Simplified Chinese Solaris User's Guide

Adobe PostScript™

040805@9495

Contents

Preface

Simplified Chinese Solaris User's Guide provides information for users, system administrators, and developers on the Solaris™ Operating System (Solaris OS). The guide also describes Simplified Chinese localized desktop tools and utilities available with the Simplified Chinese Common Desktop Environment (CDE).

About This Book

This guide offers information on using, customizing, and administering the Simplified Chinese Solaris Operating System and CDE.

How This Book Is Organized

Each chapter of this guide covers a different aspect of the Simplified Chinese Solaris software.

Part I introduces the features of the Simplified Chinese Solaris Operating System.

- Chapter 1 describes the new features and the language support that are available in this release.

Part II includes end-user information.

- Chapter 2 discusses the design of the Simplified Chinese Solaris software and provides information on the supported locales.
- Chapter 3 describes the features of the Simplified Chinese Common Desktop Environment (CDE) for windowed applications.

- Chapter 4 describes the different Simplified Chinese character input modes and provides procedures that you can use to enter and convert Simplified Chinese characters.

- Chapter 5 describes how to use the `mailx`, `talk`, `xtobif`, `stdconvtool`, and `iconv` utilities. The chapter also provides tables of code conversion modules.

- Chapter 6 describes line printer, PostScript™ printer, and laser printer support in this release.

  Part III includes information for system administrators.

- Chapter 7 provides procedures on setting a default locale, adding a new locale, and changing font directories.

- Chapter 8 discusses input method server basics and application display.

- Chapter 9 provides information on PostScript, bitmap, and TrueType fonts.

- Chapter 10 discusses printing to line printers and to PostScript printers.

- Chapter 11 provides information on TTY STREAMS, command, setup, and terminal support.

- Chapter 12 discusses Simplified Chinese text and conversion utilities. The chapter also describes the binary Compatibility Package and provides a procedure for running binary code from previous Solaris releases.

- Glossary defines key words and phrases found in the Simplified Chinese Solaris documentation.

Related Books

The following books include further details on the topics covered in this guide.

- *Solaris 10 Release Notes* offers the most up-to-date information about the features offered in this release.

- *Solaris 10 Installation Guide* describes the process of installing the Solaris Operating System.

- *Solaris Advanced User's Guide* provides information on the Solaris command line interface (CLI).

- *International Language Environments Guide* describes the internationalization and the localization features that are available for application developers in this release.

Accessing Sun Documentation Online

The docs.sun.com^{SM} Web site enables you to access Sun technical documentation online. You can browse the docs.sun.com archive or search for a specific book title or subject. The URL is `http://docs.sun.com`.

Ordering Sun Documentation

Sun Microsystems offers select product documentation in print. For a list of documents and how to order them, see "Buy printed documentation" at `http://docs.sun.com`.

Typographic Conventions

The following table describes the typographic changes used in this book.

TABLE P–1 Typographic Conventions

Typeface or Symbol	Meaning	Example
AaBbCc123	The names of commands, files, and directories; on-screen computer output	Edit your `.login` file. Use `ls -a` to list all files. `machine_name% you have mail.`
AaBbCc123	What you type, contrasted with on-screen computer output	`machine_name%` **su** `Password:`
AaBbCc123	Command-line placeholder: replace with a real name or value	To delete a file, type **rm** *filename*.
AaBbCc123	Book titles, new words, or terms, or words to be emphasized.	Read Chapter 6 in *User's Guide*. These are called *class* options. You must be *root* to do this.

Shell Prompts in Command Examples

The following table shows the default system prompt and superuser prompt for the C shell, Bourne shell, and Korn shell.

TABLE P–2 Shell Prompts

Shell	Prompt
C shell prompt	`machine_name%`
C shell superuser prompt	`machine_name#`
Bourne shell and Korn shell prompt	`$`
Bourne shell and Korn shell superuser prompt	`#`

Features Overview

This part introduces the features of the Simplified Chinese Solaris Operating System (Solaris OS).

Overview of Features

The Simplified Chinese Solaris Operating System (Solaris OS) is the internationalization and the localization of the current Solaris Operating System and the Common Desktop Environment (CDE) window system.

This chapter describes the new features and the language support that are available in the Simplified Chinese Solaris release.

New Localized Features

New to this release is the Wubi input method, support for the 3.2 version of the Unicode Standard, and improvements to the `mp` print filter. This filter replaces the `xetops` and the `xutops` utilities.

- Wubi Input Method. One of the main advantages of Wubi and other shape-based input methods is a very low repetition rate. A single Wubi code seldom represents more than one character, meaning that you can enter text more quickly.

  Under the authorization of Wangma Company, the following Wubi features are available in the Solaris 10 release:

  - GB18030-2000 character set support – The GB18030 Chinese character set is the national character encoding standard issued by Chinese government in 2000. The Wubi input method supports the GB18030-2000 character set. Wubi makes working with the smaller character sets contained in GB18030-2000 easier.

  - Easy character set switching – Solaris Wangma Wubi divides GB18030 into three character sets: GB2312, GBK and GB18030. You can use keyboard shortcuts to switch between character sets as you type.

  - New radical mechanism for Simplified and Traditional Chinese. – Patented by professor Wang Yongmin, who invented the Wubi input method, this new mechanism was developed from the old radical system, version 86. With no additional training, users of Wubi version 86 can access three times more

characters of the same encoding and the same typing rules.

- Unicode 3.2 support. The `zh_CN.UTF-8` (`zh.UTF-8`) locale has been updated to support the new 3.2 version of the Unicode Standard. The new version introduces an additional 1,016 new characters and contains various normative and informative changes.

  Unicode 3.2 also defines the following newly invalid UTF-8 byte sequences:

  - 0xED as the first byte.
  - 0xA0 to 0xBF as the second byte.

    These sequences exclude the surrogate code points between U+D800 and U+DFFF. To comply with the new definition, the Simplified UTF-8 `iconv` modules have been enhanced to detect the newly defined UTF-8 invalid byte sequences.

- In the current Solaris release, the `mp` printing utility replaces the `xetops` and the `xutops` utilities.

Note – The `xetops` and `xutops` printing utilities are no longer supported in the Solaris Operating System. The utilities were formerly used to convert Simplified Chinese text files to PostScript. The conversion enabled the printing of Simplified Chinese characters to PostScript printers with no resident Asian fonts. The `xetops` utility was used in the `zh_CN.EUC/zh` locale and in the `zh_CN.GBK/zh.GBK` locale. The `xutops` utility was used in the `zh_CN.UTF-8/zh.UTF-8` locale.

The `mp` printing utility was first released with the Solaris 9 Operating System.

Language Support

The current Solaris release builds inherent internationalization features into every localized product. Localization facilities support the ANSI C recommendations for internationalization and localization that define the locale and related categories.

Locale Attributes

A *locale* contains the culturally specific information and conventions of the language for a particular global region. Each process in the Solaris Operating System has the following set of locale attributes:

- Locale settings, which provide the `locale` and `setlocale` commands you use to list and set attributes before you start a process from the command line.

For example, the Simplified Chinese locales and the English/ASCII locale both have a category that defines the display of time and date according to the cultural format, as well as the actual Simplified Chinese or English/ASCII characters for the time and date.

- Code sets, which support coding conventions for the GB2312 and the GB18030 character sets. These sets enable you to input, display, and print Simplified Chinese text in file names, system messages, and terminal (TTY), email, and data file content.

- `htt` input method server, which handles Simplified Chinese input for the Solaris Operating System. The `htt` server receives your keyboard input and converts it to Simplified Chinese characters that are used in Simplified Chinese applications.

Simplified Chinese Locales

The Simplified Chinese Solaris Operating System provides simultaneous support for the locales in the following table. The locales look the same to the end user, but the internal character encoding is different.

TABLE 1–1 Simplified Chinese Locales

Locale	Description
`zh_CN.EUC (zh)`	Simplified Chinese EUC (GB2312)
`zh_CN.GBK (zh.GBK)`	Simplified Chinese GBK
`zh_CN.GB18030`	Simplified Chinese GB18030-2000
`zh_CN.UTF-8 (zh.UTF-8)`	Simplified Chinese UTF-8 (Unicode 3.2)

Simplified Chinese Code Sets

The following table lists supported code sets for each Simplified Chinese locale.

TABLE 1–2 Simplified Chinese Code Sets

Locale	Code Set
`zh_CN.EUC (zh)`	gb2312
`zh_CN.GBK (zh.GBK)`	GBK
`zh_CN.GB18030`	GB18030–2000
`zh_CN.UTF-8 (zh-UTF-8)`	UTF-8

Simplified Chinese Input Methods and Fonts

The Simplified Chinese Solaris Operating System provides input methods and fonts for the locales shown in the lists and tables in this section.

The following input methods are supported for the zh locale:

- New QuanPin
- New ShuangPin
- NeiMa
- QuanPin
- ShuangPin
- GB2312 NeiMa
- WangMa Wubi
- English-Chinese
- Optional code table input methods
- Input method auxiliary window support for Simplified Chinese

The following input methods are supported for the zh_CN.GB18030 locale:

- New QuanPin
- New ShuangPin
- QuanPin
- ShuangPin
- GB18030–2000 NeiMa
- WangMa Wubi
- English-Chinese
- Optional code table input methods
- Input method auxiliary window support for Simplified Chinese

For a complete list of fonts supported for the Simplified Chinese locales, see "Bitmap and TrueType Fonts" on page 120.

Input Method Auxiliary Window

The input method auxiliary window supports the following functions.

- Input method switching

- Input methods properties configuration

- Lookup tables for the following character sets:

  - GB2312
  - GBK
  - GB18030-2000
  - Unicode

- User-defined characters

- Input method help

- Virtual keyboard

Locale Categories

In the Simplified Chinese Solaris Operating System, you can use the following general and specific categories as defined by ANSI C for the Simplified Chinese and English locales:

- General `LC_ALL` setting that invokes all of the categories for locale-related aspects of the environment.
- Specific settings for particular aspects of the environment, which include the following categories:
  - `LC_CTYPE`
  - `LC_TIME`
  - `LC_NUMERIC`
  - `LC_MONETARY`
  - `LC_COLLATE`
  - `LC_MESSAGES`

For example, the Simplified Chinese and the English/ASCII locales have the `LC_TIME` category that defines the display of the time and date according to the cultural format, as well as the actual Simplified Chinese or English/ASCII characters used in the display.

End-User Information

This part includes end-user information.

Introduction to the Simplified Chinese Solaris Software

This chapter describes the design of the Simplified Chinese Solaris software. These sections also provide information on the locales supported in the Simplified Chinese Solaris environment. Topics in this chapter include:

Design of the Simplified Chinese Solaris Software

The Simplified Chinese localization of the internationalized release of CDE provides two work environments: a user environment and a developer environment. Each environment is localized to handle the linguistic and cultural conventions that are unique to the Simplified Chinese language.

- The user environment has desktop tools and a window manager (`dtwm`) that are customized to communicate in the language of a particular locale.

- The development environment provides internationalized versions of `Xlib` and Motif that programmers use to develop localized applications. For further information, see the *International Language Environments Guide*.

Setting a Simplified Chinese Locale

You can set any of the following locales when you login to your Simplified Chinese desktop:

- `C` – ASCII English environment.
- `zh` – Simplified Chinese environment in extended UNIX code (EUC).
- `zh.GBK` – Simplified Chinese environment in GBK, an extension of GB2312-80. This standard is Guo Biao Kuo in Chinese PinYin, which supports all CJK characters that are in Unicode 2.0.
- `zh_CN.GB18030` – Simplified Chinese environment in `GB18030-2000`. The new GB 18030-2000 standard, which will obsolete the GBK, supports CJK Unified Ideographs Extension A and Yi, Mongolian, Tibetan and Uigur minority scripts in Unicode 3.0.
- `zh_CN.EUC` – Symbolic link to `zh` locale.
- `zh_CN.GBK` – Symbolic link to `zh.GBK` locale.
- `zh.UTF-8` – Simplified Chinese environment in Unicode 3.2.
- `zh_CN.UTF-8` – Symbolic link to `zh.UTF-8`.

▼ How to Set the Locale at Login

You can set your default locale or change it with the following procedure.

欢迎来 nanjing
公用桌面环境(CDE)

请输入你的用户名

| 确定 | 重新开始 | 选项 ▽ | 提示 |

语言	▶	C 至 es_EC ▷	es_GT ---- 瓜地马拉语
会话过程	▷	es_GT 至 zh ▶	es_MX ---- 墨西哥语
远程登录	▷		es_NI ---- 尼加拉瓜语
命令行登录			es_PA ---- 巴拿马语
复位登录屏幕			es_PE ---- 秘鲁语
			es_PY ---- 巴拉圭语
			es_SV ---- 萨尔瓦多语
			es_UY ---- 乌拉圭语
			es_VE ---- 委内瑞拉语
			fr ------- 法语
			it ------ 意大利语
			sv ------- 瑞典语
			⊙ zh ------- 简体汉语

Common Desktop Environment
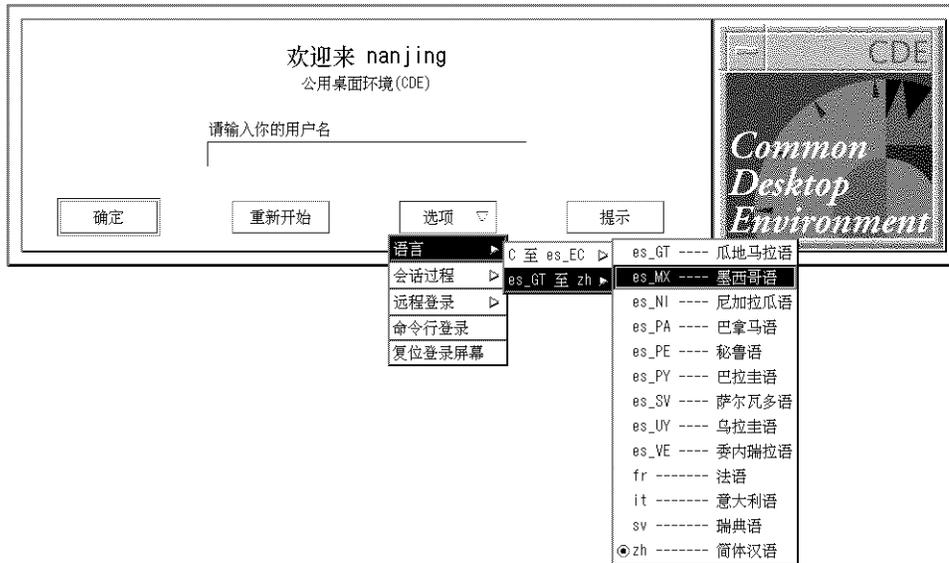
1. **Choose Language from the options menu on the login screen.**

2. **Select the C, zh, zh.UTF-8, zh_CN.GB18030, or zh.GBK locale.**
   The new locale is set for your CDE session.

CDE

The following sections provide information on the Simplified Chinese localization of the Simplified Chinese Common Desktop Environment (CDE) for windowed applications.

Simplified Chinese Graphical User Interface

The Simplified Chinese CDE Motif graphical user interface is similar in layout and design to the U.S. release of CDE. Simplified Chinese CDE supports multibyte characters and Simplified Chinese messages with Motif objects. Differences in character width and proportional spacing cause minor differences in the exact layout of some Motif objects.

All application windows that can take Simplified Chinese input include a *status area* associated with their input window to show the current conversion mode. With an input conversion mode on, as Simplified Chinese is being typed its entry point becomes a highlighted (reverse video) *preedit area* until the input is converted to Simplified Chinese or special characters and committed. Some input modes also provide conversion choices among several Simplified Chinese characters on menus.

Simplified Chinese Input/Output

To accommodate the diversity of Simplified Chinese, the Solaris software provides several different input methods for entering Simplified Chinese characters. With these methods you can enter ASCII/English characters, Simplified Chinese radicals, and Simplified Chinese characters using an ASCII keyboard or a Simplified Chinese keyboard.

Your Simplified Chinese input is stored temporarily as an intermediate representation. You then use the conversion manager program to transform the representation into a displayed character string.

The Localized CDE Desktop

The following desktop tools are available in this release. All of the tools can handle Simplified Chinese input and output.

