Tru64 UNIX

Technical Reference for Using Chinese Features

September 2002

This manual provides information on the use of the Chinese-specific features supported on the Tru64 UNIX operating system.

Product Version: Tru64 UNIX Version 5.1B or higher

Hewlett-Packard Company
Palo Alto California

Table of Contents

About This Manual

9 Other Chinese Features 9–1

Figures

Tables

About This Manual

This manual provides information for end users and programmers who want to use and develop internationalized applications in Chinese locales on the HP Tru64 UNIX Operating System.

Audience

This manual is for both new and experienced end users and programmers who are interested in the Chinese variant of the Tru64 UNIX operating system.

Organization

This manual is organized as follows:

Chapter 1	Describes the Chinese character sets supported on the Tru64 UNIX operating system
Chapter 2	Describes the Chinese codesets and the conversion among different codesets
Chapter 3	Describes the Chinese locales
Chapter 4	Describes the local hardware devices that support the Chinese locales
Chapter 5	Provides information on Chinese fonts
Chapter 6	Provides information on Chinese keyboards
Chapter 7	Describes how to input Chinese characters
Chapter 8	Introduces Chinese printing support
Chapter 9	Provides descriptions of other Chinese features

Related Documents

The following manuals supplement information in this manual:

Tru64 UNIX *Writing Software for the International Market* and Tru64 UNIX *Using International Software* describe how to create internationalized programs on Tru64 UNIX and how to use internationalization features of the operating system. These manuals are part of the Tru64 UNIX documentation set and are available on the World Wide Web at the following URL: http://www.tru64unix.compaq.com/docs/.

Programming for the World: A Guide to Internationalization (O'Donnell, Sandra Martin, 1994, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, USA) provides information about cultural and linguistic requirements around the world and the changes needed in computer systems to handle those requirements.

X Window System, Third Edition, Robert W. Scheifler and James Gettys, 1992, Digital Press, One Burlington Woods Drive, Burlington, MA, USA, 01803

Programmer's Supplement for Release 5 of the X Window System, Version 11, David Flanagan, O'Reilly & Associates, Inc., 103 Morris Street, Sebastopol, CA, USA, 95472

The Unicode Standard, Version 3.0 and Version 3.1, The Unicode Consortium, http://www.unicode.org/

Information Technology-Universal Multiple-Octet Coded Character Set, ISO/IEC 10646:2001. The basic multilingual plane defined by this standard is identical with the main body of Unicode character encoding.

Codes of Chinese Graphic Characters for Information Interchange, Primary Set (GB2312-80); National Standards Bureau of China, Beijing, 1980

Chinese Character Input Standard, GB18030-2000; National Standards Bureau of China, Beijing, 2001

Conventions

The following typographical conventions are used in this manual:

%	A percent sign represents the C shell system prompt.
$	A dollar sign represents the system prompt for the Bourne and Korn shells.
#	A number sign represents the superuser prompt.
% **cat**	Boldface type in interactive examples indicates typed user input.
File	Italic type indicates variable values, placeholders, and function argument names.
[|] {|}	In syntax definitions, brackets indicate items that are optional and braces indicate items that are required. Vertical bars separating items inside brackets or braces indicate that you choose one item from among those listed.
...	In syntax definitions, a horizontal ellipsis indicates that the preceding item can be repeated one or more times.
cat(1)	A cross-reference to a reference page includes the appropriate section number in parentheses. For example, cat(1) indicates that you can find information on the cat command in Section 1 of the reference pages.
[RETURN]	In an example, a key name enclosed in a box indicates that you press that key.
Ctrl/*x*	This symbol indicates that you hold down the first named key while pressing the key or mouse button that follows the slash. In examples, this key combination is enclosed in a box (for example [Ctrl/C]).

1

Character Sets

The Tru64 UNIX operating system supports the following Chinese character sets:

- CNS 11643
- DTSCS
- Big-5
- GB2312-80
- Extended GB
- GBK and GB18030-2000
- Unicode
- ISO/IEC 10646

The CNS 11643 and Big-5 character sets are commonly used for traditional Chinese characters. Also, the traditional Chinese Input Server optionally supports DTSCS input. The GB2312-80, GBK, Extended GB, and GB18030-2000 character sets are commonly used for simplified Chinese characters. The Unicode and ISO/IEC 10646 character sets are common to both traditional and simplified Chinese.

