



starxhong

碼齡6年

53

原創

124

粉絲

165

獲贊

162

評論

35萬+

訪問

3694

積分

266

收藏

1萬+

週排名

1萬+

總排名

等級

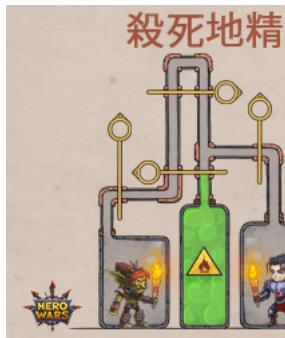


TA的主頁



私信

關注



最新文章

[ssh遠程免密訪問配置](#)[git submodule使用方法](#)[推薦系統中的最小樣本量和參數估計](#)[shell向上取整](#)[tensorflow各個版本的CUDA以及Cudnn版本對應關係](#)

分類專欄



ssh

1篇



遠程

1篇



git

1篇



VC++

27篇



Qt

8篇



機器學習

6篇



歸檔

2020

3月

2篇

2月

1篇

1月

1篇

2019

11月

1篇

10月

1篇

4月

1篇

2018

12月

1篇

11月

1篇

9月

7篇

8月

11篇


熱門文章

航空遙感圖像(Aerial Images)目標檢測數據集匯總  45499

pycharm加入import路徑  25987

Qt入門——選擇路徑打開文件並讀取文件  19723

Python庫numpy中的Broadcasting機制解析  16043

spark踩坑記——windows環境下spark安裝和運行  15270

最新評論

Faster R-CNN/R-FC...
s953926: [reply]m0_46612038[/reply]請問你解決了嗎?

Faster R-CNN/R-FC...
hongxingabc: [reply]wen_zihan[/reply]是，只用6類的訓練數據就可以了

Faster R-CNN/R-FC...
wen_zihan: 博主你好，請問如果測試部分的數據集只包含六個pascal voc的子類，那在最開始! ...

Faster R-CNN畫PR曲線
weixin_44732102: [reply]liqian007[/reply]您好我也出現了和你一樣的問題請問你當初是怎麼! ...

航空遙感圖像(Aerial Ima...
qq_43360419: 博主，我是做圖像配準，這些數據集可以用嗎

目錄

[背景](#)

[UTF-8 介紹](#)

[設計思路](#)

[string 與wstring 的轉換](#)

[轉換版本一](#)

[轉換版本二](#)

[轉換版本三](#)

QQ客服



 kefu@csdn.net

 客服論壇

400-660-0108

工作時間8:30-22:00

[關於我們](#) [招聘](#) [廣告服務](#) [網站地圖](#)

京ICP備19004658號 [經營性網站備案信息](#)

 公安備案號11010502030143

京網文〔2020〕1039-165號

Unicode與utf-8相互轉化)

9218 收藏 2 ☆

展開

請點擊：<https://blog.csdn.net/jjc918/article/details/52200478>

C++ 實現Java的hashCode函數時遇到編碼的問題。原文如下：

然而，不幸的我們接到命令，要在Linux 下支持對文案進行文案超長截斷處理。這樣的話應該怎麼做

是UTF-8編碼的。如果在本地的話可以通過本地環境配置來保證。命令行下運行locale命令，LC_CTYPE命令，應該有一行是fileencoding=utf-8。這樣我們就有了工作的基礎。

是一種變長編碼，對於一個Unicode 的字符編碼成1 至4 個字節。我們可以認為，在UTF-8 中，英文是

那我們可能想用這樣一個方案：遍歷字符串，判斷當前字節屬於中文還是英文，如果英文則對長度加1則對長度加一併跳到後面第三個字節繼續處理。達到我們需要的文案長度時，break 跳出循環，返回當

力，程序容易寫出問題。而且我們前面的假設畢竟是一般情況下（雖然概率很低），如果出現一個四字

字符都佔據相同長度，那我們的處理就會簡單多了。這時候我們想到了C++的wstring類型，wstring符個數。wstring與string一樣都是基於basic_string類模板，不同的是string使用char為基本類型，而Unicode字符的存儲，在Win下是兩個字節，Linux的實現則是四個字節，可以直接用sizeof(wchar_t)查看

tring 和wstring 的互相轉換，並用wstring 來判斷字符個數，在超長時進行截斷。

麼可以採用下面的寫法，這是最好的：

```
string& str)

::codevt_utf8<wchar_t>;
ert_typeX, wchar_t> converterX;

tes(str);

string& wstr)
```

```
= std::codecvt_utf8<wchar_t>;
ert_typeX, wchar_t> converterX;

s(wstr);
```

提供的對string 和wstring 的轉換，對Unicode 進行了語言和庫級別的支持。但這一特性在gcc/g++

— [stackoverflow](#)