Address Manager – Carries out remote operations and finds information about the systems and users on your network. This application can speed up such tasks as sending email, logging in remotely, and setting appointments on someone else's calendar.

Application Manager – Manages the tools and other software applications available on your system.

Audio Tool – Records, plays, and saves audio files in AU, AIFF, and WAV format.

Calculator – Mimics the function of a hand-held calculator.

Calendar Manager – Manages appointments and To Do lists. You can use this application to set and distribute appointment reminders.

Clock – Displays the current time in analog or digital format. You can control the display of local time with this application.

Console – Starts a `dtterm` terminal emulator as your workspace console window.

File Manager – Displays the files and folders on your system. You can move, copy, open, and delete files and folders with this application. You can also use the application to view the contents of your floppy diskettes and CD-ROMs.

Find Files – Enables you to search your system for files or folders that match specific search criteria. Your criteria might include, for example, the name, location, or the size of a file or folder you want to find.

Front Panel – Controls for access to applications and utilities and utilities on the system. The Front Panel is a centrally-located window that occupies all workspaces.

Help – Displays searchable help information for CDE.

Icon Editor – Enables you to create new icons or modify existing icons.

Image Viewer – Enables you to view, print, and save the contents of file types such as GIF, TIFF, JPEG, and PostScript. You can use the Snapshot function of this application to capture a picture of a window or another part of your screen. The picture is saved as a raster file in bitmap format.

Mailer – Handles the distribution, receipt of your electronic mail messages.

Performance Meter – Monitors various aspects of system performance.

Print Manager – Enables you to submit, view, and cancel print jobs. This application is the graphical front-end to the print command. It supports *drag-and-drop* file transfer operations.

Process Manager – Display the processes that are currently running on your workstation. The application enables to perform actions on the active processes.

Text Editor – Enables you to create and edit text files. The application is used in CDE tools such as the Mailer composition window.

Style Manager – Enables you to customize some of the visual elements and system device behaviors of your workspace environment, such as: colors and fonts, keyboard, mouse, window, and session start-up behaviors.

Terminal – Acts as an ASCII character terminal that you can use to enter UNIX® commands at a system shell prompt.

Entering Simplified Chinese Text

This chapter provides procedures and other information that you can use to enter Simplified Chinese text. The chapter discusses the following topics:

Character Input Overview

This chapter describes the Simplified Chinese Solaris input modes that you can use to enter the following kinds of characters.

- ASCII/English characters
- Simplified Chinese characters
- Special symbols

You can type all of these characters in the input areas of the following application windows:

- In terminal emulation (TTY) windows, such as Terminal windows

- In text entry areas, such as those found in the Text Editor and Mailer applications
- In dialog boxes, such as the new folder name box in the File Manager application
- In other special use subwindows, such as pop-ups

Input Window Areas

In the Simplified Chinese Operating System, application subwindows contain two areas that are used to enter Simplified Chinese characters.

In the Simplified Chinese Operating System, application subwindows contain two areas that are used to enter Simplified Chinese characters. A lookup choice window and an auxiliary window are also available in Chinese input mode.

- **Preedit area** – The text entry area that holds your character formations before you *commit* them.

Note – When you commit characters, the characters are put in the text block that is assembled for the application.
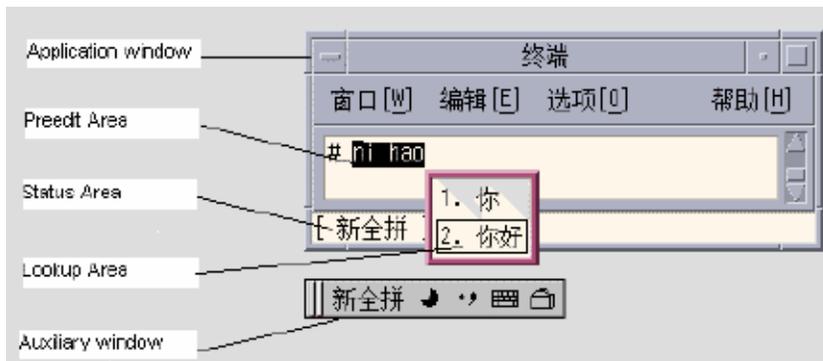
- **Status area** – The area at the lower left of the application subwindow that displays the current conversion mode and the active keyboard. Later sections in this chapter discuss keyboard switching and using the available conversion modes.
- **Lookup choice window** – A popup window that displays the conversion candidates that are available for the characters or the radicals in the preedit area.

- **Auxiliary window** – This window contains a *palette* of icons that provide you the following functions and utilities to simplify text entry and to manage input methods:

  - Input method switching
  - Chinese full-width/half-width character mode switching
  - Chinese/English punctuation mode switching
  - Input method properties setting
  - Input method selection
  - Lookup tables for GB2312, GBK, GB18030-2000, and Unicode character sets
  - Virtual keyboard

The input method auxiliary windows supports all UTF-8 locales and the following Simplified Chinese locales:

- `zh/zh_CN.EUC`
- `zh.GBK/zh_CN.GBK`
- `zh.UTF-8/zh_CN.UTF-8`

Two kinds of input methods are supported:

- Methods based on a code table such as Wubi
- Methods developed by a vendor, such as NewPinYin or NeiMa.

The following figure shows the interface model for auxiliary window support.
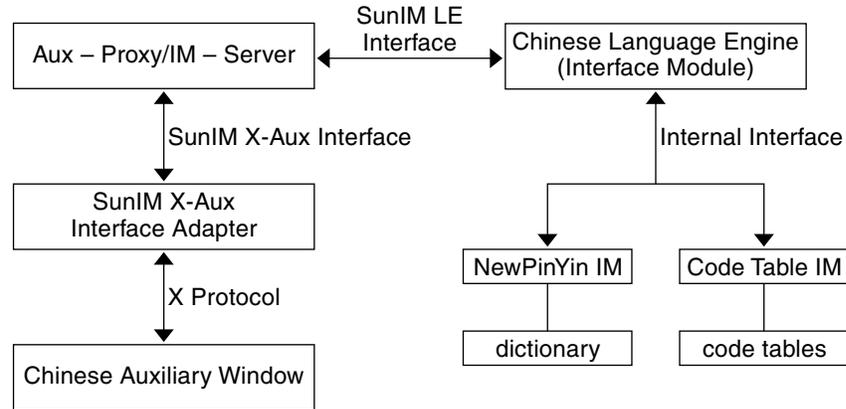
FIGURE 4–1 Interface Model for Auxiliary Window Support

Using Simplified Chinese Input Methods

This section provides procedures that you can use to select and switch between different input methods.

▼ How to Select an Input Method

1. **In the typing area, press Control-spacebar to turn on Simplified Chinese input conversion.**

   An auxiliary window appears.

2. **Select the desired input method through one of the following actions:**

   - In the status area of the application subwindow, use the function keys to switch input methods: the F2 key for the first input method, the F3 key for the second input method, and so on.

Tip – You can also press Control-Escape repeatedly until you reach the desired input method.

- Use the input method panel.
  - Click the utilities button in the auxiliary window.

    The utilities menu appears.
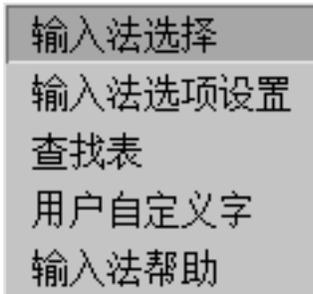
  - Click the input method selection item from the utilities menu.
    The input method selection panel appears.
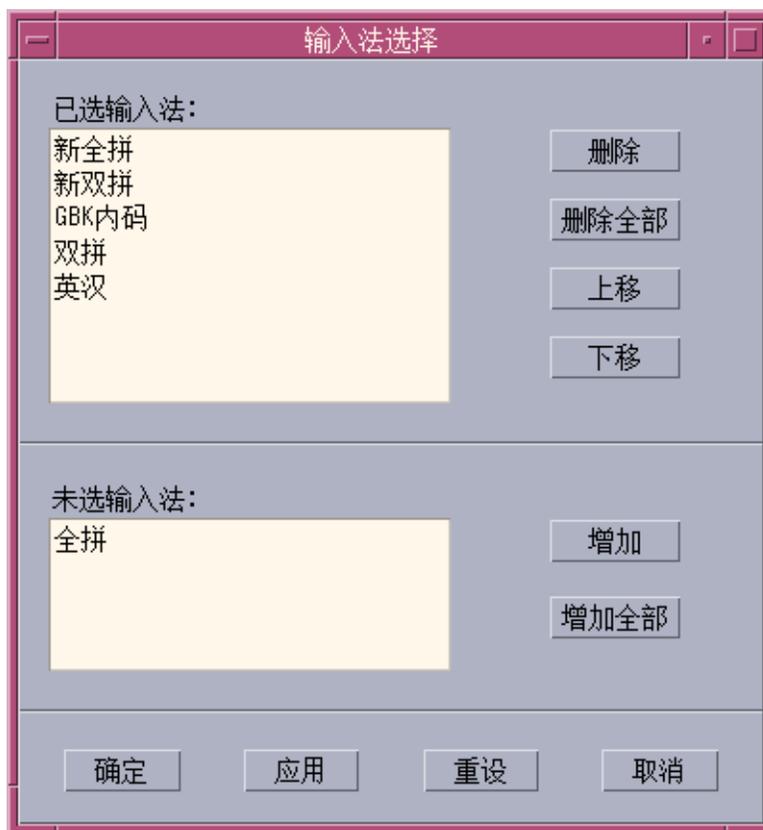
– Select the input method you want to use from the input selection panel.

After you select an input method, click OK or Apply to activate the setting. The first input method you select is the default input method.

Note – When you press Control-spacebar the default input method is selected.

If you change input methods, you can press Control-Escape to return to the default input method.

▼ How to Access Input Method Help

Help pages display in the default browser, such as Netscape Navigator™.

● **Select the input method help option from the utility menu to display the help pages in a browser.**

▼ How to Switch Between Half_width and Full_width Character Mode

● **Switch between half_width character mode and full_width character mode through one of the following actions.**

– In the status area of an application subwindow, type Shift-spacebar to switch between half_width character mode and full_width character mode.

– In the auxiliary window, click the half_width/full_width button.

The input method system is in full_width character mode when this button appears in the auxiliary window:

The input method system is in half_width character mode when this button appears in the auxiliary window:

When the system is in full_width mode, the full_width character of the input key is committed. For example, when you input an a in full_width mode, the full_width a is committed.

▼ How to Switch Between Chinese and English Punctuation Modes

● **Switch between Chinese punctuation mode and English punctuation mode through one of the following actions.**

– In the status area of an application subwindow, type Control- to switch between Chinese punctuation mode and English punctuation mode.

– In the auxiliary window, click the Chinese/English punctuation button.

The following icon indicates the input method system is in Chinese Punctuation Mode:

The following icon indicates the input method system is in English Punctuation Mode.

When you select the punctuation key in Chinese Punctuation mode, the corresponding Chinese punctuation character is commited to the application. For example, when you are in Chinese Punctuation mode and the $ symbol is selected,

the Chinese currency symbol character is committed to the application.

The punctuation keys include: , . / <> :;’”\$!^&_-

The correspondence between English keys and Chinese punctuation is mapped in the following figure.

```
!           !
"           "
$           ¥
&           —
'           '
(           (
)           )
」           ，
.           。
：           ：
；           ；
<           《
>           》
?           ?
@           ·
\           、
^           …
—           ——
```
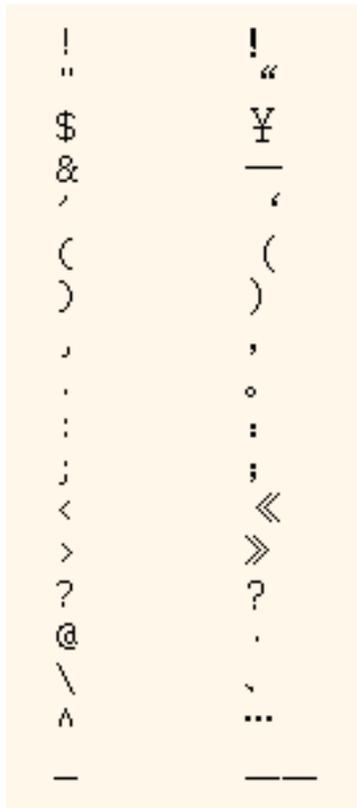
Setting Input Method Options

Four code table input options are available for the input method you select.

■ Display candidates key by key – This option causes the input method to search a dictionary table when you press a valid key. Candidates for selection then display in the lookup window.

If this option is not active, the character mapped to the key you press appears in the preedit area. When you press the spacebar, the input method engine searches the dictionary table and displays the available candidates for the character in the preedit area.

■ Display external codes – This option displays the external codes of the candidates you enter display in a lookup window.

- Automatically commit if only one candidate – This option commits the external code of a character when only one candidate is available. If this option is not selected, the external code of a character appears in a lookup window even when only one candidate is available.

- Display keymap character for every external code – This option displays the character mapped to a valid key in the preedit area when you press the key.

▼ How to Select an Input Method Option

1. **Click the input method selection item from the utilities menu.**
   The input method selection panel appears.

2. **Select an input method from the selection panel.**
   The input method options panel appears.

3. **Select an input method option.**

4. **Click OK or Apply to activate the selection.**

Input Method Utilities

After you make your input method selections, you can use the information and the procedures in this section to take the following actions:

- Activate lookup table selection
- Select a virtual keyboard
- Create user defined characters

Lookup table

From a lookup table, you can search for and select the Chinese characters you want to input. Three kinds of lookup tables available:

- Lookup tables with native encoding. A lookup table with EUC_CN encoding is provided in the zh_CN.EUC/zh_CN/zh locale. A lookup table with GBK encoding is provided in the zh_CN.GBK/zh.GBK locale, and a lookup table with GB18030 encoding is provided in the zh_CN.GB18030 locale.
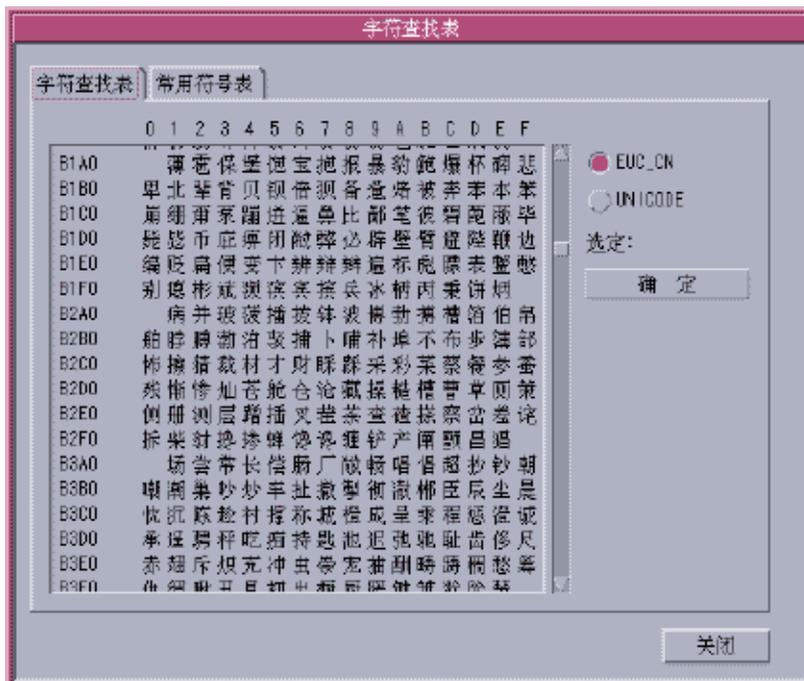
■ Lookup table with UNICODE encoding.

