

Dos和Unix文本文件的区别及转换

[el/2023/9/19 14:44:18](#)

背景：

1. 由于windows和linux对换行的标识不一样，不同系统的代码传递导致代码格式的改变中可能会带来程序无法正常编译通过的问题。因此根据一些编译的错误提示，可以定位到是文件格式的问题，要对程序的文本文件进行转换。
2. 将Windows下文件传递给Linux下打开时，在Vim显示会 ^M 字符，也是由于dos文件和unix文件对换行的表示方法不同。
3. **fgets()** 一行一行读取文件，读取windwos和linux下的文件没有问题，但是读取mac的文件却不能正常读入。

原因是fgets()是以\n来判断一行的，而mac文本文件以\r来结束一行。将mac文件转为dos文件即可正常读取。也可以用getdelim()或getline()函数来解决这个问题，因为在这两个函数中用个参数表示终结符，也就是说自己说明什么字符是终结符。

如果要做到通用的话，还得判断文本文件是什么格式的文件，也好知道传入哪一个终结符的参数。

来源：<https://blog.csdn.net/miaozk2006/article/details/82416939>

区别：

MS-Windows 默认使用的文件格式是：**dos**

Unix,Linux 默认的文件格式是：**unix**

当然也有**mac**的文件格式。

unix和dos文本文件的格式区别主要是两点：

1. 换行符 end of line (<EOL>)

文件格式	风格	Hex
dos	<CR> <NL> (\r\n)	0x0D0A
unix	<NL> (\n)	0x0A
mac	<CR> (\r)	0x0D

其中<NL> 等价于<LF>。

Windows下换行使用CRNL(\r\n)两个字符来表示，其中CR为回车（ASCII=0x0D），NL为换行（ASCII=0x0A），而在Linux下使用NL一个字符来表示。因此在Linux下使用vi来查看一些Windows下创建的文本文件时，有时会发现在每一行尾部有^M字符，其实它是显示CR回车字符。

2. 文件结束符

文件格式 Hex

文件格式 Hex

dos	0x04
unix	0x1A

查看文件格式：

1. 在vim中使用命令 `:set ff` 或 `:set fileformat` 即可返回当前文件的文件格式(fileformat)。

2. Linux命令：

```
file [filename]
```

如果是dos文件，则会返回：

```
UTF-8 Unicode text, with CRLF line terminators
```

如果是unix文件，则会返回：

```
UTF-8 Unicode text
```

文件格式转换：

1. **dos2unix** 和 **unix2dos** 工具

Linux有两个工具，可以实现文件类型相互转换。

```
dos2unix dosfile.txt
```

2. 使用**Vim**

使用Vim打开文件

DOS转UNIX：`:set fileformat=unix`

UNIX转DOS：`:set fileformat=dos`

保存退出：`:wq`

命令中 **fileformat** 也可以用 **ff** 来代替，并可以输入`:help fileformat`来获得跟多的关于选项信息。

关于Vim设置：

在Vim中有两个参数可以设置**fileformat**, **fileformats** ,具体可以通过帮助手册来了解它们。

```
:help fileformat
```

文章来源:https://blog.csdn.net/qq_37151416/article/details/111039516

本文来自互联网用户投稿，该文观点仅代表作者本人，不代表本站立场。本站仅提供信息存储空间服务，不拥有所有权，不承担相关法律责任。如若转载，请注明出处：

<http://www.ngui.cc/el/2851869.html> 如若内容造成侵权/违法违规/事实不符，请联系编程学习网邮箱：809451989@qq.com进行投诉反馈，一经查实，立即删除！

相关文章

暂无图片

[\[c语言\]_const修饰指针](#)

const修饰指针变量时候有三种情况： 1 const 修饰取值操作(修饰星号*) int a 10; const int* p1 &a; const int *p2 &a; int const* p3 &a; int const *p4 &a;以上两种指针p1,p2,p3,p4的声明等价，当const在星号 * 前面，...