1.1 CNS 11643

The CNS (Chinese National Standard) 11643 character set standard was published by the National Bureau of Standards of Taiwan in 1986 and was updated in 1992. It is also called "Standard Interchange Code for Generally-used Chinese Character" (SICGCC).

CNS 11643 provides 16 character planes for defining Chinese characters. Each character plane is divided into 94 rows and each row has 94 columns. Altogether, a total number of 8,836 characters can be accommodated in each plane. Character planes 1-11 are reserved for defining standard Chinese characters while character planes 12-16 are user-defined areas.

Figure 1–1: CNS 11643 Character Planes
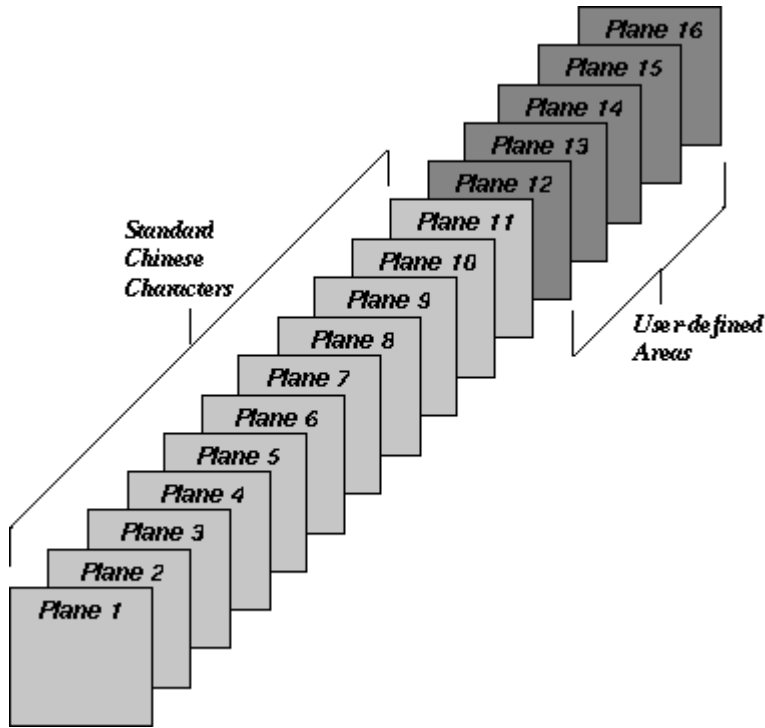
The original CNS 11643 standard, published in 1986, defines certain groups of characters only on the first and second character planes. Table 1–1 describes these groups of characters.

Table 1–1: Characters Defined in CNS 11643-1986

Character Plane	Character Type	Number of Characters
Plane 1	Special characters	651
	Control characters	33
	Frequently used characters	5,401
Plane 2	Less frequently used characters	7,650

Figure 1–2 and Figure 1–3 illustrate the positions of these characters in the first and second character planes.

Figure 1–2: CNS 11643 First Character Planes
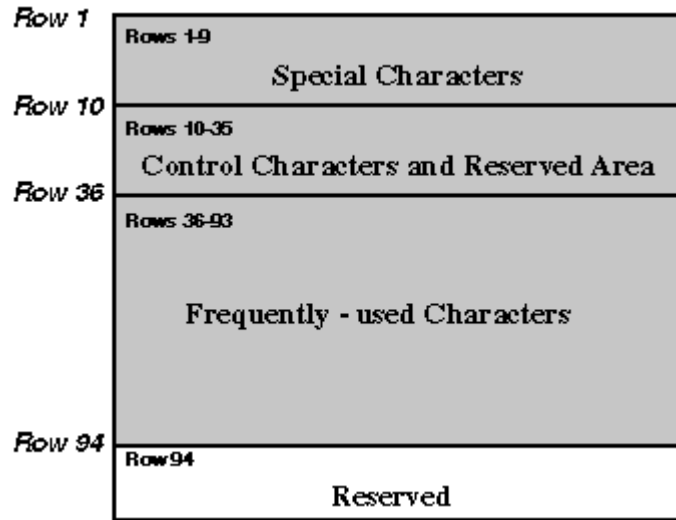
Figure 1–3: CNS 11643 Second Character Plane
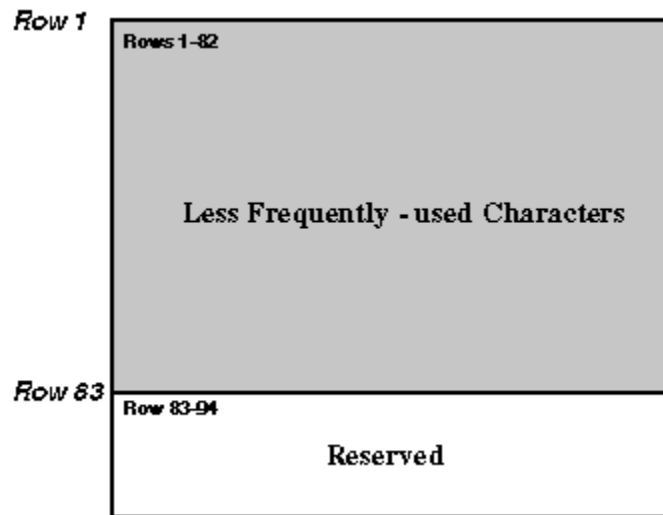
Because the CNS11643-1986 character set was not rich enough to fully meet application requirements, such as names and addresses, the information industry in Taiwan requested an expansion of the character set. In 1991, the Bureau of National Standard formed a team to study how to expand CNS 11643. On August 4, 1992, the Bureau of National Standard published the revised CNS 11643 - Chinese Standard Interchange Code (CSIC).