■ Lookup table for special characters, such as Greek characters and Mathematic symbols.
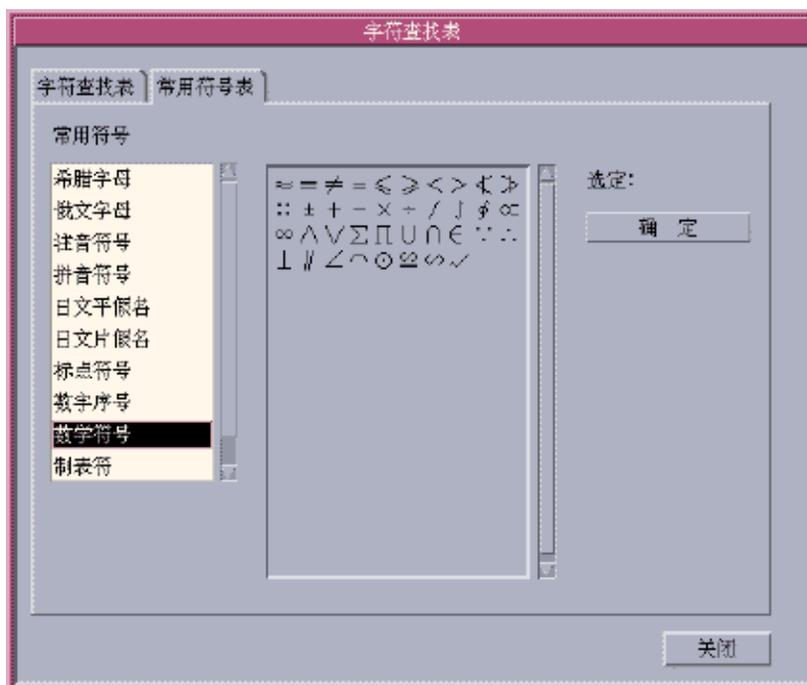
▼ How to Activate Lookup Table Selection

1. Click the utilities button in the auxiliary window.

2. Click the lookup item from the selection menu.

When you activate the lookup option, the characters that are available for a string you type in the preedit area display for selection in a lookup choice window.

▼ How to Search and Select Lookup Choices

1. Type a string you want to convert in the preedit area.

The lookup choice window appears.

You can use the following keys to search through the characters and radicals that are available for your string.

Period (.)   Moves forward to the next page of choices

Comma (,)   Moves backward through the choices

2. Type the number or letter of the label of the lookup choice you want to select.

Your choice is substituted for the string in the preedit area.

Virtual Keyboards

You can use virtual keyboards as lookup utilities to simplify the input of certain special symbols.

The Simplified Chinese system supports several virtual keyboards.

PC Keyboard

The following figure shows the PC virtual keyboard.

Greek Keyboard

The following figure shows the Greek virtual keyboard.

Russian Keyboard

The following figure shows the Russian virtual keyboard.

ZhuYin Keyboard

The following figure shows the ZhuYin virtual keyboard.

Chinese Punctuation Characters Keyboard

The following figure shows the Chinese Punctuation Characters virtual keyboard.

Number Symbol Lookup Keyboard

The following figure shows the Number Symbol Lookup virtual keyboard.

Mathematic Symbol Lookup Keyboard

The following figure shows the Mathematic Symbol Lookup virtual keyboard.

Special Symbols Lookup Keyboard

The following figure shows the Special Symbol Lookup virtual keyboard.

Table Symbol Lookup Keyboard

The following figure shows the Table Symbol Lookup virtual keyboard.

▼ How to Select a Virtual Keyboard

● **Click the virtual keyboard button in the auxiliary window.**

The virtual keyboard for the active input method appears.

User Defined Characters (UDC)

The user-defined character (UDC) editor tool enables you to draw and save new characters. Once you ascribe a character to an input method, the character can be displayed in an application.

▼ How to Activate the UDC Tool

- **Select the user defined character item on the utility menu to activate the UDC tool.**
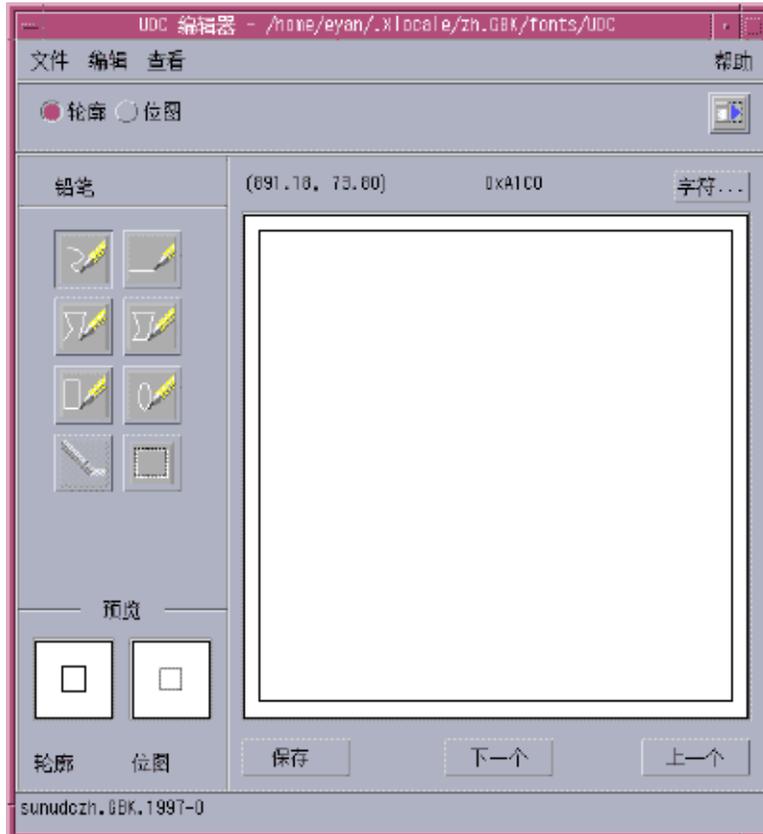
See Chapter 9 for more information about user defined characters.

Input Methods and Conversion Modes for Entering Text

This section describes the input methods and conversion modes that are available for entering ASCII/English, Simplified Chinese, and other characters.

In the `zh/zh_CN/zh_CN.EUC` locales, you can use the following function keys to access the available input methods:

- NewQuanPin, the default input method (F2)
- NewShuangPin (F3)
- GB2312 (F4)
- QuanPin (F5)
- ShuangPin (F6)
- English_Chinese (F7)
- WangMa Wubi (F8)

In the `zh.GBK/zh_CN.GBK` locales, you can use the following function keys to access the available input methods.

- NewQuanPin, the default input method (F2)
- NewShuangPin (F3)
- GBK NeiMa (F4)
- QuanPin (F5)
- ShuangPin (F6)
- English_Chinese (F7)
- WangMa Wubi (F8)

In the `zh_CN.GB18030/zh.UTF-8/zh_CN.UTF-8` locales, you can use the following function keys to access the available input methods.

- NewQuanPin, the default input method (F2)
- NewShuangPin (F3)
- GB18030 NeiMa (F4)
- QuanPin (F5)
- ShuangPin (F6)
- English_Chinese (F7)
- WangMa Wubi (F8)

ASCII Character Input

Applications start in ASCII mode and the status area of the application subwindow is blank. You can toggle ASCII mode on or off by pressing Control-spacebar or the Chinese/English key on a Chinese keyboard. When you turn off ASCII input mode, the indicator of the default input mode appears.
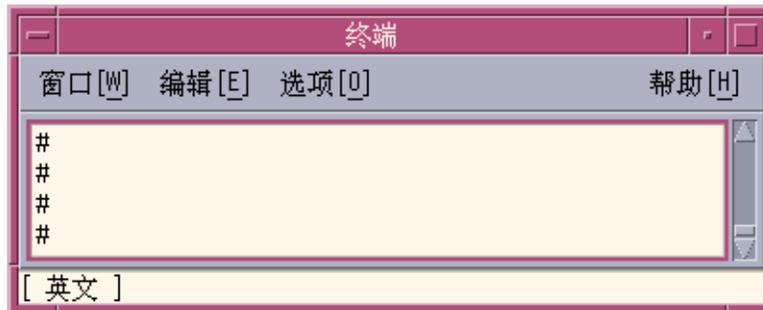
New QuanPin and New ShuangPin Input Methods

This section describes the features in the New QuanPin and New ShuangPin input methods, and how to use some of the features in the `zh_CN.EUC` and `zh_CN.GBK` locales.

PinYin is a popular input method in PRC. Various PinYin-based input methods exist. Two of these input methods, New QuanPin and New ShuangPin, provide support for the following features:

- Storing and recalling user-defined phrases
- Dynamically adjusting the frequency of lookup choices
- Typing PinYin strings up to 222 characters with the New QuanPin input method
- Typing ShengMu characters
- Entering GBK Hanzi phrases

These features are described in detail in the following sections.

▼ How to Define Phrases for Later Use

The following describes how to define the phrase **ke lin dun** and store it for later use.

1. **Select the input method.**

   Follow the steps in the procedure "How to Select an Input Method" on page 34 to select the input method.

2. **Type the phrase `kelindun` without spaces in the typing area.**

   The New QuanPin and New ShuangPin input methods insert spaces for you automatically.
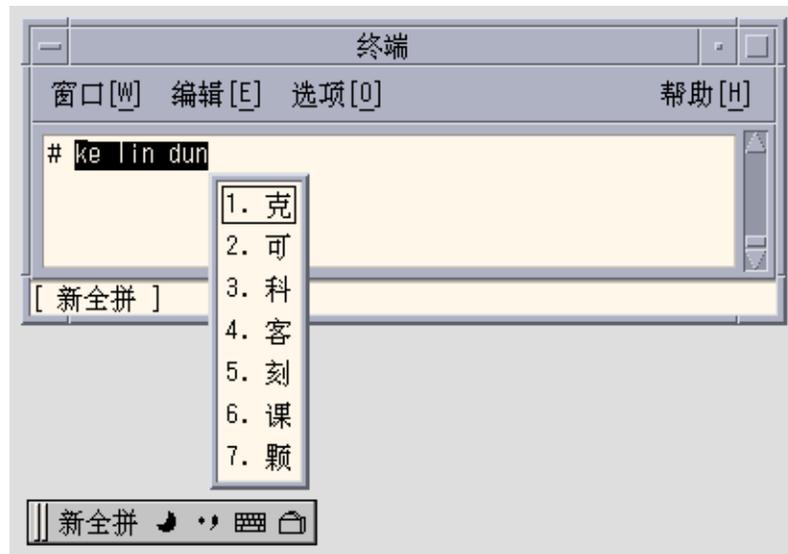
3. **Type the number that corresponds to the candidate you want to select.**

4. Select the characters of the second and third parts of the phrase.

The new phrase is defined and added to the user dictionary file. The next time you type **ke lin dun**, you will see the phrase you defined.

▼ How to Select Frequently Used Candidates

In the New QuanPin and the New ShuangPin input methods, the candidates that you select are moved to the start of the list to facilitate repeated use.

1. **Select the input method.**

   Follow the steps in the procedure "How to Select an Input Method" on page 34 to select the input method.

2. **Type sh yi.**

   Notice the order of the five available candidates.

3. **Select the fifth candidate.**

4. Type sh yi again.
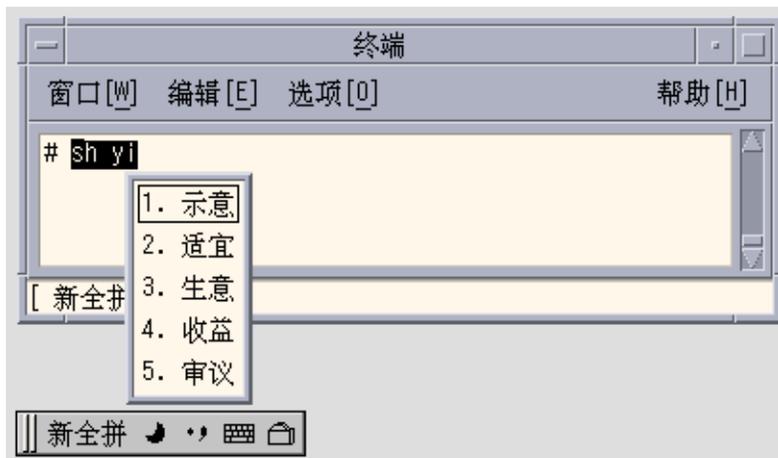
Notice that the fifth candidate has moved to the first position because you previously selected it. Frequently used candidates are promoted for faster selection.

Other Input Method Features

The NewQuanPin and New ShuangPin input methods provide support for other useful functionality.

Typing Long PinYin Strings

The New QuanPin input method accepts PinYin strings of up to 222 characters long. The following string is used in the next figure.

```
>>meiguozhongtongkelindunzhengzaitaolunhaiwanjushiwenti<<
```
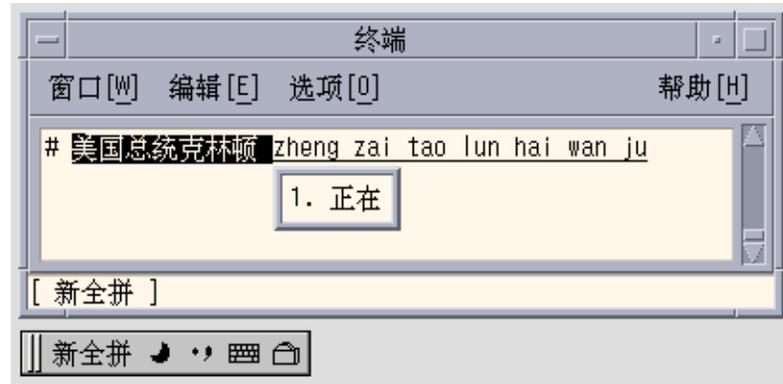

The result is the following Chinese string:

美国总统克林顿正在讨论海湾局势问题

Note – The New ShuangPin input method supports input strings of up to 30 characters.

Typing ShengMu

You can also type ShengMu only. Candidates are supplied for ShengMu, as shown in the following figure.

Entering GBK Hanzi

The zh_CN.GBK locale supports GBK by default, as shown in the following illustration:
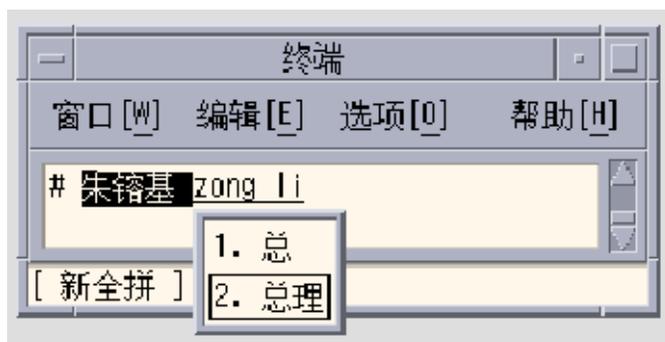
The second Chinese character in the following figure is defined only in the GBK standard.

朱镕基

Single GBK candidates are placed at the end of the list of candidates. Press Return to scroll to the GBK area.

Tip – For faster selection next time, you can define the GBK candidate as a phrase. For more information, see "How to Define Phrases for Later Use" on page 53.

Both New QuanPin and New ShuangPin support GBK Hanzi by default in the zh.GBK locale. However, because several Hanzi have the same ShengMu (the first part of PinYin), New QuanPin and New ShuangPin do not display GBK candidates if you provide only the ShengMu.

For example, typing the string rong will display GBK candidates because the string is a complete PinYin string. However, typing r alone will not display any GBK candidates because the string is only a ShengMu string.

Keyboard Definitions

This section describes the keyboard definitions that are used for the New QuanPin and New ShuangPin input methods.

Edit Keys

The following table shows the definitions of the edit keys.

Note – The preedit line is a normal X text field.

TABLE 4–1 Edit Key Definitions

Key	Definition
[a-z]	PinYin character.
Home	Moves to the start of the preedit line.
End	Moves to the end of the preedit line.

TABLE 4–1 Edit Key Definitions   *(Continued)*

Key	Definition
Left	Moves the caret in the preedit line to the left. If the left character is Hanzi, the original PinYin is displayed.
Right	Moves the caret in the preedit line to the right.
Delete	Deletes the PinYin character following the caret on the preedit line.
Backspace	Deletes the PinYin character preceding the caret on the preedit line.

Page Scroll Keys

The candidates of a PinYin string belong to the following groups:

- G1 – Highest frequency Hanzi + Long (3 or more) Cizu + Double Chinese Cizu
- G2 – GB Single Hanzi
- G3 – GBK Single Hanzi (in the `zh_CN.GBK` locale)

Some PinYin strings might have more candidates than can be displayed in the same window. In that case, use the keys described in the following table to scroll through the available candidates.

