

I18N 简介

习惯上，[internationalization](#)（国际化）简写为I18N,中间的数字为省略的字母个数。

在 Linux 系统中，通过定义一组环境变量来设置程序的语言环境，以实现对I18N的支持，这种机制称为：[locale](#)（本地化）

locale 变量

LANG	低优先级全局 locale 变量	如果下面的变量未赋值，默认使用此变量的值
LC_COLLATE	比较和排序习惯	会影响目录列表的分类显示等
LC_CTYPE	定义系统的字符处理特性	哪些字符能被视为字母、数字，等等；与中文输入关系密切
LC_MESSAGES	提示信息,错误信息, 状态信息, 标题, 标签, 按钮和菜单等	
LC_MONETARY	定义货币单位和货币型数值的格式	
LC_NUMERIC	定义非货币型数值的格式	影响到千位分隔符和小数分隔符等
LC_TIME	定义日期和时间的格式	
LC_NAME	姓名书写方式	
LC_ADDRESS	地址书写方式	
LC_TELEPHONE	电话号码书写方式	
LC_MEASUREMENT	度量衡表达方式	
LC_PAPER	定义默认的纸张尺寸	
LC_IDENTIFICATION	对 locale 自身包含信息的概述	
LC_ALL	高优先级全局 locale 变量	为此变量赋值会强行覆盖上面变量的值，不推荐

- 假如未设置以上变量，系统将采用 POSIX 作为 lcoale，也就是 C locale

locale 值

[locale](#) 变量的值有三个要素：语言代码 (Language Code)、地域代码 (Country Code) 和字符集 (Encoding)^[35]：

语言代码[_]地域代码[.字符集]

例如：

语言/国家代码	描述
en_US.ISO-8859-1	美国英语
en_US.UTF-8	
zh_CN.UTF-8	简体中文
zh_TW.UTF-8	繁体中文

字符集

众所周知，计算机中的信息，是以数字形式表示的，字符也不例外。字符以数字编号的形式存储，使用时，根据这个编号，在字符集中找到相应的字符

字符集是字符在系统内的编码方式，也就是通常所说的内码^[36]。不同的字符集有不同的编码方式，

例如“码”字，它的 GB2312编码 为 426B；UTF-8编码 为 E7A081 。如果错误的以 GB2312编码 来检索 E7A081，将会产生类似“锘跨爨”的乱码

只要系统中安装了中文字体，通过字符集的支持，便可以正常显示中文，而无需设置 locale；locale 可以使操作界面显示中文，并可以使用中文输入等

其它

另外还有一个本地化变量叫做 LINGUAS。它会影响到基于 gettext 的程序^[37]；它还能决定某些特殊软件包的本地化，比如kde-i18n和openoffice。这个变量的值为一组以空格分隔的语言代码：

```
LINGUAS="zh en"
```

^[35] 中文 Windows 系统中使用的字符集为 GB2312，而 UTF-8 是大势所趋，它能够显示比 GB2312 更多的字符

^[36] 参见/usr/share/i18n/charmaps文件

^[37] 通过编译时安装本地化文件的方式实现i18n，编译时需要Native Language Support(NLS)支持