

SUN 工作站矢量汉字处理技术

吴林 郭翔 (国家遥感中心航空遥感一部 100076)

摘要:本文分析了 SUN 工作站上的汉字处理原理及矢量字库结构,介绍了在 SUN 工作站上扩充矢量汉字处理的方法。并具有平台无关性,可用于其他基于 X Window 的操作系统。

关键字:矢量汉字 IM 服务器 IM 库 XIM 协议

SUN 工作站以其卓越的性能,强大的处理能力赢得广大用户的青睐。但 SUN 工作站的汉字处理能力,一直困扰着国内用户。虽然 SUN 工作站 Solaris 操作系统具有汉字处理能力,但非常有限。(1)仅提供两种中文字体,且字形有限(3种)。(2)受 X11 协议的限制,只能处理点阵字符,变化能力弱。Solaris 的汉字处理能力难以满足国内用户的需要,是 SUN 工作站在我国应用的一个瓶颈。笔者通过对 SUN 工作站上的汉字处理和 X11R5 中有关国际化的支持作了研究分析,解决了 SUN 工作站的矢量汉字处理,并能对汉字进行无极缩放和旋转,具有平台无关性。增强了工作站上的汉字处理能力。

一、SUN 工作站汉字处理机制

SUN 工作站使用的是基于 X 窗口系统的 OpenWindows 图形用户界面(GUI)。在其上开发汉字功能有三种方法:(1)利用 OLIT(OPEN LOOK Intrinsic Toolkit)编程环境;(2)利用 XView 编程环境;(3)直接在 X11R5 环境下开发。由于 X11R5 下开发的 GUI 具有与平台无关性,所以下面着重介绍 X11R5 下开发图形用户界面。

1. X 窗口系统输入方法

X 窗口系统输入方法(X Window System Input Method)简称 XIM,具有国际化处理机制(双字节和多字节处理),不仅能处理 ACSII 字符,且可支持各国语言的输入及处理。XIM 与 X 窗口系统相互关系如图 1 所示,包括 X 服务器、IM 库(Input Method Library)、IM 服务器(Input Method Server)。X 服务器将多个击键形成的组合字符串送到 IM 服务器中,IM 服务器再依据组合字符串映象一个双字节或多字节字符(如中文,韩文)。

2. IM 服务器

X 窗口系统中的 IM 服务器是一个独立的进程,命名为 htt。Solaris 支持下列亚洲国家和地区:中国大陆、中国台湾、日本和韩国。IM 服务器主要由下列进程组成:

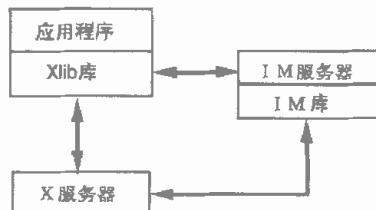


图 1

(1)事件处理,处理所有的直接来自 X 服务器的键盘事件。

(2)语言接口,通过这个接口,IM 服务器能在两种语言之间变换(英语和汉语或其他亚洲语言)。

(3)描绘子进程,支持查找选择区,状态区风格和各种预测区风格。

- 查找选择区,用于输入方法查找选择,如汉语拼音输入等。

- 状态区风格,用于状态区管理和显示,同时支持根窗口显示状态区。

- 预测区风格,它是设定字符预显示区位置和字符在插入之前的显示形式。

3. IM 库

IM 库主要包括下面三部分功能:

(1)代码转换器,将 EUC(Extends UNIX Code)转换成复合字符串或复合字符串转换成 EUC 的函数集合。

(2) 字符输出,在 X 窗口中输出双字节或多字节字符的函数集合。

(3) 字符输入,是输入与 IM 服务器通信的函数集合,通过 XIM 协议与 IM 服务器通信。

4. XIM 协议

XIM 协议(XIM Protocol)负责 IM 库和 IM 服务器之间的通信,主要用于键盘事件处理。XIM 协议将接口子进程信息送到 IM 服务器,IM 服务器响应接受到的信息。类型主要有两种: XIMP-FE-TYPE1 和 XIMP-SYNC-BE-TYPE2。

(1) XIMP-FE-TYPE1,是一种前端 IM 服务器。仅当输入变换是打开时,多个击键组合才能成一个组合字符串,并把它输入到 IM 服务器中处理(如图 2)。

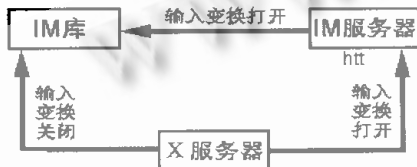


图 2

(2) XIMP-SYNC-BE-TYPE2,是一种后端同步 IM 服务器。无论输入变换是打开还是关闭,所有键盘事件都先送到 IM 服务器,IM 服务器负责送回无用的键盘事件到 IM 库(如图 3)。