TABLE 4–2 Page Scroll Key Definitions

Key	Definition
- =	Scrolls to previous/next candidate
[ ]	Scrolls to previous/next candidate
, .	Scrolls to previous/next candidate
Return	Quickly scrolls through all candidates

Select Keys

New QuanPin and New ShuangPin use the numeric selection keys.

Separator Characters

In accordance with the national PinYin standard, the separator (') is supported to avoid ambiguous interpretations of PinYin strings. For example, the PinYin string `[jiang]` can be interpreted as `[jiang]` or `[ji][ang]`. Both spellings are valid. In New QuanPin, however, `[jiang]` is interpreted only as `[jiang]`. You must use the separator and enter `[ji'ang]` for the string to be interpreted as `[ji]` and `[ang]`. New ShuangPin does not require the use of separators.

Dictionary Files

New QuanPin and New ShuangPin share two dictionary files: `PyCiku.dat` and `Ud.Ciku.dat`. In the `zh_CN.EUC` and `zh_CN.GBK` locale, the default path names are `/usr/lib/im/locale/zh_CN/data/PyCiku.dat` and `/usr/lib/im/locale/zh_CN/data/UdCiku.dat`.

Users cannot normally write to these files. However, because users can affect the way New QuanPin and New ShuangPin work through features such as frequency adjustment and user-defined phrases, you should update the dictionary files frequently.

A user's dictionary is normally located in `~/.Xlocale/PyCiku.dat` or `~/.Xlocale/UdCiku.dat`. The tilde (~) indicates the home directory of the user who starts the `htt` command. When you start New QuanPin and New ShuangPin input methods, the system locates and reads the dictionary files in the user's home directory. If a dictionary file is not found, the following system default path is used:

`/usr/lib/im/locale/zh_CN/…`

New ShuangPin Features

ShuangPin is an abbreviated form of QuanPin. ShuangPinis faster but more difficult to use than QuanPin. New ShuangPin supports all of the features, keyboard definitions, and dictionary files of New QuanPin.

Various ShuangPin keyboard mapping designs exist in PRC. The most popular three designs are ZiRanMa, Chinese Star, and Intelligent_ABC. The New ShuangPin input method supports all three of these keyboard mappings.

New ShuangPin Keyboard Mapping

The following tables contain keyboard mappings for the ZiRanMa, Chinese Star, and Intelligent_ABC keyboards.

TABLE 4–3 ZiRanMa Keyboard Mapping

Key	Definition
i	ch
u	sh
v	zh
a	a

TABLE 4–3 ZiRanMa Keyboard Mapping   *(Continued)*

Key	Definition
b	ou
c	iao
d	uang, iang
e	e
f	en
g	eng
h	ang
i	i
j	an
k	ao
l	ai
m	ian
n	in
o	o, uo
p	un
q	iu
r	uan, er
s	iong, ong
t	ue
u	u
v	v, ui
w	ua, ia
x	ie
y	uai, ing
z	ei

TABLE 4–4 CStar2.97 Keyboard Mapping

Key	Definition
u	ch

TABLE 4–4 CStar2.97 Keyboard Mapping   *(Continued)*

Key	Definition
i	sh
v	zh
a	a
b	ia, ua
c	uan
d	ao
e	e
f	an
g	ang
h	iang, uang
i	i
j	ian
k	iao
l	in
m	ie
n	iu
o	o, uo
p	ou
q	er, ing
r	en
s	ai
t	eng
u	u
v	v, ui
w	ei
x	uai, ue
y	iong, ong
z	un

TABLE 4–5 Intelligent ABC Keyboard Mapping

Key	Definition
i	ch
u	sh
v	zh
a	a
b	ou
c	in, uai
d	ua, ia
e	e
f	en
g	eng
h	ang
i	i
j	an
k	ao
l	ai
m	ue, ui
n	un
o	o, uo
p	uan
q	ei
r	iu, er
s	ong, iong
t	uang, iang
u	u
v	v
w	ian
x	ie
y	ing

TABLE 4–5 Intelligent ABC Keyboard Mapping   *(Continued)*

Key	Definition
z	iao

GBK Code Input Method

The GBK code input method uses the GBK code defined by the Chinese Internal Code Specification. This method includes all of the Chinese characters and symbols in GB2312-80, and other CJK Chinese characters in GB 13000-1. Each Chinese character or symbol is identified by a four-hexadecimal digital internal code defined in the Chinese Internal Code Specification.

▼ How to Type GBK Code Text

This procedure describes how to use the GBK codes to type Chinese characters and symbols.

1. **Open a Terminal window.**

2. **In the Terminal window, press Control-spacebar to turn on Chinese input conversion.**

3. **Press F4 to select the GBK code input method.**

    The status area shows that the GBK code input mode is on.
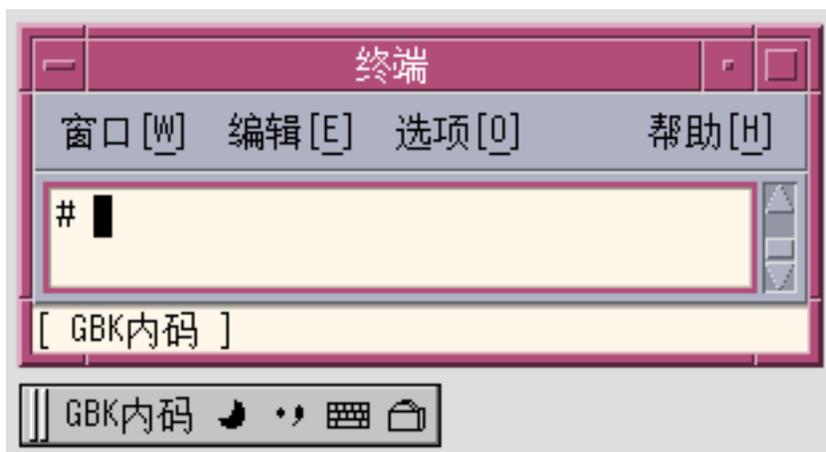
4. Type the first three of the four keys that represent the character to display. In this example, type `b0a` of the string `b0a1`.

The first three letters are visible in the preedit area.

5. Type the fourth key.

The character automatically replaces the letters in the preedit area.

GB2312 Code Input Method

The GB2312 code input method uses the GBK code defined by the Chinese Internal Code Specification. This specification includes all of the Chinese characters and symbols in GB2312-80, and other CJK Chinese characters in GB 13000-1. Each Chinese character or symbol is identified by a four-hexadecimal digital internal code defined in the Chinese Internal Code Specification.

▼ How to Type GB2312 Code Text

This procedure describes how to use the GB2312 codes to type Chinese characters and symbols.

1. **Select the input method.**

   Follow the steps in the procedure to select the input method.

   The status area shows that the GB2312 code input mode is on.
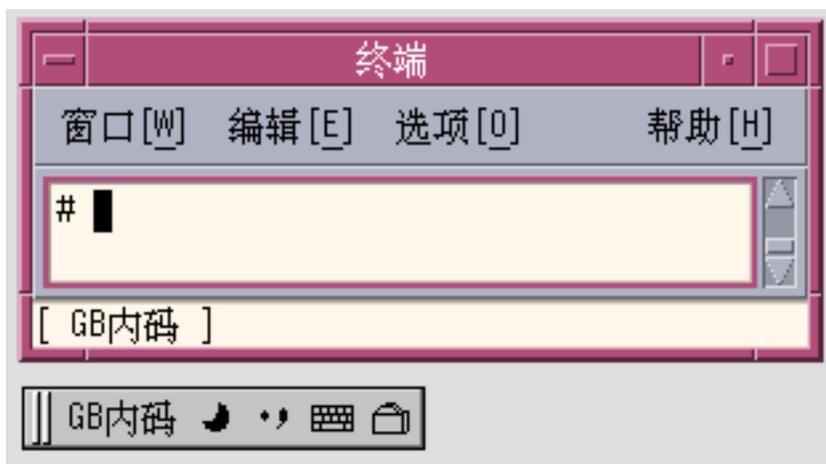
2. Type the first three of the four keys that represent the character to display. In this example, type `b0a` of the string `b0a1`.

The first three letters are visible in the preedit area.

3. Type the fourth key.

The character automatically replaces the letters in the preedit area.

GB18030 Code Input Method

The GB18030 code input method uses the GB18030 code defined by the Chinese Internal Code Specification. This method includes all of the Chinese characters and symbols in GB2312-80, and other CJK Chinese characters in GB 18030. Each Chinese character or symbol is identified by a four-hexadecimal or eight-hexadecimal digital internal code defined in the Chinese Internal Code Specification.

▼ How to Type GB18030 Code Text

This procedure describes how to use the GB18030 codes to type Chinese characters and symbols.

1. **Select the input method.**

   Follow the steps in the procedure "How to Select an Input Method" on page 34 to select the input method.

   The status area shows that the GB18030 code input mode is on.

2. **For example, to input Chinese GB18030 character with code 0xb0a1, press the first three of the four keys that represent the character to display. In this example, type `b0a` of the string `b0a1`.**

   The first three letters are visible in the preedit area.

3. **Type the fourth key.**

   The character automatically replaces the letters in the preedit area.

4. To input a Chinese GB18030 character with code 0x82358538, press the first seven of the eight keys that represent the character to display. In this example, type `8235853` of the string `82358538`.

The first seven numbers are visible in the preedit area.

5. Type the last key.

The character is automatically committed to the window.

QuanPin Input Method

The QuanPin input method requires up to six keystrokes to type each Chinese PinYin character. QuanPin maps PinYin phonetics to single lowercase Roman letters. You can use the QuanPin input method to type individual Chinese characters in both the zh_CN.EUC and zh_CN.GBK locales.

▼ How to Type QuanPin Text

This procedure describes how to use the QuanPin input method to type the character that represents the Full PinYin word fang. For information on making the lookup choices used in this procedure, see "How to Search and Select Lookup Choices" on page 44.

1. **Select the input method.**

   Follow the steps in the procedure "How to Select an Input Method" on page 34 to select the input method.

2. **Type the four keystrokes fang.**
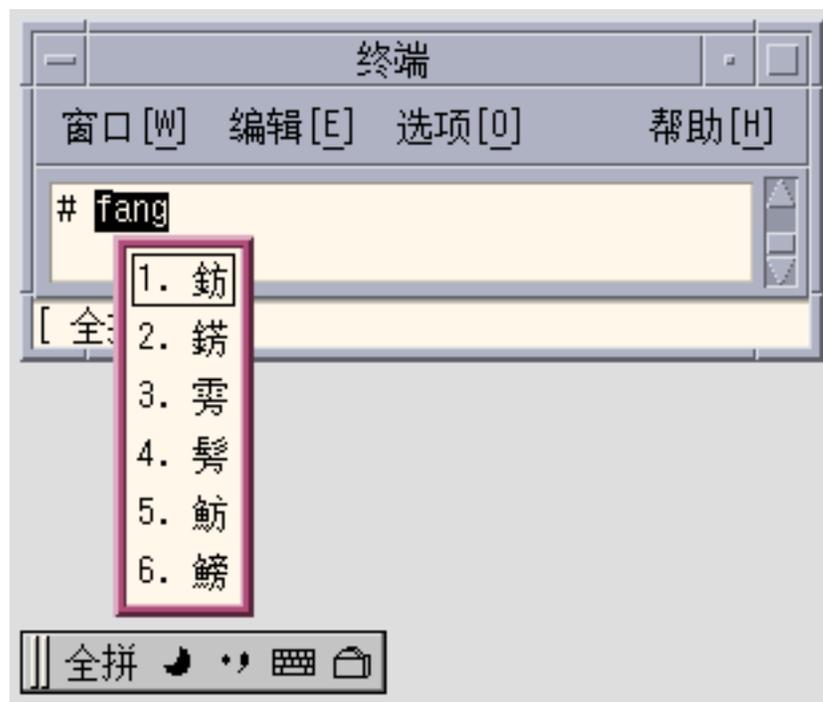
3. Type 1 to select the corresponding GBK Chinese character in the lookup choice list.

Your choice is substituted for the Full PinYin string in the preedit area.

English_Chinese Input Method

You can use the English_Chinese input method in both `zh_CN.EUC` and `zh_CN.GBK` locales. With this method, you type English words of up to 15 keystrokes that are mapped to Chinese phrases. For each keystroke, a lookup window displays characters that match your input. To select a character, you type the number that corresponds to your lookup choice. For more information, see "How to Search and Select Lookup Choices" on page 44.

▼ How to Type English_Chinese Text

The following procedure shows you how to use this input method to enter the Simplified Chinese phrase for the English word, **world**.

1. **Select the input method.**

   Follow the steps in the procedure "How to Select an Input Method" on page 34 to select the input method.

2. **Type the five keystrokes `world`.**

3. **Type 3 to select the corresponding Chinese phrase from the lookup choice list.**
   Your choice is substituted for the English string in the preedit area.

▼ How to Use Wildcard Characters in English_Chinese Lookups

You can use the wildcard characters asterisk (*) or question mark (?) to search a system dictionary. The * stands for one or more letter. The ? represents only one letter.

1. **To search for all the English words that end with `lution`, type input `*lution`. The lookup choice window appears as shown in the following figure.**
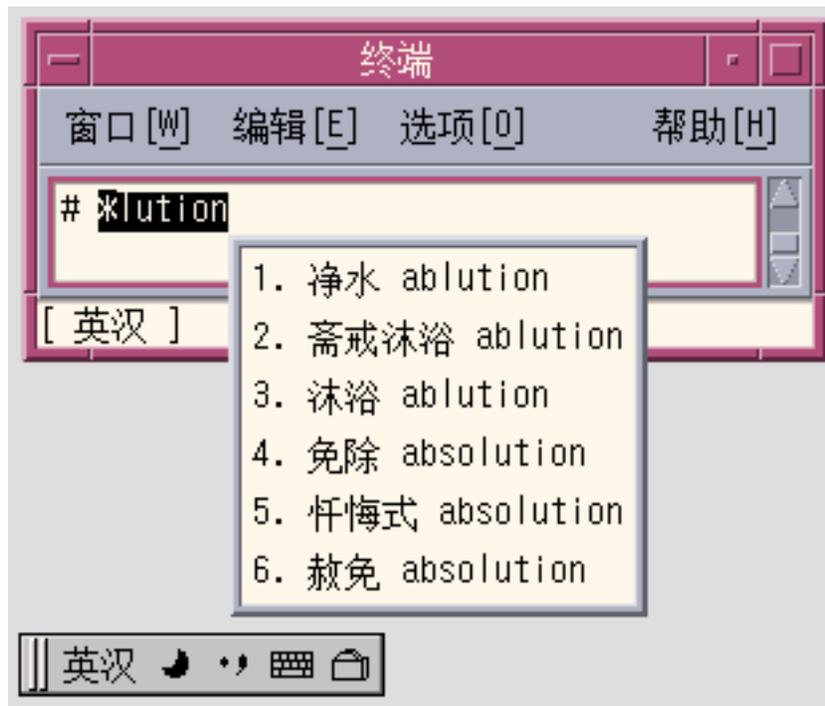
2. **To search for all three-letter English words which begin with `c`, type `c??`.**
   The lookup choice window appears as shown in the following figure.

終端

窗口[W]　编辑[E]　选项[O]　　帮助[H]

\# C???

