第32卷 第19期 计 算 机 工 程 2006年10月
Vol.32 № 19 Computer Engineering October 2006
·软件技术与数据库· 文章编号：1000—3428(2006)19—0046—03 文献标识码：A 中图分类号：TP316

基于红旗 Linux 的多语种操作系统的设计

李 莉[1,3]，缪 成[2]，吾守尔·斯拉木[1]
(1. 新疆大学信息科学与工程学院，乌鲁木齐 830046；2. 同济大学经济管理学院，上海 200092；
3. 同济大学电子与信息工程学院，上海 200092)

摘 要：介绍了新疆地区广泛使用维吾尔、哈萨克、柯尔克孜等少数民族语言与汉语在计算机处理方面的差异。以 Linux 系统的国际化框架为基础，设计出了可以支持维、哈、柯、汉、英的多种语言文字输入显示的多语种 Linux 操作系统。并介绍了民文输入法、复杂文本层等主要模块的实现技术。

关键词：Linux；本地化；多语言；复杂文本层；输入法

Design of Multilanguage OS Based on Red Flag Linux

LI Li[1,3], MIAO Cheng[2], WSHUR · Silamu[1]
(1. College of Information Science and Engineering, Xinjiang University, Urumqi 830046; 2. School of Economic Management, Tongji University, Shanghai 200092; 3. School of Electronics and Information Science, Tongji University, Shanghai 200092)

【Abstract】This thesis outlines the differences between the Chinese and minority language of Uighur, Kazak, Khalkhas, which are generally used in Xinjiang Province. On the basis of internationalization framework of Linux, the paper designs multilanguage Linux system which supports input and display of Uighur, Kazak, Khalkhas, Chinese and English. It also introduces implementation of master module, such as input, complex text layout.

【Key words】 Linux; Localization; Multilanguage; Complex text layout; Input method

Linux 以价格低廉、安全性好、配置灵活等优点在世界上得到广泛使用，我国政府和信息产业主管部门也在全面推广 Linux 操作系统。在新疆维吾尔自治区，维吾尔、哈萨克、柯尔克孜等少数民族文字（以下简称为民文）使用较为广泛，并且它们也经常与汉字一起混合使用。但是目前国内外开发的 Linux 发行套件只能处理汉文与英文，这严重地阻碍了 Linux 在新疆地区的使用和普及。针对这一情况，我们以国内普遍使用的中文红旗 Linux 操作系统为原型，开发出具有全民文界面，可以同时输入显示民文、汉文、英文等文字，符合民文编辑习惯的多文种 Linux 操作系统。

1 民文的语言特点和多文种软件基本概念

维吾尔文、哈萨克文和柯尔克孜文这 3 种文字均属阿拉伯语系的文字，却有着比较复杂的使用特性，主要包括：

(1)与上下文内容相关的显现形式。民文字母会根据邻近字母的属性具有不同的显现形式，一般来说分为 4 种显现形式：独立型，首写型，中写型和尾写型。在这里引出了两个概念：民文名义字符及其变形显现形式，名义字符是指标识民文字母的编码；变形显现形式指在语义上还是名义字符，但它却有着和名义字符不同的书写形式。

(2)双向性。民文文字书写习惯是靠右对齐，从右向左书写，与汉文、英文等相反，而且在新疆地区使用的操作系统需要能够同时处理显示民文与汉文的混合文本。这里定义两个概念：逻辑顺序和可示顺序。字符的逻辑顺序与字符输入顺序一致。对于汉英文字符逻辑顺序与可示顺序是一致的。但是对于民文文本，二者并不一致。

(3)变形组合字型。这是民文的一种特殊的书写形式，它以一个字型来显示民文的两个字母，如维文的 ﺑ 字母与 ﺠ 字母连写时显示的字型是 ﺒﺠ，以一个字型代替两个字母。

2 Linux 国际化框架和系统的总体设计方案

为使 Linux 操作系统可以支持民文、汉文与英文的处理，必须依靠操作系统的国际化与本地化机制。Linux 国际化的核心是 NLS(National Language Support)子系统，它的总体框架如图 1 所示。该子系统构建在基于 ASCII 码的 Linux 核心上，为世界上不同地域、不同语言环境的应用提供国际化本地化支持。

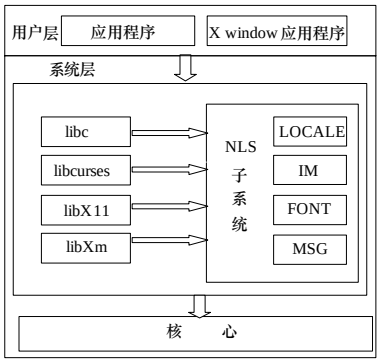

图 1 NLS 子系统结构

操作系统中所有支持国际化结构的实用程序，包括 X Window 应用程序都是建立在这个基础上。包括存储各种不同语言文化特征和特定字符编码处理功能的本地化数据库(LOCALE)；文字输入(IM)体系架构；支持多种文字字符显示

