇	2014年 37篇
2013年 29篇	2012年 73篇
2011年 80篇	2010年 110篇
2009年 8篇	

目录

- 一. 有哪些编码
- 二. Unicode是什么的缩写
- 三. Unicode是什么
- 四. Unicode的历史
- 五. Unicode的最新标准
- 六. UCS是什么的缩写
- 七. UTF是什么的缩写
- 八. UCS与UTF的关系是怎样的?
- 九. UTF有哪些常见规范 (也就是编码格...
- 十. UTF的描述UTF-8和UTF-16编码的规...
- 十一. IETF是什么缩写
- 十二. IETF组织的作用是什么
- 十三. RFC文档的作用是什么
- 十四. Unicode与ASCII, 以及GB的兼容...