1. 出租汽车 cab
2. 出租马车 cab
3. 计程车 cab
4. 车上司机室 cab
5. 下流人 cad
6. 无赖 cad

[英汉

英汉
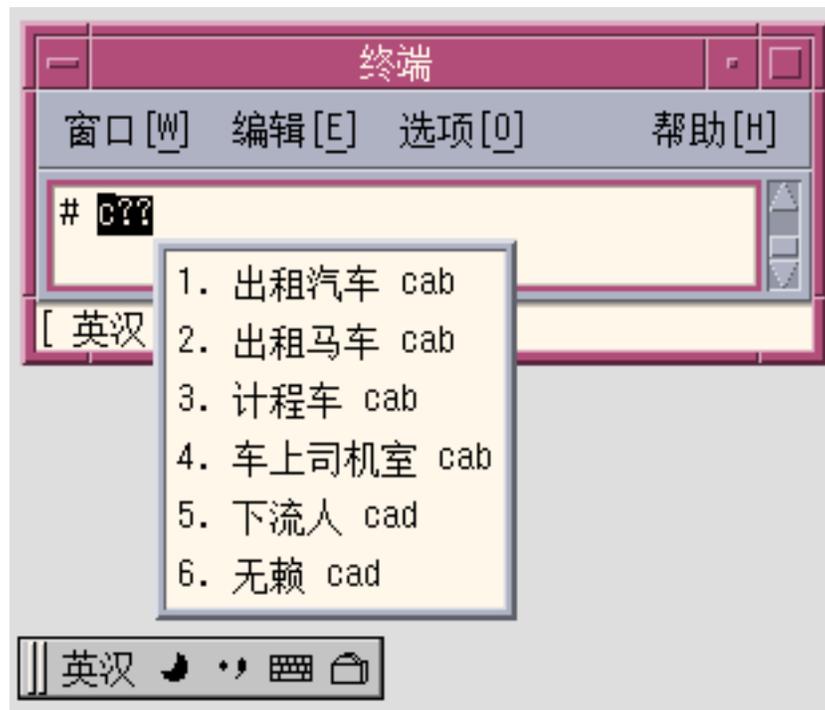
WangMa Wubi Input Method

Wubi is a popular input method in China. The encoding rule used in the Wubi input method is based on the radical or stroke shape of Chinese characters.

One of the main advantages of Wubi and other shape-based input methods is a very low repetition rate. The lower repetition rate, a feature not found in PinYin-based input systems, means that only one or two Chinese characters are represented by a Wubi key sequence. Because a single Wubi code seldom represents more than one character, you can enter text more quickly.

Wubi is built on the GB18030-2000 character set standard, a graphemic encoding system. Almost all Chinese, Kanji, and Hanja characters can be encoded with the GB18030-2000 standard.

This section describes the following features included in this release.

- GB18030-2000 character set support

- Easy character set switching
- New radical mechanism for Simplified and Traditional Chinese
- Three-level progressive identification code
- Phrase input and professional word galleries
- Help key
- Fault tolerance code
- Word-phrase association
- Properties settings

GB18030-2000 Character Set Support

The GB18030-2000 character set is a national encoding standard issued by the Chinese government in 2000. The encoding length set by the standard is one, two, or four bytes. GB18030-2000 includes 6,763 standard Simplified Chinese characters, 13,053 Traditional Chinese (Big5) characters, 3,000 characters used in Hong Kong, and 21,003 GBK characters. The Wubi input method supports the GB18030-2000 character set, which makes it working with the smaller character sets contained in GB18030-2000 easy. See "Easy Character Set Switching" on page 78.

For example, if you type the letters gigg and scroll pages to the end, you will find a GB18030 character shown in the following figures:

Easy Character Set Switching

Solaris WangMa Wubi divides the GB18030-2000 character set into smaller sets of commonly used Chinese characters.

- GB2312, which contains 6,763 characters

- GBK, which contains 21,003 characters
- GB18030-2000, which contains 27,533 characters

When you enter text, you can use the following keyboard shortcuts to switch between character sets.

- To use the GB2312 character set, press Control-Shift-1.
- To use the GBK character set, press Control-Shift-2.
- To use the GB18030-2000 character set, press Control-Shift-3.

Because GB18030-2000 is a relatively new standard, support in Wubi for the GB2312 and GBK character sets ensures backward-compatibility with earlier standards. You might prefer to work in the GB2312 or GBK character set because of improved performance and lower repetition rates.

New Radical Mechanism for Simplified and Traditional Chinese

The new radical, or root, mechanism is a patented technology invented by professor Wang Yongmin who invented Wubi. Professor Yongmin developed from the mechanism from version 86, the old radical system. The mechanism has evolved into a new encoding system compatible with both Simplified and Traditional Chinese. Users of Wubi version 86 can work with three times more characters, using the same encoding and typing rules, without additional training.

Three-Level Progressive Identification Code

One of the main features of Wubi is the last-stroke grapheme identification codes that distinguish between characters of a similar shape. The identification codes are assigned according to the shape of the last radical of the character. The purpose of identification codes is to help users master the Wubi input method at three different levels.

- In level A, for beginning users, all three graphemic types with less than four codes have identification codes.

- In level B, for intermediate users, only the left-right shaped Chinese characters have identification codes.

- In Level C, for advanced users, identification codes are not used.

Phrase Input and Optional Professional Word Galleries

Wubi supports phrase input. In addition to individual characters, entire phrases can be assigned Wubi codes. In addition to 90,000 basic phrases, there are 11 professional word galleries, similar to glossaries, for each of the following industries:

- Traffic and transportation
- Computer and household electronics
- Economy and finance
- Medicine and health
- Mining and metallurgy
- Foreign trade and travel
- Military affairs and national defense
- Law and aesthetics

Galleries also exist for place names and for idioms.

You can select word galleries that contain between 3,000 and 20,000 entries. in the Preferences dialog box.

For example, when you choose the Medicine and Health phrase gallery and type the word `mino`, medical phrases are listed for selection.

Encoding Help Feature

The Solaris Wubi input method supports encoding hint features. As you type, the character encoding appears in the Select Repetition Code Window. This feature can help you master the encoding methods and codes of Chinese characters. In addition, you can use the uppercase or lowercase Z key as a wildcard at any time. Z is the only key not mapped to a character in Wubi. To help you learn to use Wubi, you can press the Z key to query the system for input codes.

For example, when you can type `azzd` to search all characters or phrases with a Wubi code that begins with the letter A and ends with the letter D.

Fault Tolerance Code

According to the preferences you set, the fault tolerance code feature can increase the probability that the system will provide the correct character even when you make a typing mistake.

Word-Phrase Association

The word-phrase feature is another productivity aid. The system provides a list of characters that are most likely to follow the character just selected. Instead of typing a code, the system provides a list of likely options from which you can choose the correct character. This feature is also accessed in the Preferences dialog box.

For example, when you type the letters `iuxx`, the Chinese character ×Ì is automatically committed to application. After the character appears in application window, a new candidate window will display and the phrases which begin with this Chinese character will be listed in this candidate window.
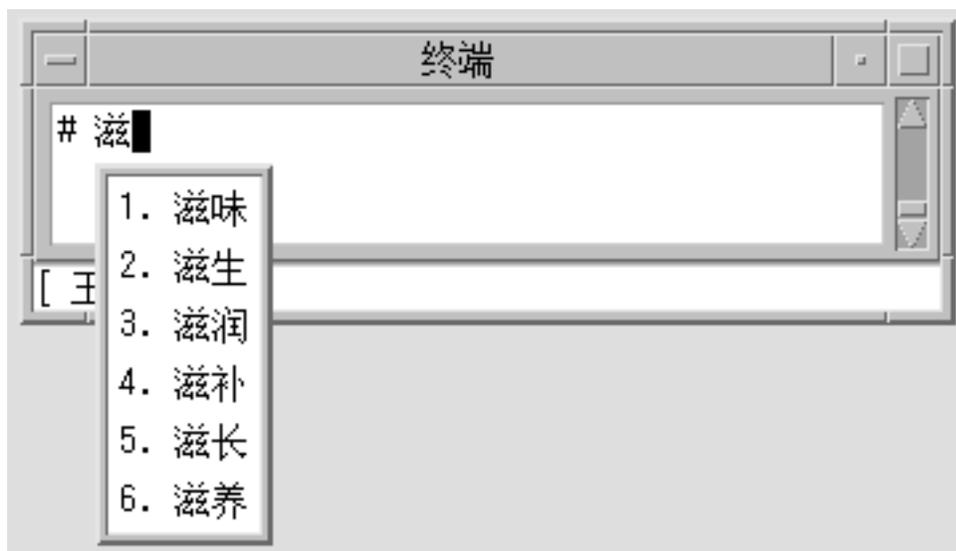
Properties Settings

You can make the following settings in the Properties dialog box:

- Character sets: GB2312, GBK, or GB18030
- Professional word galleries
- Identification code mode
- Display the Wubi code for a candidate
- Display the candidates after each keystroke
- Association of characters with phrases
- Fault tolerance code
- Display characters and phrases with the same code
- Display the key prompt in the preedit area

The following graphics show the settings in the Properties dialog box.

输入法选项设置

已选输入法：   输入法选项：

全拼
双拼
英汉
王码五笔

汉字字集：   GB18030（27533 汉字）

专业词库：   医药卫生

识别码模式：   A 模式（带识别码）

☑ 五笔编码提示

☐ 即打即现

☑ 字词联想

☑ 允许容错码

☑ 允许字词重码

☐ 键对应提示

确定   应用   重设   取消

输入法选项设置

已选输入法：   输入法选项：

全拼
双拼   汉字字集： GB18030（27533 汉字） ⊟
英汉
王码五笔   专业词库： 医药卫生 ⊟

                       识别码模式： A 模式（带识别码）   ⊟
                                  B 模式（左右型汉字带识别码）
               ☑ 五笔编码提   C 模式（不带识别码）

               ☐ 即打即现

               ☑ 字词联想

               ☑ 允许容错码

               ☑ 允许字词重码

               ☐ 键对应提示

     确定   应用   重设   取消

Code Table Input Method Interface

The current Solaris Operating System provides a code table input method interface that enables Chinese users to add new input methods into their system.

Creating a Code Table

A code table is a plain text dictionary file that contains a list Chinese characters, words, and phrases that are mapped to input keystrokes. When you type the specified keystrokes, the associated characters, words, and phrases appear for selection.

The code table file contains the following sections:

- [Description] This section lists the distinguishing characteristics of the code table.

    Name: Code table name.
    Encode: UTF-8, GB, GB2312, GBK, or BIG5 encoding used the code table.
    WildChar: Wild character used for input codes.
    UsedCodes: Valid characters for input.
    MaxCodes: Maximum number of input codes for one item.

- [Comment]

- [Key_Prompt] This section identifies the prompt string of an input key. The prompt string appears in the preedit area of the application subwindow.

- [Function_Key] This section describes the behavior specified function keys.

    PageUp: Scroll up a list of selection items.
    PageDown: Scroll down a list of selection items.
    BackSpace: Delete an input code.
    ClearAll: Clears all the input areas, such as preedit area and the lookup area.

- [Phrase] This section associates input codes with corresponding Chinese phrases. The phrases must be separated by spaces. The format of each line is:`keystroke_sequence word1 word2 word3 ....`

- [Single] This section associates input codes with corresponding Chinese characters. The format of each line is: `keystroke_sequence Characterlist`. The characters of the `Characterlist` are not separated by spaces.

- [Options] This section specifies the options that you toggle on or turn off for the code table input method.

    HelpInfo_Mode: Display help information.

    KeyByKey_Mode: Display lookup candidates key by key or only when the spacebar is pressed.

    KeyPrompt_Mode: Display the prompt string of the input key in the preedit area.

    AutoSelect_Mode: Commit the lookup choice automatically when only a single candidate is available.

    SelectKey_Mode: Select numbers, uppercase letters, or lowercase letters.

The following example shows a code table file.

```
# HANZI codetable input table

[Description]
Name:          英汉
Encode:        GB2312
UsedCodes:     -ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz
WildChar:      *?
MaxCodes:      15

[Function_Key]
PageUp:        ,<
PageDown:      .>

[Options]
KeyByKey_Mode:         ON
HelpInfo_Mode:         ON
AutoSelect_Mode:       OFF
KeyPrompt_Mode:        ON
SelectKey_Mode:        Number

[Phrase]
Chinese 中国人 中国人的 中国的 中文 汉语
Christ   基督 救世主
Christian      信徒 基督徒 基督教徒 基督教的
```

▼ How to Create a New Code Table Input Method

1. Create and edit the code table source file.

Prepare the code table source file to define the mapping of characters, words, or phrases to input keystrokes.

2. Convert the source code table file to binary format.

Use the txt2bin utility to convert the code table text file to binary file.

```
# /usr/lib/im/locale/zh_CN/common/txt2bin \
source_codetable_file binary_codetable_file
```

Note – You can find the txt2bin and bin2txt are utilities in the directory:
/usr/lib/im/locale/zh_CN/common/

---

3. Add the code table to the input method specification file,
/usr/lib/im/locale/zh_CN/sysime.cfg.

For example, if your new code table binary file is called newim.data, add the entry newim to the input method specification file, sysime.cfg.

4. **Restart the htt input method server by typing the following commands as root.**

```
# /etc/init.d/IIim stop
```

```
# /etc/init.d/IIim start
```

The new input method is ready to use when you log in to the system.

Localized Applications

The following sections in this chapter describe the utilities and applications that you use in the Simplified Chinese Solaris Operating System.

- "Using the `mailx` Utility With Simplified Chinese Characters" on page 91
- "Using `talk` With Simplified Chinese Characters" on page 92
- "The `xtobdf` Utility" on page 92
- "`stdconvtool` Utility" on page 92
- "`iconv` Command" on page 94

Using the `mailx` Utility With Simplified Chinese Characters

The `encoding` variable in `.mailrc` does not affect Mailer behavior, but it does set `mailx` function. To send email in formats other than 7-bit ASCII, such as to send (8-bit) Simplified Chinese characters you must use `/usr/SUNWale/bin/mailx`. The `encoding` variable in `.mailrc` sets encoding formats for `mailx`.

International transmission conventions require that header information use only ASCII characters. Simplified Chinese characters should not be used in the header (including the Subject line) with Mailer or with `mailx`.

Note – The `/bin/mailx` application has not been localized and cannot send or receive Simplified Chinese characters.

Using `talk` With Simplified Chinese Characters

To use Simplified Chinese characters with `talk`, the `/usr/SUNWale/bin/talk` application is required because `/bin/talk` is not localized.

The `xtobdf` Utility

Simplified Chinese Solaris software provides this BDF font generator:

`xtobdf` — Convert from font in X server to font in BDF

`stdconvtool` Utility

The `sdtconvtool` graphic user interface utility enables file conversion between various code sets. The `sdtconvtool` functionality is similar to `iconv`.

The following figure shows the `stdconvtool` panel.
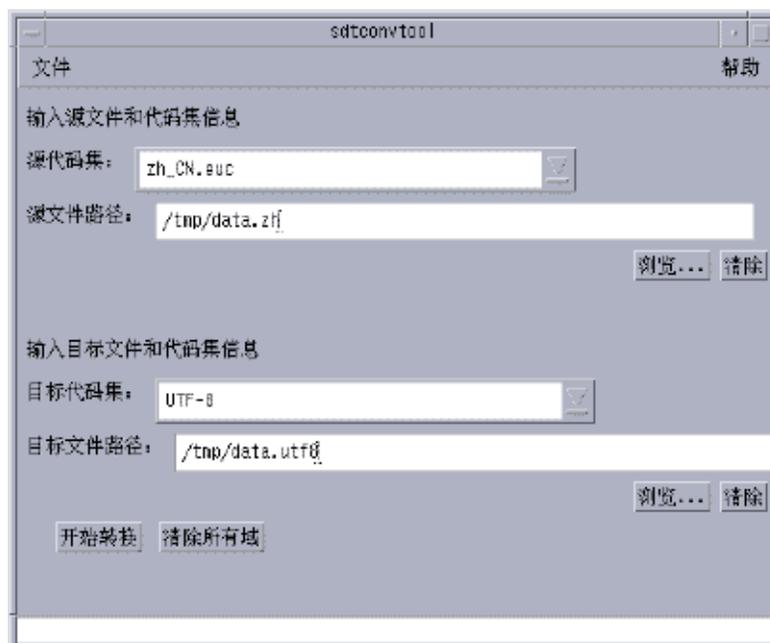
▼ How to Convert a File Using the `stdconvtool` Utility

1. **Select the code set of the file to be converted.**

   Scroll through the pull-down list and select the code set of the file to be converted.

2. **Enter the path of the file to be converted.**

   You can enter the path manually in the source file path area, or you can use the browse button to find and select the file.

3. **Select the code set to for the converted file.**

   Select the target code set.

4. **Enter the path for the target file.**

   You can enter the path manually in the target file path area, or you can use the browse button to find and select the file.

5. **Click the start conversion button.**

`iconv` Command

The `iconv` command converts the characters or sequences of characters in a file from one code set to another. The command then writes the results to standard output. The Simplified Chinese Solaris software includes special filters for the `iconv` command.