The revised CNS 11643, called CNS 11643-1992, defined 651 special characters, 33 control characters and 48,027 Chinese characters, as shown in Table 1–2.

Table 1–2: Characters Defined in CNS 11643-1992

Character Plane	Character Type	Number of Characters
Plane 1	Special characters Control characters Frequently used characters	651 33 5,401
Plane 2	Less frequently used characters	7,650
Plane 3	Rarely used characters (EDPC Part I)	6,148
Plane 4	Used for residency system, ISO 2nd edition DIS 10646 Han characters, 171 EDPC Part II Characters	7,298
Plane 5	Rarely used characters	8,603
Plane 6	Variants based on the Ministry of Education publications (less than, or equal to, 14 strokes)	6,388
Plane 7	Variants based on the Ministry of Education publications (greater than 14 strokes)	6,539

Planes 5, 6, and 7 are based on Taiwanese Ministry of Education publications.

Because the number of characters defined in CNS 11643-1992 is far greater than those required for general use, the revised CNS 11643 is called "Chinese Standard Interchange Code (CSIC)".

_____ **Note** _____

In this release, the new characters added to CNS 11643-1992 are not supported. Only the characters defined in CNS 11643-1986 and DTSCS (which will be described in the next section) are supported.

_____

1.2 DTSCS

In addition to CNS 11643, the operating system supports the DIGITAL Taiwan Supplemental Character Set (DTSCS). Currently, only the EDPC Recommended Character Set, which defines a total of 6,319 characters, is included in DTSCS. EDPC Recommended Character Set was first published by the Electronic Data Processing Center of Executive Yuen in June, 1988.

Figure 1–4: EDPC Recommended Character Set
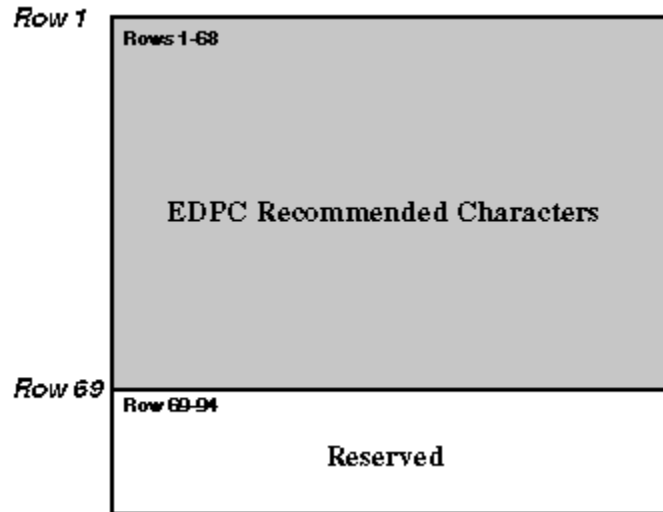
As a de facto standard, computer vendors support the EDPC Recommended Character Set and assign it to CNS 11643 character plane 14.

In the revised CNS 11643-1992, the 6,319 characters in the EDPC Recommended Character Set are assigned to the third and fourth character planes of CNS 11643, as described in Table 1–3 and shown in Figure 1–4.