[阅读更多...](#)

暂无图片

[C语言实现泛型 C11关键字_Generic](#)

在 C11 标准中，_Generic 关键字让 C 语言如同 C 等面向对象程序设计语言一样，支持轻量级的泛型编程设计。 1 泛型是什么 泛型编程（generic programming）是程序设计语言的一种风格或范式。泛型允许程序员在强类型程序设计语言中编...

[阅读更多...](#)

暂无图片

[迅雷下载的资源，文件夹有大小，但是打开没有文件](#)

问题描述： 刚刚使用迅雷下载资源，下载成功后，在迅雷中右键打开文件夹没有响应，但是在迅雷里播放该资源文件建立的视频是可以播放的。于是我切换到迅雷的下载文件夹中，发现有两个同名的文件夹，鼠标放上去都...

[阅读更多...](#)

暂无图片

[ssh连接VitrualBox虚拟机](#)

在Virtual Box里安装了Ubuntu 20.04虚拟机，打算在Windows上使用ssh连接，按照网上的教程，发现怎么都无法连接上虚拟机，报错： ssh_exchange_identification: read:

Connection aborted 。经过排查发现是Ubun...

[阅读更多...](#)

暂无图片

[\[C语言\]二维数组与指针](#)

二维数组与指针 1 二维数组初始化 以下几种赋值方式均可。 `int a[4][4] {0};int a[4][4] {1,2,3,4,5,6};int a[4][4] {{1,2,3},{4,5},6} //其他元素都为0;int b[][4] {{1,2},{3,4}}; //第一维可以省略;第二维不能省略2 遍历 //用一个for遍历 int c[3][4]; fo...`

[阅读更多...](#)

暂无图片

[\[C语言\]函数指针](#)

本篇是《C指针：基本概念、核心技术及最佳实践》（Pointers In C : A hands on Approach) 第七章“函数指针”的读书笔记。 1 定义方式 函数指针可以实现动态调用函数，适用与在程序运行时才决定调用哪个函数的情形，可以把函数指针...

[阅读更多...](#)

暂无图片

[Monkey Patch failed: Error: Command failed: 解决方案](#)

Vs code 的界面字体看久了，有点腻，想要换成Idea界面那种字体，于是我找到了插件“Custom UI”，“Custom UI”插件依赖于”Monkey Patch“ 插件。 在启用”Monkey Patch“插件的时候报错： Monkey Patch failed:...

[阅读更多...](#)

暂无图片

[修改VS code 界面字体](#)

修改VS code 的界面字体，不是修改编辑框里面的字体。 需要安装两个插件：“Custom UI”和”Monkey Patch“。 在插件商店就可以找到。 我们只需要安装“Custom UI”就可以了，安装了“Custom UI”就会自动安装”Monkey Patch“。 ...

[阅读更多...](#)

暂无图片

[\[c++\] std::pair的用法](#)

pair是c中的一种模板类型，存储两个数值类型；既可以是已经存在的，也可以是自己定义的数值类型。 pair将2个数据组合成一组数据；当需要这样的需求时就可以使用pair；如stl中的map就是将key和value放在一起保存。另一个应...

[阅读更多...](#)

暂无图片

[\[c++\] insert和emplace的区别](#)

emplace()和emplace_back()是c11新引入的方法。 emplace()也是在iter处插入元素；其他元素后移。 insert和emplace的区别在于；insert是插入一个完整的对象；而emplace是先调用该对象的构造函数生成的对象；再把这个对象插入vecto...

[阅读更多...](#)

最新文章

- [CARLA平台+Q-learning的尝试 \(gym-carla\)](#)
- [在gym的MountainCar环境下，用图片帧作为状态训练DQN网络](#)
- [Docker的常见用法与Carla的使用](#)
- [Carla语义分割图数据的进阶处理](#)
- [基于meanshift的目标跟踪详细解读代码（卡尔曼滤波及CAMshift）](#)
- [如何注册ActiveX控件，并在jsp页面调用](#)

Copyright @ 2022~2023