



北京邮电大学

Beijing University of Posts and Telecommunications



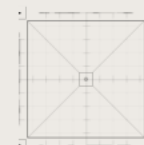
显示乱码问题

蒋砚军 北京邮电大学计算机学院

► 显示乱码问题



- 屏幕显示乱码问题
- 文本文件格式问题
- 汉字编码问题
- 终端配置：配色和中文乱码



► 屏幕显示乱码问题



■ 现象

`cat /bin/bash`或`head -n 1 /bin/bash`

导致屏幕乱码，键盘输入也只能看到画表用的框框符

■ 原因

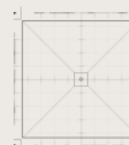
`/bin/bash`文件不是文本文件，这些二进制内容逐字节送到终端，凑巧一个序列被终端理解为一个转义序列，执行动作：修改了终端字符集

■ 字符集

- ◆ ASCII字符集（7比特128种符号）
- ◆ 选择其他字符集，相同码值代表了那个字符集之中的某个符号

■ 解决方法

- ◆ 断开连接，重新登录
- ◆ 注意不要把非文本信息在终端输出



► 文本文件格式问题



■ 现象

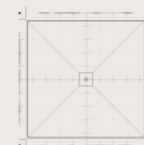
Linux下的文本文件在Windows用notepad打开，所有内容粘在一行内

■ 原因

Linux和Windows的文本文件的存储格式不同

■ 解决方法

- ◆ 在Linux完成格式转换后再用Windows处理（首选此方案）
- ◆ 用Windows中可容忍Linux文本文件格式的其他编辑器



► 文本文件格式



■ Windows和Linux文本文件的行尾不同

- Linux: 行尾处仅存换行字符
- Windows: 行尾处存回车和换行两个字符

例:文件mini.txt, 第一行为ab, 第二行为xyz

- Windows中, 文件的大小为9字节

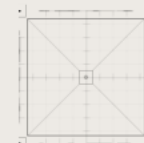
61 62 0d 0a 78 79 7a 0d 0a

- 在Linux中, 文件的大小为7字节

61 62 0a 78 79 7a 0a

■ 相关命令

- ◆ dos2unix/unix2dos, todos/frodos (两种格式间转换)
- ◆ file (查看文件的数据类型, 供参考)



► 中文编码问题



■ 问题

从Windows迁来的文件，只有在显示中文字符时是乱码。Linux本生的中文信息正常显示。

■ 原因

中文GBK与UTF8不兼容

■ 英文字符集

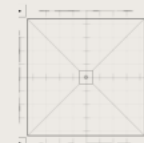
7比特ASCII码，字节高位为0的，后面7位是英文ASCII码

■ 中文字符编码GBK

Windows默认中文编码方案，两个字节表示一个汉字，字节的高位为1，以区别于ASCII码

■ 中文字符编码UTF8

许多Linux默认中文编码方案，三个字节表示一个汉字，字节的高位为1，以区别于ASCII码



► 中文编码问题



■ 检查系统设置

与语言有关的环境变量LANG应为en_US.UTF-8

```
env | grep LANG
```

若有问题，设置方法为：

```
export LANG=en_US.UTF-8
```

■ 命令iconv：中文字符编码的转换

```
iconv -f gbk -t utf8 (from GBK to UTF8)
```

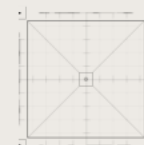
```
iconv -f utf8 -t gbk (from UTF8 to GBK)
```

```
$ echo "汉字" | od -t x1
```

```
0000000 e6 b1 89 e5 ad 97 0a
```

```
$ echo "汉字" | iconv -f utf-8 -t gbk | od -t x1
```

```
0000000 ba ba d7 d6 0a
```



► 终端配置：配色和中文编码



■ 问题

- ◆ 终端不是默认的黑底白字，ls, grep, vi等输出信息无配色
- ◆ 中文显示的是乱码

■ 调整终端仿真软件的配置

根据所用的终端仿真软件的不同，调整**终端类型**和**字符编码**，具体方法与所用软件有关。

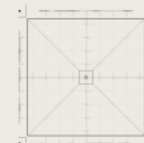
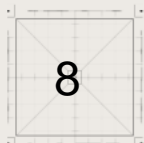
■ 以Windows下SecureCRT为例

会话选项==>终端==>仿真（**终端类型**选“Linux”，勾选“ANSI颜色”）

会话选项==>终端==>仿真（**字符编码**选“UTF-8”）

会话选项==>终端==>仿真==>字体（选“黑体”，**字符集**选“中文”）

最后一项如果未选择中文字体；或者选了中文字体，字符集未选成中文。汉字显示时仍然是乱码。





北京邮电大学

Beijing University of Posts and Telecommunications



谢谢