性，那麼第二個版本可以用unique_ptr 來管理內存，這樣可以避免直接操作指針的尷尬，程序更加安

```
string& str) {

    1;
    UTF-8");
    p(new wchar_t[len]);
    (), len);
    ;

string& w_str) {

    () * 4 + 1;
    S.UTF-8");
    (new char[len]);
    _str(), len);
```

_t 为 4 个字节，对于 s2ws，wstring 的长度肯定小于等于 string 的长度，而对 ws2s，string 的长度-1 是预留给字符串的结束符 ‘\0’ 。

可以在命令行用 locale 查看当前系统的语言环境设置，LC_CTYPE 指语言符号及其分类 。网上很多版本二个参数用空字符串，会使用系统当前默认的 locale 设置。但是这样有个问题，也许你写出来的程序在服务器上的 locale 不一定是 utf8，所以这里要强制设置为 en_US.UTF-8。

中对多字节字符串和宽字符串的互相转换函数，依赖于当前 locale 中所指定的字符编码。

就只能使用下面的 new/delete 了。

```
string& str) {
```

```
1;  
UTF-8");  
n];  
n);
```

```
string& w_str) {
```

```
( ) * 4 + 1;  
S.UTF-8");  
  
, len);
```

下来的处理就很简单了。实现处理函数 FormatText，然后加入 main 函数测试，完整代码如下：

```
10;  
  
string& str) {
```

```
1;  
  
n];  
n);
```

```
string& w_str) {
```

```
( ) * 4 + 1;  
  
, len);
```

```
txt) {
```

```
txt << std::endl;  
txt);  
" << w_txt.size() << std::endl;
```

```
<< (*txt).size() << std::endl;43 | if (w_txt.size() > kTextSize) {
TextSize);
```

```
xt << std::endl;
```

```
好玩等你";
```

```
xt);
```

```
t;
```

```
;
```

```
爆...";
```

文结束

ong/p/3785791.html 这种方法是通过比编码大小但方式将中英文分开，显得生硬，但在某些特殊情况
计算必须英文为1个，其余都为2个）。

/402283/stdwstring-vs-stdstring 这个stackoverflow问题帮我们理解string和wstring的区别。

1万+

关注私信

收藏2 手机看 ...

收起全文 ^

幸福稻飯丸

餓了嗎?

VISIT !

评论

文章成绝版了！ 9个月前

Just Coding!

8044

lovecj6185的专栏

344

基本知识, java, 系统软件, url, 工具软件等。 在下面的描述中,

语言，是因为它的class文件采用UTF-8，而JVM运行时使用UTF-16（至	yinbucheng的博客 591
编码集 信息最终都表示为一个二进制的字符串。每一个二进制位（bit）有0和1两	雪峰的专栏 4287
钱	
9-1,GBK,UTF-8编码及相互转换 默认是iso-8859-1.所以，request.getParameter()获得的值必须转码，除	yanjiang212的专栏 2万+
ICODE,UTF-8,GB2312,GBK) VSI、UNICODE、UTF-8、GB2312、GBK、DBCS、UCS.....是不是看的	stay hungry,stay foolish ! 781
主要包括以下几个方面：编码基本知识，java，系统软件，url，工具软件	Rachael™(©野穀子®)的专栏 938
编码集 3211首先讲一下几种字符的编码方式:1. ASCII码我们知道，在计算机内	u012179540的专栏 353
字符用一个字节进行表示，值小于等于 127。汉字用两个字节进行表示，	YellowMax 3859
22的博客 - CSDN博客	11-19
3的博客 - CSDN博客	9-30
<div><div><div>the Industry's Best Octal Industrial High Side</div><div><ul style="list-style-type: none">● Robustness and Smart Diagnostics● Reduces Power and Heat Dissipation● Flexibility● Compact 6mm x 6mm QFN Package</div></div></div>	
8859-1,GBK,UTF-8编码及相互转换 j212/article/details/2946837[/url]因为url传送默认编码是容器的编码，	WessonGao的博客 38
乌托邦 - CSDN博客	11-8
ixin_33826609的..._CSDN博客	11-16
难了 别工作的话，又有哪个机构会站出来给我推荐工作？ 笔者冷静分析多方数	CSDN学院 3万+
猎猫、圈内好友，以及年过35岁的几位老程序员.....舍了老脸去揭人家伤	启舰 46万+
isery的博客 - CSDN博客	8-29
亟(UTF-8/GBK)_运..._CSDN博客	5-5
动点赞，自动滑屏，自动切换视频 费，自动滑屏，自动切换视频代码如下auto();var	qq_40618664的博客 2万+
他们的学习方法 ;，原来成功都是有迹可循的。	敖丙 26万+
栏 - CSDN博客	10-1
§ - CSDN博客	11-23
<div><div><div>the Industry's Best Octal Industrial High Side</div><div><ul style="list-style-type: none">● Robustness and Smart Diagnostics● Reduces Power and Heat Dissipation● Flexibility● Compact 6mm x 6mm QFN Package</div></div></div>	
	沉默王二 35万+

，有什么推荐的学习网站吗？最近很浮躁，手头的一些网站都看烦了，想

见 - 蚂蚁搬家 - CSDN博客

10-28

Java成神之路 24万+

秋招中得到前东家华为（以下简称 hw）的赏识，当时秋招签订就业协

端开发，最近他们公司新来了一个技术总监，这位技术总监对技术细节很

成为别人眼中的大神 敖丙 5万+

如果你是学生可以看看，提前熟悉起来。...