基金项目：国家“863”计划基金资助项目（2003AA1Z2110）；新疆维吾尔自治区高技术基金资助项目（200412108）；新疆大学青年教师基金资助项目（QN040123）
作者简介：李 莉(1977－)，女，博士生、讲师，主研方向：中文信息处理；缪 成，博士生、助研；吾守尔·斯拉木，教授
收稿日期：2005-11-10 **E-mail：**maomi_xj@sina.com

的字体(FONT)选择匹配机制；不需要重新编译，就可以实现显示当地语言的本地化文本信息(MESSAGE)等。

设计维、哈、柯、汉、英多语种操作系统的首要问题是选择系统的内码，候选的编码标准有 UTF－8 和 GB18030，它们都较好地支持 ASCII 码，同时也涵盖了新疆地区所使用的民文、汉文字符。GB18030 在 4 字节区支持了民文编码并且目前 Linux 已经较好地支持了 GB18030 编码。但是由于 GB18030 不是系统默认处理编码，在处理字符，如字符输入，字符编码在客户程序之间通信，就需要经常将字符在 UTF-8 编码与 GB18030 编码之间转换，增加了系统开销，而对于中文系统，因为要兼容以前大量使用 GB 编制的文档，所以需要使用 GB18030 编码，但对于民文不存在这种问题。所以系统最终采用了 UTF-8 作为系统内码。

至此，我们设计出 Linux 多语种操作系统的本地化环境变量名，如维吾尔语的环境变量为 ug_CN.UTF-8。在此基础上，以 Linux 的 NLS 作为系统设计的总体框架，通过制作民文本地化环境，集成民文输入模块、安装民文字体和民文界面翻译文件(MESSAGE)，在图形函数库中加入民文的复杂文本变形显示支持，建立了一个完善的民、汉文多语种操作系统。系统总体结构如图 2 所示。

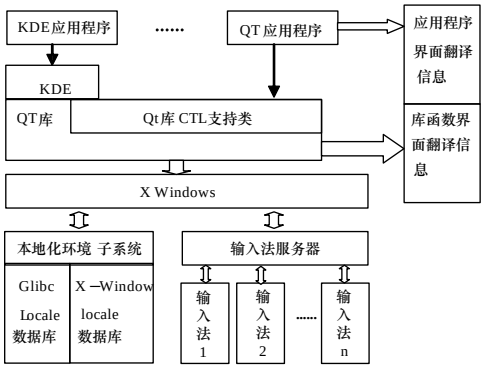

图 2 系统总体设计结构

(1)本地化环境子系统:包括 glibc 的 locale 和 X Window 的 locale 环境，它们影响基本函数行为，提供了程序运行的本地化环境，为整个系统提供了与民文的文化特性有关的描述信息。

(2)自适应多语言输入法子系统：为系统提供统一的汉文、民文的输入服务，该子系统同时提供了民文键盘映射与输入法服务器模式的民文输入，通过用户的设置可以方便地进行纯民文输入和民文与汉文的混合输入。

(3)民文界面信息子系统：为系统提供民文界面信息，这一部分包括应用程序的民文界面的翻译和主菜单、桌面条目的民文翻译和民文字体安装配置等。

(4)民文复杂文本图形库支持子系统：在图形库的层次上提供民文输出时的变形显示、从左向右书写等复杂文本层（Complex Text Layout, CTL）功能的支持。

对于民文版本的红旗 Linux 操作系统，民文的本地化环境数据库格式、民文界面资源的翻译合并等都与中文版本差别不大，所以在本文不作介绍，主要介绍民文复杂文本图形库支持子系统和自适应多语言输入法子系统的设计与实现。

3 主要子系统的设计与实现

3.1 民文复杂文本图形库支持子系统

开发民文操作系统最困难的是民文靠右对齐，从右向左书写，自动选型等特点的实现。对于这一类复杂文本特性，X Window 体系并没有给出支持，而是一般交由上层的图形函数库处理，在 Gnome 中由 pango 负责处理，而在 KDE 中由 Qt 库进行复杂文本的处理。因为 RedFlag 使用的是 KDE 桌面环境，这里以 Qt 库为例说明民文这部分特性的实现。

3.1.1 民文自动选型功能的实现

自动选型是按照民文字母连接属性及相邻字母连接属性的不同来确定其显示字型。在 Qt 库中由 QTextEngine 类负责对显示字符串进行分析，按字符的属性不同将显示字符串分为 item(字符块)。再以字符块为单位取得字符显示属性，计算绘制坐标，进行显示等工作。它的成员函数 shape()会根据 item 中字符对应的文本种类调用 scriptEngines 数组(也被称为 JumpTable)中与该文本对应变形处理函数。我们就是在 scriptEngines 插入了一个民文自动选型的变形函数，进行显示前的自动选型操作。它的处理过程是:

(1)对有左右映射的字符，计算、显示其映射字符。

(2)按照（表 1）民文变形规则表，计算民文的显示字型。

(3)根据名义字符编码及其上一步所计算出的显示字型，查表得到字符的显现字型编码。

(4)根据前后字型是否可以变形组合，将可以组合的字符替换为合体字符编码。

表 1 民文变形规则表

规则	内容	字符属性
规则 1	名义字符连接属性是透明属性，不影响连接行为	显现字符使用独写型
规则 2	名义字符连接属性是右连属性，右侧字符是可以与左边字符相连时	显现字符使用尾写型
规则 3	名义字符连接属性是双连属性，右侧字符是可以与左相连，左侧字符是可以与右相连时	显现字符使用中写型
规则 4	名义字符连接属性是双连属性，右侧字符是可以与左相连，但左侧字符不能与右相连时	显现字符使用尾写型
规则 5	名义字符连接属性是双连属性，右侧字符不可以与左相连，但左侧字符可以与右相连时	显现字符使用首写型
规则 6	其他情况	显现字符使用独写型
规则说明：(1)应用规则的优先级从高到低，只有前面的规则都不适用时，才应用下一个规则。(2)每一个字符都有连接属性。包括右连、双连、者透明属性		

3.1.2 民文从右向左书写与靠右对齐功能的实现

在 Qt 图形库显示字符时，先应用双向算法对民文字符块倒序，使民文字符串的逻辑顺序变为显示顺序，我们用大写字母来代表民文，那么处理前的民文顺序如果是“ABCDE”，处理后的民文顺序为“EDCBA”。同时将其应用的字符块顺序上，就能支持民文嵌套汉文或汉文嵌套民文的处理。

其次因为文本块靠右对齐与靠左对齐是对称的，只是在 X 坐标上相差一个宽度。在此通过动态坐标映射算法来计算每一个文本块靠右对齐的绘制坐标，以图 3 为例，Bounding Rect 代表字符绘制区，Text2 代表一行中的第 2 个文本块，moveBy 表示 Text2 在靠右对齐下状态下 X 坐标需要移动的距离。Right()、width()等分别代表相应区域在靠左齐下方式下的坐标值

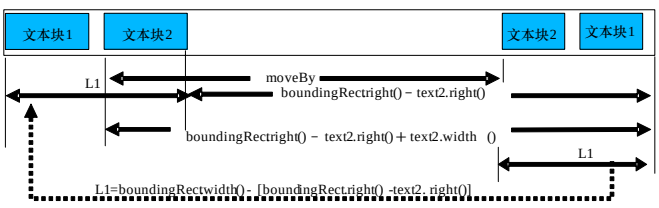

图 3 坐标值变换示意图

L1=boundingRect.width()-[boundingRect.right()-text2.right()]，它代表了文本块最远一侧与对齐点之间距离。在靠左对齐与靠右对齐这两种方式下，L1 的值是相等的。

boundingRect.right()–text2.right() + text2.width() 在靠左对齐方式下代表了 Text2 左侧边界到字符绘制区右侧（也就是靠右对齐方式下对齐点）之间的距离。

在从靠左对齐向靠右对齐方式转变时，Text2 的 X 坐标需要移动的距离：

moveBy=boundingRect.right()–text2.right()＋text2.width()-L1

=2*(boundingRect.right()–text2.right())+text2.width()-boundingRect.width

通过以上方式将以靠左对齐计算出的坐标值变换为靠右对齐的绘制坐标，从而可以在统一的框架下实现靠左对齐与靠右文本块的绘制坐标计算。

3.2 自适应多语言输入法子系统

民文像英文一样是字母文字。因此在做民文输入法时可以使用键盘映射方式。Linux 在 XKB 扩展状态下最多可以加载 4 组键盘映射表。在本系统中将英文、维吾尔文、哈萨克文与克尔柯孜文分别设计成 4 组键盘映射表，用一个修饰键在组之间切换。以维文、英文的键盘映射表文件简述如下：

```
xkb_keymap "uighurKey" {
        …//第 1 部分设计键盘类型，在此省略。
          xkb_symbols            { //配置键盘映射编码表
      name[Group1]= "US/ASCII";   //第 1 组命名
      name[Group2]= "uighur";   //第 2 组命名
key     <AC08> {   [ k,   K ], //第 1 组键盘映射表中“k”键大小
//写状态的编码定义。
          [   Arabic_kaf,   0x010006C6   ]      }; //第 2 组键“k”键
//大小写状态的编码定义。
```