Table 1–3: Mapping of EDPC Recommended Character Set to CNS 11643-1992

EDPC Characters	Character Plane	Number of Characters
Part I	Plane 3	6,148
Part II	Plane 4	171

1.3 Big–5

The Big-5 character set, though not a national standard, is commonly used by the Taiwan information industry, particularly in the PC and workstation market. The Big-5 character set was designed to meet the requirements of five major software vendors in Taiwan. Since its publication, much software and hardware, and many peripheral devices have been developed to support Big-5.

Big-5 is very similar to the first two planes of CNS 11643-1992. The frequently used Chinese characters (5,401) defined in the two character sets are exactly the same except that their positions in the code table are different. For the less frequently used Chinese characters, Big-5 defines two more characters in addition to the 7,650 characters defined in the second character plane of CNS 11643, and their positions in the code table are also different.

1.4 GB2312–80

The GB2312-80 character set is a standard published by the State Bureau of Standardization of the People's Republic of China (PRC) in 1980 and put in force in May, 1981.

GB2312-80 defines 7,445 characters, including 6,763 Chinese characters in the following categories:

- Graphic symbols

  GB2312-80 defines 682 graphic symbols placed in rows 1-9.

- Level 1 characters

  GB2312-80 defines 3,755 frequently used characters placed in rows 16-55.

- Level 2 characters

  GB2312-80 defines 3,008 less frequently used characters placed in rows 56-87.

The GB2312-80 code table is divided into 94 rows (Qu), numbered from 1 to 94. Each row has 94 columns (Wei), also numbered from 1 to 94. Figure 1–5 illustrates the GB2312–80 character set.

Figure 1–5: GB2312-80 Character Set

1.5 Extended GB

The extended GB character set provides 8,836 (94 x 94) code points for defining user-defined characters. The 8,836 code points are divided into two regions:

- User-Defined Area — Spans rows 1-87 and provides 8,178 code positions.

- User-Defined (reserved) Area — Spans rows 88-94 and provides 658 code positions. This area is where users define special and frequently used user-defined characters.

The extended GB code table is similar to the GB2312 code table. It is divided into 94 rows and each row has 94 columns.

1.6 GBK and GB18030-2000

The GBK character set is an extension to the GB 2312-80 character set. GBK includes all of the simplified Chinese (Hanzi) characters specified by the ISO 10646 standard (also known as the GB13000:1.93 character set) that are not already included in GB 2312:80. GBK is therefore defined as a normative annex of GB13000.1-93.

The GB18030-2000 character set, defined by the Chinese National Standard organization, further extends GBK by means of 4-byte code points. That is, the GB18030-2000 character set has 4-byte encoding in addition to the 1-byte and 2-byte encoding of GBK.

GB18030-2000 incorporates GBK support for the Hanzi characters specified by Unicode Version 3.0 and the ISO/IEC 10646-2001 standard.

1.7 Unicode

The Unicode Standard, Version 3.0 specifies a universal character set (UCS) that contains definitions for 49,194 characters and includes a Private Use Area for vendor-defined or user-defined characters. The character set has the following features:

- All characters are treated as 16-bit units.

- Each 16-bit unit has an abstract character identity.

- Certain sequences of 16-bit characters in a text stream are transformed into other characters, called composed characters.

- All characters have properties, such as base, numeric, spacing, combination, and directionality. The Unicode standard provides rules for ordering characters with different properties so that parsing of character sequences is unambiguous.

- The relationship between Unicode characters and the glyphs in the native language script that users see, type, or print is not necessarily one-to-one. A glyph may be mapped to a single abstract character or a composed character. Conversely, more than one glyph can be mapped to a character.

- The ISO 8859-1 character set occupies the first 256 code positions (and the ASCII character set the first 128 positions) of the UCS.

1.8 ISO/IEC 10646

The ISO/IEC 10646 standard, which is specified in Information Technology-Universal Multiple-Octet Coded Character Set, ISO/IEC 10646, allows characters to be specified as either 32-bit units or as 16-bit units like Unicode. In their 32-bit form, the 16-bit character values in Unicode are zero-extended through a second 16-bit unit to conform to ISO/IEC 10646.