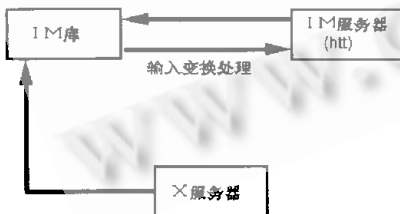


图 3

二、矢量字库

矢量字库由两部分组成:一是每个汉字的索引信息,包括汉字形在字库中的偏移位置及长度;二是汉字字形数据,包括汉字每个笔划点数及坐标。

矢量字库中每个汉字的索引是定长的(6字节),其中前4字节为偏移量,后两个字节为长度。矢量字库中符号(前16区)与汉字(后72区)是分开存放的,且区位码机内是高位置1,因此索引信息在字库中首址可由下式求出:

$$\text{汉字索引信息首址} = ((\text{区码} - 0xb0) * 94 + \text{位码} - 0xa1) * 6;$$

$$\text{符号索引信息首址} = ((\text{区码} - 0xa1) * 94 + \text{位码} - 0xa1) * 6;$$

根据字形的索引信息,读取压缩的字形数据,还原时根据字节第6、7位组合的4种状态(00、01、10、11)分别进行解码。

三、SUN 工作站上的矢量汉字处理方法

SUN 工作站上增加矢量汉字处理方法,需要进行以下步骤:

1. 初始化地区语言环境

程序使用 setlocale 函数初始化地方语言环境。setlocale 检查环境变量 LANG,LANG 定义为 zh,ko 或 zh-TW 表明设置了地区语言环境,其中 zh 表示中国大陆版,ko 表示韩国,zh-TW 表示中国台湾。其语法关系如下:

```
(void) setlocale(LC-ALL, NULL); /* 依据 LANG
设置地区语言环境 */
```

或

```
(void) setlocale(LC-ALL, "zh"); /* 直接设置汉字
环境 */
```

2. 应用程序与 XIM 服务器连接

在设置了地区语言环境后,应用程序需调用 Xlib 中 XOpenIM 函数打开输入方法,然后使用 XCreatIC 函数创建输入上下文 IC。输入上下文 IC 是指抽象的表示一个特殊输入进程的状态。连接上 XIM 后,用 XmbLookupString 函数从输入方法中获取复合双字节字符串。这一过程如下所示:

```
KeySym    keysym;
XIMStyle  xim-mode = XIMPreeditPosition | XIMSta-
tusArea; /* 初始化预测区位置和状态区风格 */
XIM       im;
XIC       ic;
...
```

```

...
if ((im = XOpenIM(display, rdb, NULL, NULL)) ==
NULL) { /* 显示服务器上并打开 IM 服务器 */
    printf("Error: XOpenIM()! \n");
    exit(0); }
...
...
ic = XCreateIC(im,
    XNInputStyle, xim-mode, /* 设置预测区位置和
状态区风格 */
    XNClientWindow, window, /* 设置客户窗口 */
    XNFocusWindow, client, /* 设置聚焦窗口 */
    XNPreeditAttributes, preedit-attr, /* 设置预测区
风格 */
    NULL);
...
...

```

3. 键盘处理获取国标码

X 中处理键盘是以事件方式处理, X 服务器接受到一个事件, 判断此事件是否为键盘事件, 如果是则调用键盘处理进程。

```

for(;;) {
    XNextEvent(theDisplay, &event); /* 处理事件 */
    switch(event.type) {
        case FocusIn: /* 客户聚焦 */
            ...
        case KeyPress: /* 键盘处理 */
            count = XmbLookupString(ic, /* 输入上下文 */
                (XKeyPressedEvent *)&event, /* 键盘事件 */
                string, /* 输入返回的字符串 */
                len, /* 字符串长度 */
                &keysym, /* 键盘系统 */
                &lus-status);

```

string 返回输入的字符串即国标码。

4. 双字节字符输出

使用 XmbLookupString 函数返回的 string (在此为国

标码), 将国标码转换为区位码之后, 按照计算索引地址公式, 从矢量字库中取出字形数据的偏移地址以及汉字长度, 再根据偏移地址及长度从字形数据部分读出汉字笔划各端点坐标值, 然后应用 Xlib 函数、XFillPolygon 函数描绘出汉字。在实际应用中以 ReadHZ 函数、CreatePoint 函数、Dcode 函数对矢量汉字库进行解码与输出, 字库采用 UC DOS 矢量汉字库。输出子程序如下:

```

/* 获取矢量点坐标 thePoints 和点数 theNpoints */
...
/* 利用 XFillPolygon 函数输出矢量汉字 */
XFillPolygon(theDisplay, theWindow, theGC, /* 指
向显示器、窗口和图形上下文 GC */
    thePoints, /* 矢量坐标点(x,y)阵列 */
    theNpoints, /* 矢量点数 */
    theshape, /* 说明形状参数帮助服务器提高性能, 有三种选择 Complex、Convex 和 Nonconvex */
    theMode); /* 说明坐标模式, CoordModeOrigin 为绝对坐标,

```

...

使用上述方法, 扩充了矢量汉字处理功能, 一改 SUN 工作只处理点阵字符风格, 增强了汉字字体、字体变化能力, 改善工作站的中文环境, 方便中国用户使用。此方法还适用于 SGI、HP、DEC 等使用 X 窗口的工作站环境, 具有很强的适用性(有需要源程序的读者, 可与作者联系)。

参考文献

- [1] Asian Solaris 2. 4 Application Developer's Reference Manual. SUN Microsystem Inc. 1995.
- [2] 张倪, 莫斌等. Motif 与图形用户开发工具. 北京, 清华大学出版社, 1994.
- [3] 刘鹏, 叶欣等. Xlib 编程手册. 北京, 海洋出版社. 1992

(来稿时间: 1997 年 8 月)