这里的按键名“AC08”、字符编码名“Arabic_kaf”都是 XKB 扩展协议中的定义，如果字符编码在 XKB 扩展协议中没有定义，则通过 0x0100XXXX 的方式可以定义输入编码为 XXXX 的字符。

键盘映射民文输入法具有效率高、稳定性高的优点，但是这种输入法在民文、汉文同时输入时，输入法之间切换很不方便，因为从民文输入向汉文输入切换时，需要用右 alt 键先切换到英文，再由输入法服务器的输入法切换方法(通常是 ctl+space 键)由英文切换到汉文，反之亦然。为了提高系统的适用性，我们又开发基于红旗 Linux 输入法服务器 rfinput 的民文输入法，开发了民文输入法动态链接库模块，将民文与汉文输入程序都作为输入法服务器中的一种输入法模块，统一运行在输入法服务器 rfinput 下。

但是由于 rfinput 是以 GB18030 作为程序内码，它的码表文件中汉字存储的全部是 GB18030 的编码，如果要将其全部移植成 UTF-8 编码，工作量较大，还需要修改输入法模块的代码。因此采用了多重本地化运行环境的技术，单独编制的民文输入法模块采用与 rfinput 一致的内码 GB18030，并使 rfinput 及其加载的输入法模块独立运行在 zh_CN.GB18030 的环境下。以下是 rfinput 运行配置文件，

```
[Desktop Entry]
Name=Autorun
Name[zh_CN]=rfinput
Comment[zh_CN]=Autorun - rfinput
Exec=LC_ALL=zh_CN.gb18030 rfdock//将 rfdock 和 rfinput 的运
行环境设置为 gb18030，rfdock 是 KDE 托盘程序，在其中自动启动
rfinput 运行。
…
```

通过以上方式，当用户只需要输入纯民文字符时可以选择民文键盘映射输入法。而在同时输入民、汉文时可以选择 rfinput 提供的民汉文输入法。从而可以广泛地适应于不同用户的需要，方便进行民文字符的输入。

4 系统的测试和结论

通过以上方式在红旗 Linux 的基础上研发的维、哈、柯、汉、英多语种 Linux 操作系统，在保证红旗 Linux 操作系统所有功能的基础上，支持民文语言习惯与用户界面，提供了具有广泛适用性的民文输入法，在使用 Qt 和 KDE 编辑的图形程序中实现了民文、中文混合输入显示、民文自动选型、从右向左书写等功能。在图形方式下，可以像中文一样地使用民文建立文件名、文件夹、搜索维文文件等功能。实现了维、哈、柯、汉、英多语种红旗 Linux 操作系统的目标。

参考文献

1 戴庆厦，许寿椿，高喜奎. 中国各民族文字与电脑信息处理[M]. 北京：中央民族出版社, 1991: 83-94.

2 “九五”国家重点科技攻关项目. 开放系统中文处理应用平台(98-779-01-02)[R]. 中国科学院软件研究所开放系统与中文信息处理中心, 2000-09.

3 GB18030-2000. 信息交换用汉字编码字符集基本集扩充[M]. 北京：中国标准出版社, 2000.

4 The Linux Globalization Specification[EB/OL]. http://www.li18nux.org/li18nux2k/.

5 XKB Library and Protocol Reference Manuals[EB/OL]. ftp://ftp.X.org/pub/R6.4/XC/doc/hardCopy/XKB/.

~~~~~~

（上接第 45 页）

主要功能，而且为测试者提供了一套构造具体测试环境所需要的核心组件以及许多可供扩展的接口。通过已经实现的分布式仿真测试环境可以看出，基于这个框架，测试者可以很快地开发出适合特定测试项目的可靠性测试环境，这可以极大地提高测试环境的开发效率，从而达到进一步提高可靠性测试自动化程度的目标。

参考文献

1 Lyu, Michael. Handbook of Software Reliability Engineering[M]. Mc Graw Hill and IEEE Computer Society Press, 1996.

2 阮　镰，刘　斌，陈雪松. 软件可靠性测试及其环境[J]. 测控技术, 2000, 19(2).

3 刘　斌，高小鹏，陆民燕等. 嵌入式软件可靠性仿真测试系统研究[J]. 北京航空航天大学学报, 2000, 26(4).

4 Broekman B, Notenboom E. 张君施，张思宇，周承平译. 嵌入式软件测试[M]. 北京：电子工业出版社, 2004.

5 Jorgesen P C. 韩　柯，杜旭涛译. 软件测试(第 2 版)[M]. 北京：机械工业出版社, 2003.

6 Fayad M E, Johnson R E. 姜晓红，李　岩，赵爱东译. 特定领域应用框架：行业的框架体验[M]. 北京：电子工业出版社，2004.
~~~~~~