2

Codesets and Codeset Conversion

The operating system fully supports the following Chinese codesets by including locales and codeset conversion support:

- DEC Hanyu (Section 2.1)

- Taiwanese EUC, Extended UNIX Code (Section 2.2)

- Big-5 (Section 2.3)

- DEC Hanzi (Section 2.4)

- GBK (Section 2.5)

- GB18030 (Section 2.6)

- UTF-8 (Section 2.10)

The operating system also provides conversion support for the following codesets:

- Telecode (Section 2.8)

- Shift Big-5 (Section 2.7)

- UTF-16 (Section 2.9)

- UCS-4 (Section 2.9)

- HKSCS (Section 2.11.4)

2.1 DEC Hanyu

The DEC Hanyu codeset, denoted by `dechanyu`, consists of the following character sets:

- ASCII
- CNS11643, the first and second character planes
- DTSCS
- User-Defined Characters

DEC Hanyu uses a combination of single-byte, 2-byte, and 4-byte data to represent ASCII characters, symbols, or ideographic characters.

2.1.1 ASCII CODE

All ASCII characters can be represented in the form of single-byte 7-bit data in DEC Hanyu. That is, the Most Significant Bit (MSB) of ASCII characters is always set off.

2.1.2 CNS 11643 Code

Each CNS 11643 character is represented by a 2-byte code in DEC Hanyu, which complies with the CNS 11643 standard. The MSB of the first byte is always set on while that of the second byte can be on for the first character plane or off for the second character plane. See Figure 2–1.

Figure 2–1: DEC Hanyu Encoding of CNS 11643 Planes
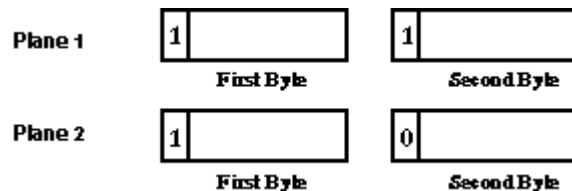
The first byte of a CNS 11643 code determines the row number of the character, while the second byte determines its column number. Table 2–1 illustrates the code range of a CNS 11643 code.

Table 2–1: CNS 11643 Code Range in DEC Hanyu

Character Plane	1st Byte (hexadecimal)	2nd Byte (hexadecimal)
Plane 1	A1 to FE	A1 to FE
Plane 2	A1 to FE	21 to 7E

The following formulas illustrate the code of a CNS 11643 character in relation to its row and column numbers.

CNS 11643 Plane 1 character:

First byte = A0 + row number
Second byte = A0 + column number

CNS 11643 Plane 2 character:

First byte = A0 + row number
Second byte = 20 + column number

For example, if a character is positioned at the first column of the 36th row on CNS 11643 Plane 1, its encoding value is calculated as follows:

First byte = A0 (hex) + 36 = C4 (hex)
Second byte = A0 (hex) + 01 = A1 (hex)

Its encoded value is C4A1.

Similarly, if a character is positioned at the first column of the 36th row on CNS 11643 Plane 2, its encoding value is calculated as follows:

First byte = A0 (hex) + 36 = C4 (hex)
Second byte = 20 (hex) + 01 = 21 (hex)

Its encoded value is C421.

Figure 2–2 illustrates the division of a 2-byte code space and the position of CNS 11643 characters.

Figure 2–2: Code Space for CNS 11643 in DEC Hanyu
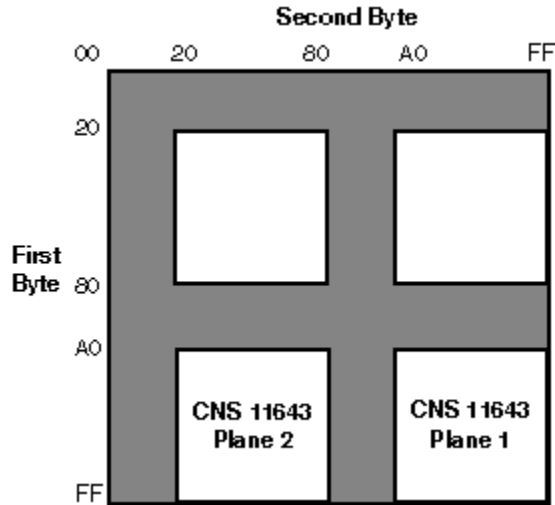
2.1.3 DTSCS Code

Each DTSCS character is represented by a 4-byte code in DEC Hanyu. The first two bytes are the leading codes, namely 0xC2 and 0xCB, which are used as a designator sequence for the DTSCS character set. The MSB of the third and fourth bytes is set on for the EDPC Recommended Character Set. See figure 2–4.