If no conversion exists for a particular character, the character is converted to the underscore _ in the target code set. The following options are supported:

- `-f` *from-code* — Symbol of the input code set
- `-t` *to-code* — Symbol of the output code set

The following table lists the code set conversion modules that are supported in Simplified Chinese Solaris software. For more information, see *Solaris 10 Reference Manual Collection* `iconv`(1).

TABLE 5–1 Simplified Chinese `iconv` Code Conversion Modules (`zh` locale)

Code	Symbol	Target Code	Symbol
ISO2022–CN	`zh_CN.iso2022-CN`	UTF–8	`UTF-8`
UTF–8	`UTF-8`	ISO2022-CN	`zh_CN.iso2022-CN`
zh.GBK	`zh_CN.gbk`	ISO2022-CN	`zh_CN.iso2022-CN`
zh.GBK	`zh_CN.gbk`	UTF–8	`UTF-8`
GB2312-80	`zh_CN.euc`	ISO 2022-7	`zh_CN.iso2022-7`
ISO 2022-7	`zh_CN.iso2022-7`	GB2312-80	`zh_CN.euc`
GB2312-80	`zh_CN.euc`	ISO 2022-CN	`zh_CN.iso2022-CN`
ISO-2022-CN	`zh_CN.iso2022-CN`	GB2312-80	`zh_CN.euc`
UTF-8	`UTF-8`	GB2312-80	`zh_CN.euc`
GB2312-80	`zh_CN.euc`	UTF-8	`UTF-8`
GB2312-80	`zh_CN.euc`	BIG5	`zh_TW-big5`
BIG5	`zh_TW.big5`	GB2312-80	`zh_CN.euc`
HZ-GB-2312	`HZ-GB-2312`	GB2312–80	`zh_CN.euc`
GB2312–80	`zh_CN.euc`	HZ-GB-2312	`zh_CN.euc`

TABLE 5–2 Simplified Chinese `iconv` Code Conversion Modules (`zh.GBK` locale)

Code	Symbol	Target Code	Symbol
UTF-8	`UTF-8`	GBK	`zh_CN.gbk`
GBK	`zh_CN.gbk`	UTF-8	`UTF-8`
GBK	`zh_CN.gbk`	BIG5P	`zh_TW-big5p`
GBK	`zh_CN.gbk`	BIG5HK	`zh_TW-big5hk`
GBK	`zh_CN.gbk`	ISO-2022-CN	`zh_CN.iso2022-CN`
ISO2022-CN	`zh_CN.iso2022-CN`	GBK	`zh_CN.gbk`
GBK	`zh_CN.gbk`	BIG5	`zh_TW-big5`
BIG5	`zh_TW-big5`	GBK	`zh_CN.gbk`
BIG5P	`zh_TW-big5p`	GBK	`zh_CN.gbk`
BIG5HK	`zh_TW-big5hk`	GBK	`zh_CN.gbk`
HZ-GB-2312	`HZ-GB-2312`	GBK	`zh_CN.gbk`
GBK	`zh_CN.gbk`	HZ-GB-2312	`zh_CN.gbk`
HZ-GB-2312	`HZ-GB-2312`	UTF-8	`UTF-8`
UTF-8	`UTF-8`	HZ-GB-2312	`HZ-GB-2312`

TABLE 5–3 Simplified Chinese `iconv` Code Conversion Modules (`zh_CN.GB18030` locale)

Code	Symbol	Target Code	Symbol
UTF-8	`UTF-8`	GB18030-2000	`zh_CN.gb18030`
GB18030-2000	`zh_CN.gb18030`	UTF-8	`UTF-8`
GB18030-2000	`zh_CN.gb18030`	BIG5HK	`zh_HK-big5hk`
GB18030-2000	`zh_CN.gb18030`	BIG5P	`zh_TW-big5p`
BIG5HK	`zh_HK-big5hk`	GB18030-2000	`zh_CN.gb18030`
BIG5P	`zh_TW-big5p`	GB18030-2000	`zh_CN.gb18030`

The following `iconv` code conversion modules are located in `/usr/lib/iconv`:

For the `zh` locale:

- `zh_CN.euc%zh_TW-big5.so`
- `zh_TW-big5%zh_CN.euc.so`

For the `zh.GBK` locale:

- `UTF-8%zh_CN.gbk.so`

- `zh_CN.gbk%UTF-8.so`
- `zh.CN.gbk%zh_CN.iso2022-CN.so`
- `zh_CN.iso2022-CN%zh_CN.gbk.so`
- `zh_CN.gbk%zh_TW-big5.so`
- `zh_TW-big5%zh_CN.gbk.so`

In the following example, an EUC mail file is converted to ISO 2022-CN:

`system% **iconv -f zh_CN.euc -t zh_CN.iso2022-CN** *mail.euc* **>** *mail.iso2022-CN*`

For further information, see the `iconv(3C)` and the `iconv_zh(5)` man pages. These
utilities can be used for converting files for printing.

Setting Up Solaris Simplified Chinese Printing Facilities

The Simplified Chinese Solaris Operating System supports printing Simplified Chinese output through the following types of printers:

- Line printer with built-in Simplified Chinese fonts
- PostScript-based printer with built-in scalable fonts
- Any PostScript-based printer for bitmap printing

Note – Review the manufacturer's documentation on installing the printer before you complete the procedures in this chapter.

Line Printer Support

For the Simplified Chinese Solaris Operating System to run a line printer, the printer must recognize EUC.

Using the EUC to GB Code Filters

A printer that does not support EUC needs filters that convert EUC files for printing. Use the commands in this section to print EUC files to non-EUC printers.

The following commands install the printer lp1 on port ttya. The commands signal the print service that lp1 accepts only GB format files.

```
# lpadmin -p lp1 -v /dev/ttya -I GB
# accept lp1
# enable lp1
```

See the lpadmin(1M) man page for more information.

You can use an lpfilter command shown in the following example to print files with formats that are not supported by the printer. The command line signals the print service that a converter called *filter-name* is available through the filter description file named in *pathname*.

lpfilter -f *filter-name* **-F** *pathname*

The following example shows the output of *pathname* for a converter called euctogb. The *pathname* filter converts the default input type to GB with the euctogb converter.

```
Input types: simple
Output types: GB
Command: euctocgb
```

To print an EUC file, use a command line such as the following.

system% **lp** EUC-filename

To print a GB format file, use a command line such as the following.

system% **lp -T GB** *GB-filename*

Laser Printer Support

An application must have the mp utility to print Simplified Chinese characters.

Using the mp Utility

The mp utility supports all Asian locales including UTF-8 locales. As a printing filter, mp generates a properly formatted version of the file content in PostScript format. Depending on the locale's system font configuration for mp, the Postscript output file contains glyph images from a scalable or a bitmap system font. The mp utility is enhanced in this release to print files of a certain type for each locale. For more information, see the mp(1) man page.

You can use a command such as the following to print a file with Simplified Chinese characters. The file might also include ASCII/English characters.

system% **mp** *filename* | **lp -d printer**

System Administrator Information

This part includes system administrator information.

CHAPTER **7**

Configuring a Simplified Chinese Solaris System

You can use the procedures this chapter to set a system-wide default locale or to add a new locale from the Solaris DVD and CD media. You also change the font path set by the Xsession script included in the Simplified Chinese Solaris Operating System. The following sections are included in this chapter.

- "Setting the Default Locale" on page 101
- "Using Aliases to Set the Locale" on page 102

Setting the Default Locale

You can make a system-wide change to the default locale with the following procedure.

▼ How to Set the Default Locale

1. **Become superuser.**

2. **Edit the `/etc/default/init` file.**

3. **Add or update the `LANG` variable with the appropriate locale.**

   LANG=*locale*

   You can set the LANG variable with one of the following locales: zh, zh.GBK, zh_CN.GB18030, zh.UTF-8

4. **Instruct all users to exit CDE.**

5. **Kill the `dtconfig` process.**

   # **/usr/dt/bin/dtconfig -kill**

6. **Reboot the system.**

   ```
   # reboot
   ```

Using Aliases to Set the Locale

You can use aliases to change locales in a terminal window without typing long command lines or editing your `.dtlogin` file and running `source` every time.

▼ How to Set Locale Aliases for the Korn Shell

● **Put aliases such as the following in your `.profile` file.**

```
alias chn "export LANG=zh; /bin/stty cs8 -istrip defeucw"
alias chn_GBK "export LANG=zh.GBK; /bin/stty cs8 -istrip \defeucw"
alias asc "export LANG=C; /bin/stty cs7 istrip defeucw"
alias chn "export LANG=zh; /bin/stty cs8 -istrip defeucw"
```

▼ How to Set Locale Aliases for the C Shell

● **Put aliases such as the following in your `.cshrc` file.**

```
alias chn_EUC "setenv LANG zh; /bin/stty cs8 -istrip defeucw"
alias chn_GBK "setenv LANG zh.GBK; /bin/stty cs8 -istrip \defeucw"
alias asc "setenv LANG C; /bin/stty cs7 istrip defeucw"
alias chn_UTF-8 "setenv LANG=zh.UTF-8; /bin/stty cs8 -istrip defeucw"
```

Managing the `htt` Input Method Server

This chapter describes the basic display features of the input method server. The input method server (IM server) handles Simplified Chinese input for Simplified ChineseSolaris software. The IM server receives keyboard input and makes the input available as Simplified Chinese characters to Simplified Chinese Solaris applications. The IM server can serve any internationalized X Window application that uses X Window Input Method (XIM) application program interfaces (API) to receive language input.

The following sections of this chapter explain the basic display features of the IM Server.

Input Method Server Basic Properties

The Simplified Chinese Solaris Operating System starts the IM server automatically when you start the Simplified Chinese in a locale that requires an input method. The IM server continues to run and service applications that are started and connected to it.

Simplified Chinese Solaris applications that use the IM server for Simplified Chinese character input typically find the IM server running when they start. To user the IM server's service for language input, the IM server should be running before an application is started. If the IM server is not running when an application is started, the application might not be able to get the input service even if IM server is started later.

Three `htt` Processes

The input method server comprises three related processes running together. The processes are `htt`, `htt_xbe`, and `htt_server`. One process controls the input method server properties, another controls the population of the input method server (that is, a "watchdog" process), and the third handles input methods of clients.

Usually, you don't need to know the details of these three processes. When `htt` is started, `htt_xbe` and `htt_server` are started automatically.

`htt` Application Display

The following figure shows a typical application display when using the `htt` input method.
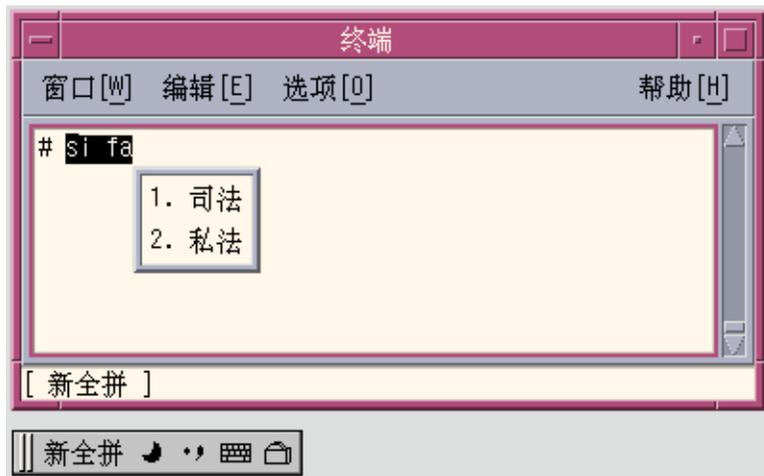
How and When `htt` Is Started

The `htt` process must be running before an application starts in order for the application to use the `htt` input method server to receive Simplified Chinese input. The `C` locale does not require `htt`. If you change the default locale from `C` to another locale that requires `htt`, you must start `htt` in same terminal window. You can start `htt` with a command such as the one shown in the following example.

```
system% /usr/openwin/bin/htt -if twle -xim htt_xbe &
```

When the locale is set to `zh`, `htt` is started automatically when the Simplified Chinese windowing environment starts. In CDE, `htt` is started from the `/usr/dt/config/Xsession.d/0020.dtims` script. This script is executed by `Xsession`, which runs at every initialization stage upon a user's login from `dtlogin`. This script ensures that `htt` is started before other applications in CDE.

Using the `iiim` Server in `zh.UTF-8` and `zh_CN.UTF-8` Locales

The Internet Intranet Input Method Server (iiim) supports the European Local Input Methods that use the Compose Key. The iiim server supports the Asian Remote Input Methods using IIIM Protocol. The default Language Engine is English/European which supports ASCII and some European Languages such as German and French.

To switch to Chinese input methods, press Control-spacebar.

To switch to other language engines, click the left mouse button in the status area.

The following language engine list appears:

- Cyrilllic
- Greek
- Thai
- Arabic
- Hebrew
- Hindi
- Unicode Hex
- Unicode Octal
- Lookup
- Simplified Chinese

The following language engines appear if their corresponding locales are installed:

- Japanese — if `ja` locale is installed
- Korean— if `ko` locale is installed
- Traditional Chinese — if `zh_TW` locale is installed
- Traditional Chinese (Hong Kong)— if `zh_HK` locale is installed

After the Simplified Chinese language engine is selected, you can use the following function keys to switch between different input methods:

- F2 — New PinYin
- F3 — New Shuang Pin
- F4 — QuanPin
- F5 — ShuangPin

- F6 — English_to_Chinese

For more detailed information about Input Methods, see Chapter 4.

Fonts

This chapter describes the PostScript, TrueType and bitmap fonts supported for use in the Simplified Chinese Solaris Operating System. The following sections discuss how to use and to edit the fonts.

Display PostScript System (DPS)

The Simplified Chinese Solaris Operating System provides PostScript fonts in the Display PostScript System (DPS). This section describes the fonts and the DPS facilities. For further details, see *Programming the Display PostScript System with X*, published by Adobe® Systems.

Using Simplified Chinese PostScript Fonts and DPS Facilities

The Simplified Chinese Solaris Operating System includes the fonts listed in the following table.

TABLE 9–1 Simplified Chinese Solaris Operating System PostScript Fonts

Font Name	Description
Song-Medium	Alias of Song-Medium-EUC; can be used like a Roman font.
Song-Medium-EUC	Song-Medium font, EUC encoding, horizontal display; can be used like a Roman font.
Kai-Medium	Alias of Kai-Medium-EUC; can be used like a Roman font.
Kai-Medium-EUC	Kai-Medium font, EUC encoding, horizontal display; can be used like a Roman font.
FangSong-Medium	Alias of FangSong-Medium-EUC; can be used like a Roman font.
FangSong-Medium-EUC	FangSong-Medium font, EUC encoding, horizontal display; can be used like a Roman font.
Hei-Medium	Alias of Hei-Medium-EUC; can be used like a Roman font.
Hei-Medium-EUC	Hei-Medium font, EUC encoding, horizontal display; can be used like a Roman font.

The following figure shows a sample of the Song-Medium font.

窗口[W]　编辑[E]　选项[O]　　　　　　　提示[H]

```
nanjing# /usr/openwin/demo/dpsexec
PostScript(r) Version 2015.108
(c) Copyright 1984-1996 Adobe Systems Incorporated.
Typefaces (c) Copyright 1981 Linotype-Hell AG and/or
its subsidiaries.
All Rights Reserved.
PS>/Song-Medium 30 selectfont
PS>160 110 moveto
PS>(中国) show
PS>
```

Display PostScript Executive

中国

Creating Composite Roman and Simplified Chinese Fonts

You can create composite fonts using any a Roman font and any of the Simplified
Chinese fonts. For example, the following PostScript code defines a composite font,
Times-Italic+Kai-Medium. The Times-Italic is used for the ASCII characters and
Kai-Medium font is used for the Simplified Chinese characters.

```
/makeEUCfont {
        /AsianFont exch def
        /WestFont  exch def
        /NewFont   exch def

        /CIDInit /ProcSet findresource
        begin
```

```
          NewFont [ AsianFont WestFont ]
          beginrearrangedfont
            1 usefont
            1 beginbfrange
              <00> <7e>  <00>
            endbfrange
          endrearrangedfont
        end
  } bind def

  /LC_Times-Roman /Times-Roman /Kai-Medium makeEUCfont
```

Using Simplified Chinese Fonts in DPS Programming

You can use Simplified Chinese fonts just as you use Roman fonts in DPS wrap
definitions. The following code sample creates the display in the Hello World figure.

```
  defineps PSWDisplayText(char *text)
          /pointSize 50 def
          /Helvetica pointSize selectfont
          (Hello World) stringwidth pop 2 div neg 0 moveto
          (Hello World) show

          /cpSize 40 def
          /Song-Medium cpSize selectfont
          (text) stringwidth pop 2 div neg pointSize neg moveto
          (text) show
  endps
```

You can call PSWDisplayText(*Chinese text*) in a C program to display the
designated Chinese text.

