

什么是codepage和iocharset

在配置内核的VFAT文件系统配置项下，有2个选项需要输入，缺省的值是：

<*> VFAT (Windows-95) fs support

(437) Default codepage for FAT

(iso8859-1) Default iocharset for FAT

那么，codepage和iocharset是什么呢？它们其实都是为了同一个目的：如何解释文件名的字符编码。

先看看内核源码中的文档 Documentation/filesystems/vfat.txt 中的说明：

```
codepage=### -- Sets the codepage number for converting to
shortname
```

```
characters on FAT filesystem.
```

```
By default, FAT_DEFAULT_CODEPAGE setting is used.
```

```
iocharset=name -- Character set to use for converting between the
encoding is used for user visible filename and 16 bit
Unicode characters. Long filenames are stored on disk
in Unicode format, but Unix for the most part doesn't
know how to deal with Unicode.
```

```
By default, FAT_DEFAULT_IOCHARSET setting is used.
```

codepage 也就是以前PC机上的OEM代码页，用来表示字符码128～255的字符。字符码0～127之间是标准的ASCII码，大家都一样；而128～255则是 某些语言的字符或特殊的图形符号，由IBM规定了一些代码页定义。codepage用于短文件名，即8.3格式的DOS文件名。

Windows文件系统的长文件名是用16位的Unicode存储的，而Linux文件系统不支持16位的字符，所以需要在两者之间转换，这就是iocharset要完成的功能。

三、如何选择codepage

codepage 可以选择内核缺省的437，即美国英语代码页。如果要与Windows系统共用文件系统，则需要选择磁盘分区所使用的代码页。在Windows系统的命令窗口下，执行chcp命令，可以查询磁盘的代码页。chcp命令也可以修改Windows系统磁盘分区的代码页。对于简体中文Windows系统，选择 936，简体中文（GBK）。

四、如何选择iocharset

在内核中，iocharset缺省是iso8859-1，不使用中文时，可以选择这个缺省值。如果需要使用中文，则可以选择utf-8。utf-8是UCS represented in 8-bit text，是一种转换码，可以表示Unicode的编码和32位的UCS编码。

关于Unicode和UCS的说明，可以看<http://www.nada.kth.se/il8n/ucs/unicode-iso10646-overview.html>

五、Native Language Support

在内核配置的VFAT选项下只是设置了codepage和iocharset的缺省值，还需要在File Systems --- Native Language Support 配置项选择需要编译进内核的语言支持，否则无法使用。前面提到的无法安装U盘的问题，就是因为设置VFAT的codepage为437，但却没有把437 代码页编译进内核造成的。

一般来说，我们可以选择下面的几项：

```
| --- Base native language support
| (iso8859-1) Default NLS Option
| <*>    Codepage 437 (United States, Canada)
| <*>    Simplified Chinese charset (CP936, GB2312)
| <*>    ASCII (United States)
| <*>    NLS ISO 8859-1 (Latin 1; Western European Languages)
| <*>    NLS UTF-8
```

六、mount命令的选项

内核设置的是VFAT文件系统的codepage和iocharset的缺省值，在mount命令没有指定codepage和iocharset选项时使用。在执行mount命令时，可以在命令行中用 -o 参数指定当次安装VFAT文件系统所用的codepage和iocharset，如：

```
mount -o codepage=850,icharset=iso8859-1,utf8 /dev/sdb1 /mnt
```

注意(from Documentation/filesystems/vfat.txt):

There is also an option of doing UTF-8 translations with the utf8 option.

NOTE: "iocharset=utf8" is not recommended. If unsure, you should consider the following option instead.

utf8=<bool> -- UTF-8 is the filesystem safe version of Unicode that is used by the console. It can be enabled for the filesystem with this option. If 'uni_xlate' gets set, UTF-8 gets disabled.

<bool>: 0, 1, yes, no, true, false

附录: Accessing the short file names

(from: <http://ryanlee.wikidot.com/mount>)

The Linux vfat module hides the short file names. There is no way round this - it is only possible to see either the short names or the long names!

To see the short names (in cases where there is a long name too) the partition can be mounted using the msdos module. You may need to install the kernel module (kernel-module-msdos). Then mount command still requires a codepage value, but iocharset and utf8 are not used:

```
mount -t msdos -o codepage=850 /dev/sdb1 /mnt
```

Select the code page as before. The file system will show only the short names, and only valid short names may be used for new files. The behaviour when using an invalid name may be unexpected - the name simply gets truncated.