沉默王二 16万+

到三线城市洛阳工作，至今已经 6 年有余。一不小心又暴露了自己的实际

钱

呈序员!

沉默王二 8万+

dotNet全栈开发 7万+

羊毛，一时没明白是怎么回事。后来上百度搜索了一下，原来真有这回

我从3K变成了40K Leo的博客 · 7万+

需要太多的知识面，然后我一个做了运维朋友告诉我大错特错，他就是从

只要AI人才，高中毕业都行！ CSDN学院 6726

司事的代码，大部分代码都写得很漂亮，严谨的同时注释也很到位，这令

ThinkWon的博客 9万+

有什么区别？什么是 Linux 内核？Linux 的基本组件是什么？Linux 的体

纯洁的微笑 9万+

，我呆过的两家互联网公司，几乎每月每天每分钟都有黑客在公司网站上

shenjian58的博客 5万+

1) 中国学生的数学成绩, 平均下来看, 会比国外好? 为什么? (2) 男生

本篇博客。

乙! CSDN学院 1万+

直接给公司带来数10亿市值蒸发损失，并引发一段“狗血恩怨剧情”，

B面人生是怎样的？
约是：为人木讷、薪水超高、工作枯燥.....然而，当离开工作岗位，撕去

ThinkWon的博客 13万+

什么是SQL? 什么是MySQL?数据库三大范式是什么mysql有关权限的表

一个老下属，跟了我很多年，做事勤勤恳恳，主动性也很好。但随着公司

沉默王二 8万+

写出一份专业的技术简历呢？我总感觉自己写的简历太烂了，所以投了无

3y 8万+

崩溃的代码注释？

无法阅遍无数的程序代码，不知道有多少次见到那些让人既感到无语又奔

九章算法的博客 3914

但人才缺口超500万...企业：无人可用！ 。先说好消息。德勤发布《全球人工智能发展白皮书》，预计到2025	CSDN学院 7132
别说大部分程序员只会写 3 年代码，恐怕就没有程序员这种职业。...	沉默王二 4万+
l? 在离职了几个月后，前公司的领导联系到他，希望他能够返聘回去，他很	江南一点雨的专栏 5万+
删除？	申城异乡人 2万+
未毕业的同学，尽力地通过校招向大公司挤，但凡挤进去，你这一生会容	启舰 2万+
之后的三大版本JVM、JRE和JDK的关系什么是跨平台性？原理是什么	ThinkWon的博客 13万+
快！你怎么看？ 他坚信 Python 可以做任何事情。其实我是不服的，因为我在某网站看	CSDN学院 6万+
卜导图才发现，我错了	龙跃十二 5万+
对的结构原理 Spring的AOP和IOC是什么？它们常见的使用场景有哪些？	做人还是低调点 2万+
么流行迟到？ 鲜事，早九晚五是常态，但有很多企业由于不打卡，员工们10点、11点才	九章算法的博客 1955
我经历了什么？ 卜)：@Conditional是做什么的?@Conditional多个条件是什么逻辑关	路人甲Java 5万+
	郑晖的博客 1万+
桌面壁纸) 、Ixml、Beautiful Soup4	Zhangguohao666的博客 2万+
2500种以上的高级语言，程序员们大呼“学到头秃”。程序员一边面临	CSDN资讯 2万+
J，30岁以上的程序员该何去何从？ 有点感慨，大龄程序猿该何去何从。...	程序猿学社的博客 16万+
ffer	敖丙 5万+
工作，简历看的不下于万份这篇文章会用实例告诉你，什么是差的程序员	启舰 10万+
职后这几题我记一辈子 的事情。我原以为从学校出来之后，除了找工作有测试外，不会有任何与	EnjoyEDU的博客 3万+
不清楚 Java 输入类 nextLine 与 next 两个方法的区别 的参数会在方法参数中直接给出，而且需要的返回值直接在方法中 return	Geffin的博客 4万+
7	3y 1万+
数最低工资最高工资人头人头百分比	juwikuang的专栏 6215
公司架构组？想不想成为项目组的负责人？想不想成为spring的高手，	路人甲Java 3万+
ual Basic .NET JavaScript PHP SQL Go语言 R语言 Assembly language Swift Ruby Objective-C Delphi/Object Pascal Unity3D	