The Simplified Chinese Solaris software provides TrueType support in DPS.

BDF to PCF Conversion

You must convert BDF fonts to PCF format to make the fonts usable in Simplified Chinese Solaris applications.

▼ How to Convert BDF Fonts to PCF Format

1. **Use the `bdftopcf` command to convert a BDF font to PCF format.**

   ```
   system% bdftopcf -o myfont14.pcf myfont14.bdf
   ```

2. **Replace the BDF file in the following directory with the converted PCF file:**
   `$OPENWINHOME/lib/locale/`*locale-name*`/fonts`

▼ How to Install and Check the Edited Font

1. **Compress the `.pcf` font file and move it to your font directory.**

   ```
   system% compress myfont14.pcf
   ```

2. **Run the following commands in your font directory.**

   The `.bdf` file should not be in the font directory.

   ```
   system% cat >> fonts.alias
   -new-myfont-medium-r-normal--16-140-75-75-c-140-gb2312.1980-0
   Myfont-Medium14
   ^D
   system% mkfontdir
   system% xset +fp `pwd`
   ```

3. **Use a command line such as the following to view your font.**

   ```
   system% xfd -fn Myfont-Medium14
   ```

All Simplified Chinese fonts have XLFD font names. You can use the `xlsfonts` utility to display the font names.

```
system% xlsfonts | grep gb2312
```

Creating Characters With `sdtudctool`

The `sdtudctool` utility supports bitmap, Type 1, and CID fonts. This section contains instructions for using the `sdtudctool` to create, input, and print new characters for the Simplified Chinese locales in the following code ranges.

TABLE 9–2 Code Points for Locales

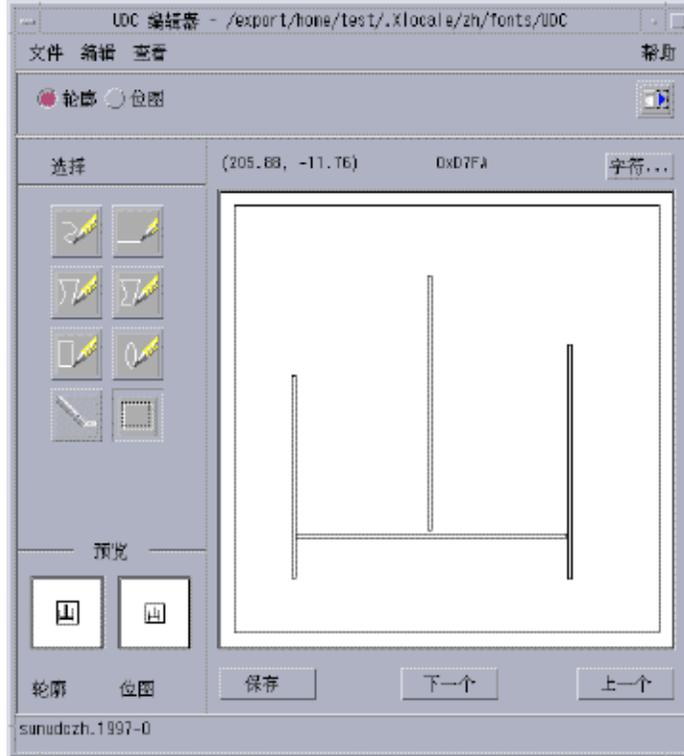| Locale	Code Ranges
zh	0xD7FA-0xD7FE
zh.GBK	0xA140-0xA7A0
	0xAAA1-0xAFFE
	0xF8A1-0xFEFE
zh_CN.GB18030	0xA140-0xA7A0
	0xAAA1-0xAFFE
	0xF8A1-0xFEFE

▼ How to Use the `sdtudctool`

You can use the following procedure to create characters in the `zh` locale.

1. **To start the utility, type the following command:**

   ```
   system% sdtudctool
   ```

The following window appears.

2. **Click the Character button.**

   The following dialog box appears:

3. **Specify the code position of the new character.**

4. **Create the outline or bitmap for the character using the drawing tools.**

5. **Click save from the file menu to save the character.**

   The new character(s) are saved in $HOME/.Xlocale/{locale}/fonts/UDC..

▼ How to Specify `sdtudctool` Options

1. **Choose options from the file menu.**

   The options dialog box appears.

2. **Turn on the append font path option to append the UDC font aliases to your font path.**

3. **Turn on save font path option to add the font files to the system font path.**

   The new font files located in `$HOME/.Xlocale/{locale}/fonts/UDC`) will be listed in the `$HOME/.OWfontpath` file and made available to the X server the next time you log in.

   If Save Font Path is off, the font files will still be saved in `$HOME/.Xlocale/{locale}/fonts/UDC`, but will not be added to the system fontpath the next time you log in.

4. **Choose one or more of the following point sizes for the bitmap fonts: 12, 14, 16, 20, 24.**

▼ How to View a Character Reference

1. **To view the character reference, click the icon below the help menu.**

   The following window appears.
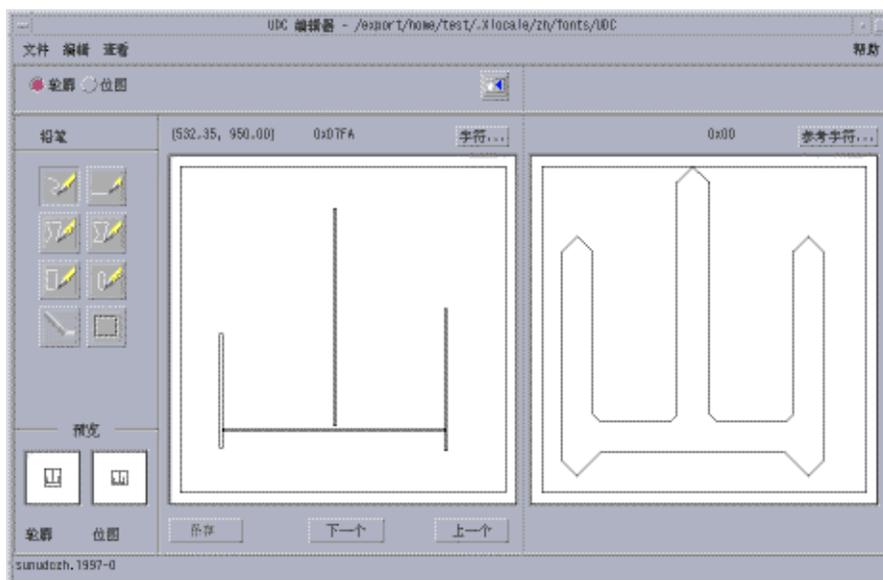
2. **To close the window, click the icon again.**

▼ How to Register the UDC to the NewPinYin Input Method

1. **Choose save from the file menu or choose the open dictionary tool option from the file menu. The `sdtudc_register` tool panel appears.**

2. Input the PinYin in the reading column.

Note – The only the valid PinYin can be registered to NewPinyin input method.

---

3. Click Confirm button

The following dialog box appears.
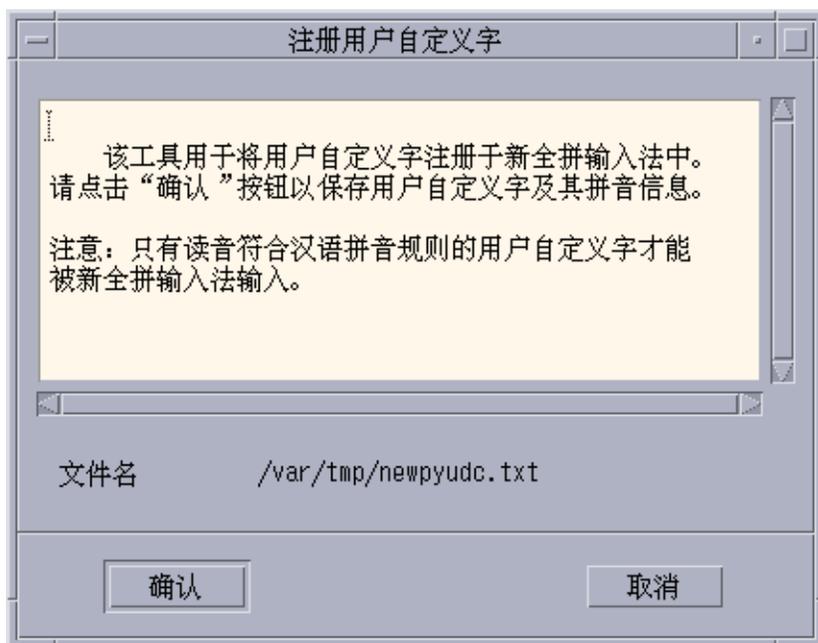
4. **The UDC and its PinYin will be registered in `/var/tmp/newpyudc.txt`. Click OK in the dialog to confirm**

▼ How to Type the New Characters

Use the NewPinyin or NeiMa input method to type the UDC's PinYin spelling for the characters you create.

1. **Type `shan` in NewPinyin mode.**

2. **Type `D7FA` in GB NeiMa mode.**
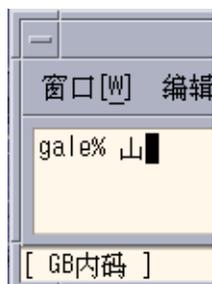
▼ How to Print the New Characters

1. **Set the fontpath for `mp` utility.**

   ```
   system% setenv UDC_BASE $HOME/.Xlocale/zh
   ```

2. **Print a text file which contains UDCs by specifying the configuration file for UDC.**

   ```
   system% /usr/openwin/bin/mp -u/usr/lib/lp/locale/zh/mp/mp.conf.U\
    udc.txt > udc.ps
   ```

Changing Font Directories

The `Xsession` script in the `/usr/dt/bin` directory includes the following font path: `/usr/openwin/lib/locale/`*locale*`/X11/fonts`, where *locale* is either `zh`, `zh.GBK`, or `zh_CN.GB18030`. To dynamically add a different font directory path, type the following commands.

```
system% xset +fp font_directory-path
system% xset fp rehash
```

▼ How to Add a Font Directory Path

● **Type the following commands to dynamically add a different font directory path.**

```
system% xset +fp font_directory-path
system% xset fp rehash
```

Bitmap and TrueType Fonts

The tables in this section list the TrueType and bitmap fonts that are available for the `zh_CN.EUC`, the `zh_CN.GBK`, and the `zh_CN.GB18030` locales.

The following table shows the TrueType fonts for the `zh`locale.

TABLE 9–3 TrueType Fonts for the `zh_CN.EUC` Locale

Full Family Name	Subfamily	Format	Vendor	Encoding
Fangsong	R	TrueType	Hanyi	GB2312.1980
Hei	R	TrueType	Monotype	GB2312.1980
Kai	R	TrueType	Monotype	GB2312.1980
Song	R	TrueType	Monotype	GB2312.1980

The following table shows the bitmap fonts for the `zh` locale.

TABLE 9–4 Bitmap Fonts for the `zh_CN.EUC` Locale

Full Family Name	Subfamily	Format	Encoding
Song	B	PCF (14,16)	GB2312.1980
Song	R	PCF (12,14,16,20,24)	GB2312.1980

TABLE 9–5 TrueType Fonts for the `zh_CN.GBK` Locale

Full Family Name	Subfamily	Format	Vendor	Encoding
Fangsong	R	TrueType	Zhongyi	GBK
Hei	R	TrueType	Zhongyi	GBK
Kai	R	TrueType	Zhongyi	GBK
Song	R	TrueType	Zhongyi	GBK

The following table shows the bitmap fonts for the `zh_CN.GBK` locale.

TABLE 9–6 Bitmap Fonts for the `zh_CN.GBK` Locale

Full Family Name	Subfamily	Format	Encoding
Song	R	PCF (12,14,16,20,24)	GBK

The following table shows the TrueType fonts for the `zh_CN.GB18030` locale.

TABLE 9–7 TrueType Fonts for the `zh_CN.GB18030` Locale

Family Name	Subfamily	Format	Vendor	Encoding
FangSong	R	TrueType	FangZheng	GB18030–2000
Song	R	TrueType	FangZheng	GB18030–2000
Hei	R	TrueType	FangZheng	GB18030–2000
Kai	R	TrueType	FangZheng	GB18030–2000

The following table shows bitmap fonts for the `zh_CN.GB18030` locale.

TABLE 9–8 Bitmap Fonts for the `zh_CN.GB18030` Locale

Family Name	Subfamily	Format	Encoding
Song	R	PCF (12,14,16,20,24)	GB18030–2000

Simplified Chinese Printing Facilities

This chapter describes how to print Simplified Chinese text to a line printer. The chapter also provides information on how to use the mp utilities to print to a PostScript printer or to a line printer.

Printing Chinese Output From a Command Line

This section provides information that you can use to print from a command line with the lp command and the mp program.

Printing With a Line Printer

You can use the commands in this section to print files of different formats to a line printer.

● **Use this command to print an EUC file.**

system% **lp** *EUC_filename*

● **Use this command to print a GB format file to a printer that supports GB format.**

system% **lp -T GB** *GB_filename*

Printing With the mp Utility

The enhanced mp print utility is available in the current Simplified Chinese Solaris release. You can use mp to print zh files in zh locale or UTF-8 files in zh.UTF-8 locales.

The output from the utility is standard PostScript that you can send to any PostScript printer.

Use the following command to print a file with the mp utility.

```
system% mp filename | lp
```

You can also use the utility as a filter because mp accepts stdin stream.

```
system% cat filename | mp | lp
```

You can set the utility as a print filter for a line printer. For example, the following command sequence signals the print service that the printer lp1 accepts only mp format files. The command also installs the printer lp1 on port /dev/ttya. See the lpadmin(1M) man page for more details.

```
system% lpadmin -p lp1 -v /dev/ttya -I MP
system% accept lp1
system% enable lp1
```

You can use the following command to add a filter for an LP print service.

```
system% lpfilter -f filtername \
-F pathname
```

The filter converts the default input type to PostScript output with the /usr/bin/mp command.

To print a UTF-8 text file, use a command line such as the following.

```
system% lp -T MP UTF-8-file
```

For more details on the mp(1), see the man page.

Using mp as an Xprt (X Print Server) Client

Using mp as an Xprt client enables mp to print the output of any printer connected to a network supported by an Xprt print service. As an Xprt client, mp supports PostScript and many versions of PCL.

The Xprt client attempts a connection to an Xprt server based on the following rules:

- When the -D *printer_name@machine[:dispNum]* or -P *printer_name@machine[:dispNum]* options are used with the mp command, mp attempts to connect to an Xprt print service on *machine[:dispNum]* with *printer_name*.

  If the above attempted connection to *machine[:dispNum]* fails or if the argument given to -D or -P is just *printer_name*, then the mp command checks the XPSERVERLIST for Xprt servers that support the *printer_name* argument. For example:

  ```
  system% setenv XPSERVERLIST "machine1[:dispNum1] machine2[:dispNum2] \
  machine3[:dispNum3]"
  ```

- If no server is found using above rules, mp checks for an XPDISPLAY environment variable set to *machine[:dispNum]*. For example:

  ```
  system% setenv XPDISPLAY "machine[:dispNum]"
  ```

- If the XPDISPLAY variable is not set or if the variable is invalid, mp tries to connect to the default display :2100. If the default display value is also invalid, mp exits with an error message.

  The /usr/lib/lp/locale/C/mp directory contains .xpr print page sample layout files for Xprt client. The sample files are for 300 dpi printers. If the target printer has a different dpi value, the dpi value of the sample files is automatically converted to the resolution of the target printer.

TTY Environment and Support

This chapter assumes that you are familiar with the following information.

- The Solaris Operating System communicates with external devices using STREAMS and `ioctl`.

- Different terminal types are supported by `termcap` and `terminfo`.

This chapter includes the following sections:

Refer to the `termio`(7I) man page for background information on STREAMS and TTY drivers.

Terminal Support

The Simplified Chinese Solaris Operating System supports Chinese terminals (EUC-GB). The terminals should have built-in Simplified Chinese fonts and input methods.

Installing a Terminal

If you have not added a terminal to your system before, install a terminal in ASCII mode only.

▼ How to Install a Terminal

You can use the serial ports option from the Admintool menu to configure serial ports for terminals. Serial ports provide the easiest means of installing a terminal.

1. Become superuser.

```
system% su
```

2. Type admintool.
The admintool menu appears.

```
system# admintool
```

3. Select the serial ports icon.

4. Select edit on the serial ports menu.
The modify service submenu appears.

5. On the modify service submenu, select enabled, baud rate 9600. Then enter the terminal type.

细节: ● 基本 ○较多 ○深入

波特率: 9600

终端类型: vt200

重置   取消

▼ How to Set Up a ttya Port from the Command Line

Use the following procedure is required tto set up a terminal on ttya port from the command line.

1. Determine the port monitor version number.

```
# ttyadm -V
```

The port monitor version number will display.

2. **Enter the following commands and substitute the port monitor version number for** *ver*.

   ```
   # pmadm -r -p zsmon -s ttya
   # sacadm -a -p zsmon -t ttymon -c /usr/lib/saf/ttymon -v ver
   ```

   For more information, see pmadm(1M) and sacadm(1M) man pages.

3. **Use the `pmadm` command that matches your terminal type to add a login service.**

   For EUC terminals, use the following command.

   ```
   # pmadm -a -p zsmon -s ttya -i root -fu -v ver -m "`ttyadm -S y \
     -T terminal_type -d /dev/ttya -l 9600 -m ldterm,ttcompat -s \
     /usr/bin/login`"
   ```

4. **Turn on the terminal.**

   Refer to the documentation that accompanies the terminal.

5. **Log in to the terminal.**

6. **Check to see that the installation is correct.**

   ```
   # setenv LANG locale
   # /bin/stty cs8 -istrip defeucw
   ```

Note – These values show that the operating system is set to communicate with the terminal in 8-bit no-parity mode. Make sure the terminal is set up in 8-bit no-parity mode. Refer to the terminal's setup manual for the proper way to set terminal options.

▼ How to Verify TTY Set Up

Do the Use the following procedure to verify that your TTY is properly set up .

1. **Type the `/bin/stty` command with the `-a` option.**

   ```
   system% /bin/stty -a
   ```

2. **If the `cs8`, `-istrip` values set in the previous procedure are not listed, use the following command to set them.**

   ```
   system% /bin/stty cs8 -istrip defeucw
   ```

This is the last step in setting up a terminal. The default setting of a Simplified Chinese terminal is the completion-code mode. Continue to the next section to install a Packed-code TTY.

Backward Compatibility

This chapter contains information for making programs backward-compatible with earlier versions of Asian Solaris software. Every utility described is supported. For this version of Solaris, you are encouraged to use the XPG4 internationalization APIs described in *International Language Environments Guide*.

Simplified Chinese Test Utilities

These utilities test various aspects of the Simplified Chinese (GB-2312-80) national standard character set. They also assume that the character being tested is part of the national standard character set.

The arguments for the functions in these tables must be a character in WC, `wchar_t`. For more information, see the `cctype(3x)` man page.

TABLE **12–1** Simplified Chinese Character Classification Functions

Routine	Description
ischanzi	Returns true if it is a Hanzi ideogram in GB-2312-80.
iscaccent	Returns true if it is an accent notation in GB-2312-80.
iscphonetic	Returns true if it is a phonetic symbol in GB-2312-80.
iscpinyin	Returns true if it is a Pinyin symbol in GB-2312-80.
iscalpha	Returns true if it is a Roman alphabetic in GB-2312-80.
iscdigit	Returns true if it is a Roman digit in GB-2312-80.
iscnumber	Returns true if it is a number in GB-2312-80.

Routine	Description
isclower	Returns true if it is a Roman lowercase in GB-2312-80.
iscupper	Returns true if it is a Roman uppercase in GB-2312-80.
iscblank	Returns true if it is a white space character from GB-2312-80.
iscspace	Returns true if it is a space character from GB-2312-80.
iscgen	Returns true if it is a graphic or general symbol in GB-2312-80.
iscsci	Returns true if it is a scientific symbol in GB-2312-80.
iscline	Returns true if it is a ruled line symbol in GB-2312-80.
iscunit	Returns true if it is a unit character in GB-2312-80.
iscparen	Returns true if it is a right or left parenthesis in GB-2312-80.
iscpunct	Returns true if it is a punctuation character in GB-2312-80.
iscgreek	Returns true if it is a Greek character in GB-2312-80.
iscrussian	Returns true if it is a Russian character in GB-2312-80.
iscspecial	Returns true if it is a Greek or Russian character in GB-2312-80.
ischira	Returns true if it is a Japanese Hiragana character in GB-2312-80.
isckata	Returns true if it is a Japanese Katakana character in GB-2312-80.

Two additional routines for Simplified Chinese, iscgb and isceuc, test for characters from the GB-2312-80 character set. The iscgb routine expects a wide character, and isceuc expects a GB-2312-80 character in EUC format. For more information, see the cctype(3x) man page.

TABLE 12–2 General Simplified Chinese General Character Classification Functions

Routine	Description
iscgb	Returns true if it is in GB-2312-80.
isceuc	Returns true if it is a GB-2312-80 character in EUC format.

Simplified Chinese Conversion Utilities

This section describes functions for wide character and string input and output, character classification, and conversion functions for the Simplified Chinese character sets. Solaris 2.7 software implements a wide character library for handling Simplified Chinese character codes according to industry standards.

Routines that have Chinese language-specific dependency are in their own language-specific library, which is linked with the corresponding C compiler option. Simplified Chinese Solaris `libcle` is linked with `-lcle`

Refer to the appropriate `man` pages for more information.

Asian Solaris software defines WC as a constant-width, four-byte code. WC uses the ANSI C data type `wchar_t`, which Solaris software defines in `wchar.h` as follows:

```
typedef long wchar_h;
```

In Solaris software, `long` is four bytes.

Conversion Utilities

The conversion functions described in this section are available, but you should use `iconv()` as a standard function.

Simplified Chinese Solaris software provides facilities for various conversions, for example:

- Characters within a code set, such as converting uppercase ASCII to lowercase.
- Between different conventions for national standard character sets, such as GB and EUC.
- Between code formats (such as converting between EUC and WC).

Programs using the general multibyte conversion utilities should include the header files `widec.h` and `wctype.h`. Simplified Chinese Solaris specific routines (such as isc*xxx*) are declared in `zh/xctype.h`.

Programs using general multibyte conversion utilities should include three header files: `wctype.h`, `widec.h`, and `zh/xctype.h`.

The *locale*/`xctype.h` file declares the Chinese locale-specific routines, which have names of the form isc*xxxx*:

As with the classification functions described in the previous section, the use of these functions can be controlled by the `setlocale` function (described elsewhere in this and other chapters).

Locale-specific conversion routines (such as Chinese `cgbtoeuc`) are contained in the `libcle` library:

This library can be linked during compilation using the C compiler option `-lcle`.

Conversion Within a Code Set

The multibyte conversion functions are similar to the one-byte conversion functions `toupper` and `tolower`. These functions convert wide-characters to other wide characters. For more information on conversion routines, see the man pages for `wconv(3)` and `cconv(3)`.

The following routines are in the regular Chinese C library.

TABLE 12–3 Simplified Chinese Case Conversion Functions (declared in `zh/xctype.h`)

Function	Description
`tocupper`	Converts code set1 Roman lowercase to uppercase
`toclower`	Converts code set1 Roman uppercase to lowercase

Conversion Between Simplified Chinese Code Sets

In the Simplified Chinese character sets, the Roman characters and numbers in code set 0 are repeated in code set 1. The following functions test wide characters.

TABLE 12–4 Simplified Chinese Code Set Conversion Functions

Function	Description
`atocgb`	Converts alphabetic or numeric characters in ASCII (code set0) to the corresponding characters in GB-2312-80 (code set1).
`cgbtoa`	Converts alphabetic or numeric characters in GB-2312-80 (code set1) to the corresponding characters in ASCII (code set0).

For further information on these functions, see the man page for `cconv()(3x)`.

Conversion for Simplified Chinese Character Codes

The following routines do character-based code conversion on the GB-2312-80 character set. They convert characters and strings between EUC format and GB-2312-80 format. To use these routines, the library `libcle` must be linked using the C compiler option `-lcle`. For further information, see the `cconv(3)` man page.

TABLE 12–5 Simplified Chinese Character-Based Functions

Function	Description
cgbtoeuc	Converts a character in GB-2312-80 format (7 bit) to EUC format
scgbtoeuc	Converts a string in GB-2312-80 format (7 bit) to EUC format
sncgbtoeuc	Converts part of a string in GB-2312-80 format (7 bit) to EUC format
euctocgb	Converts a character in EUC format to GB-2312-80 format (7 bit)
seuctocgb	Converts a string in EUC format in GB-2312-80 format (7 bit)
sneuctocgb	Converts a part of a string in EUC to GB-2312-80 format (7 bit)

Binary Compatibility Package (BCP)

Applications compiled under Chinese OpenWindows 2.*x* or Solaris 1.*x* or SunOS 4.*x* systems have different binary formats than the current Chinese Solaris release. Older applications can nevertheless be run under the current Chinese release without being recompiled by using its included binary compatibility package (BCP).

Note – SUNWowbcp must be included in your system configuration in order for you to run the following commands. See your system administrator for installation.

Running Earlier Binary Code

The following BCP command runs the compiled binary code of earlier SunOS4.*x*, Solaris 1.*x*, or Chinese OpenWindows 2.*x* applications without recompilation, However, OpenWindows V2 Chinese applications will display no input server status region. As shown in the following examples, the command calls the application by its old name (*old_application_name*) and sets the basic locale, input language, and display language using the older version's specific locale name (*old-locale*):

```
system% old_application_name -lc_basiclocale
old-locale -lc_inputlang old-locale
\ -lc_displaylang old-locale
```

The following example shows the command for running the compiled binary code of an earlier version of the textedit application in the current Simplified Chinese Solaris environment:

```
system% textedit -lc_displaylang chinese -lc_basiclocale chinese \
    -lc_inputlang chinese
```

Due to incompatibilities between Simplified Chinese Solaris 2.*x* and 1.*x* applications, you cannot cut and paste Chinese characters between them.

Glossary

ANSI   American National Standards Institute. ANSI proposes standard definitions for different computing languages. The most recent standard for the C language, prepared by the ANSI C X3J11 Committee, includes library functions for computing with multibyte characters for international usage, as well as a new data type, `wchar_t`, for dealing with four-byte characters. This standard is not completed, so it is referred to as the "proposed ANSI C standard," or ANSI C-X3J11.

ASCII   American Standard Code for Information Interchange. ASCII is a seven-bit code containing English upper and lowercase letters, punctuation, numbers and control codes. The eighth bit in each byte is used by different applications for parity checking, communication and message passing protocols, compacting data, or other purposes. Applications that are intended to be internationalized cannot use this bit and use multiple code sets or multibyte characters, and utilities that handle multiple code sets or multibyte characters.

category   In the Traditional Chinese Solaris documentation set, category is related to localization. A category is a portion of a country's language representation and cultural conventions. For instance, the date is often represented in the U.S. as *month*, *day*, *year*; while in another country it might be *day*, *month*, *year*. The date and time can be thought of as one category of a local language. Categories also refer to the program categories, the environment variables that are related to categories, and the ANSI localization tables for each category.

character set   A set of elements used for the organization, control, or representation of data. Character sets may be composed of alphabets, ideograms, or other units. Character sets may contain other character sets, which causes unclear boundaries. For example, the CNS 11643 character set contains English, Greek, and Chinese character sets in addition to Chinese radicals and many other characters.

CNS	Taiwan's Chinese National Standard. This standard is the Taiwan analogue to ASCII. In this document set, CNS refers to the code set defined by CNS 11643. It contains the Chinese characters, phonetic symbols and radicals, control codes, punctuation, and western alphabets, including Roman and Greek characters. Each character is two bytes long, with the highest or most significant bit of each byte set to zero. In other words, CNS uses the lower seven bits of each byte. Due to the size of the Taiwan Chinese character set, the character sets are divided into multiple codeplanes, with the default plane containing the most commonly used characters. ISO 2022 provides mechanisms for shifting from one codeplane to another.
	After its revision in 1992, CNS 11643 defines 48,000 characters, which are divided among codeplanes 1-7, codeplanes 8-16 are undefined, but are included in the code set architecture. Codeplanes 1 and 2 (common and rarely used characters) are unaffected by the revision. Characters that were in codeplane 14, a provisional user-defined plane, have been standardized into codeplane 3, with the overflow in codeplane 4.
code set	A set of unambiguous rules that establishes a character set and the one-to-one relationship between each character in the character set and its bit representation. For example, the English character set, including punctuation and numbers, can be mapped to the ASCII code set in such a way that each character corresponds to only one bit code, and no bit code corresponds to more than one character. A code set is also called a coded character set.
commit	Characters entered in the preedit area that are put in the text block which is assembled for the application.
EUC	Extended UNIX Code. Describes four code sets modelled on ISO-2022. Each code set can contain one or more different character sets, like the Hangul and Hanja character sets in KS C 5601. The four code sets are referred to as code sets 0, 1, 2, and 3. In this guide, these code sets are sometimes abbreviated as cs0, cs1, cs2, and cs3. Other internationalization efforts sometimes call these code sets g0, g1, g2, and g3. code set 0 is also called the primary code set, and code sets 1, 2, and 3 are called the supplementary code sets. In the Korean and Chinese implementations of the EUC codes, the primary code set (cs0) contains ASCII and begins with a zero in the most significant bit.
EUC-CNS	The EUC representation of CNS 11643. For code set 1, this standard is the normal CNS code with a one in the most significant bit of each byte. In other words, EUC-CNS equals CNS plus 0x8080. For example, the CNS character 0x212A becomes the EUC-CNS character 0xA1AA. Or in binary, 00100001 00101010 becomes 10100001 10101010. For code sets 2 and 3, characters are also prefixed by single shift bytes SS2 and SS3. In addition, code set 2 requires a codeplane byte. The code of a

code set-2 character is SS2 followed by codeplane byte followed by EUC-CNS. The codeplane byte is plane number added to 0xA0. For example, plane 2 has codeplane byte 0xA2.

ISO   International Standards Organization. Composed of a number of professional societies and companies, this organization studies and makes recommendations on internationalization issues. ISO 2022 proposes and describes the Extended UNIX Codes. Other ISO proposals include the European 8-bit code and communication protocols for internationalization.

locale   A locale describes a language or cultural environment. Its setting affects the display or manipulation of language-dependent features. Traditional Chinese Solaris software provides C for U.S.A, zh_TW for Traditional Chinese extended UNIX code, and zh_TW.BIG5 for the Traditional Chinese Big5 locale.

POSIX   Portable Operating System for Computer Environments. An IEEE standards group comprising seven committees that create documents for standardizing and internationalizing UNIX. POSIX document 1003.1 deals with the kernel and system calls. 1003.2 concerns the C-shell and standard libraries. The other five deal with real-time computing, communications and networking, and other issues.

Unicode   The international character set and encoding developed by the Unicode Consortium.

Wide character code (WC)   A constant-width four-byte code, called WC in Asian Solaris documentation, for the internal representation of EUC codes using the new ANSI-C data type wchar_t. Although EUC does not specify limits on the size of the supplementary code sets (code set 0 is always one byte), WC specifies a character as four bytes. Standardizing on four bytes takes up more memory space than necessary if the environment is primarily ASCII, but this practice also speeds processing time for strings of mixed characters. The 1000th character always begins at byte 4000 (and the 0th character starts at byte 0). This practice is useful for any type of indexing in applications.

X/Open   X/Open started as a consortium of international UNIX vendors from Europe, USA, and Asia. It is now one of the major standards organizations like POSIX and ANSI; source of *X/Open System Interface Portability Guide*.

Index