Figure 2–3: DEC Hanyu Encoding of DTSCS Characters
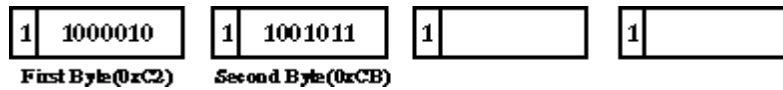
Figure 2–4 illustrates the 4-byte code space and the position of DTSCS characters.

Figure 2–4: Code Space for DTSCS in DEC Hanyu
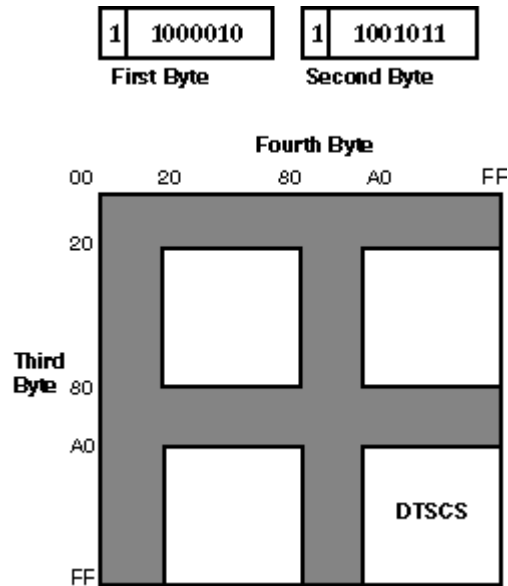
2.1.4 User Defined Characters

In addition to the CNS11643 and the DTSCS character sets described above, DEC Hanyu provides 3,587 positions for User Defined Characters (UDC). The positions for UDCs are those unused (but not reserved) code points on the CNS 11643 first and second character planes. Therefore, the encoding of UDC is exactly the same as that of CNS 11643 except that they occupy different regions, as shown in Table 2–2.

Table 2–2: UDC Code Range in DEC Hanyu

Character Plane	Number of UDC	Code Range
Plane 1	145	FDCC - FEFE
Plane 1	2,256	AAA1 - C1FE
Plane 2	1,186	F245 - FE7E

2.2 Taiwanese EUC

Taiwanese EUC (Extended UNIX Code), denoted as `eucTW`, is another codeset to support CNS 11643. The design of Taiwanese EUC allows the 16 character planes of CNS 11643 to be encoded in a unified way. A stream of data encoded in Taiwanese EUC can contain characters defined in ASCII and the 16 character planes. Figure 2–5 illustrates the encoding of Taiwanese EUC.

Figure 2–5: Encoding of Taiwanese EUC

[0-7E]				ASCII
[A1-FE]	[A1-FE]			CNS 11643 Plane 1
SS2	A2	[A1-FE]	[A1-FE]	CNS 11643 Plane 2
SS2	A3	[A1-FE]	[A1-FE]	CNS 11643 Plane 3
SS2	A4	[A1-FE]	[A1-FE]	CNS 11643 Plane 4
SS2	A5	[A1-FE]	[A1-FE]	CNS 11643 Plane 5
SS2	A6	[A1-FE]	[A1-FE]	CNS 11643 Plane 6
SS2	A7	[A1-FE]	[A1-FE]	CNS 11643 Plane 7
SS2	A8	[A1-FE]	[A1-FE]	CNS 11643 Plane 8
SS2	A9	[A1-FE]	[A1-FE]	CNS 11643 Plane 9
SS2	AA	[A1-FE]	[A1-FE]	CNS 11643 Plane 10
SS2	AB	[A1-FE]	[A1-FE]	CNS 11643 Plane 11
SS2	AC	[A1-FE]	[A1-FE]	CNS 11643 Plane 12
SS2	AD	[A1-FE]	[A1-FE]	CNS 11643 Plane 13
SS2	AE	[A1-FE]	[A1-FE]	CNS 11643 Plane 14
SS2	AF	[A1-FE]	[A1-FE]	CNS 11643 Plane 15
SS2	B0	[A1-FE]	[A1-FE]	CNS 11643 Plane 16

Taiwanese EUC uses the Single-Shift 2 control character (SS2) and an additional byte to specify a character plane. The only exception is the first plane, which does not require leading codes. Instead, two bytes specify a character's position on the first plane. The first byte determines its row number, while the second determines its column number. The MSBs of the two bytes are set on.

In this release, only the characters defined in the first and second planes of CNS 11643 and those in the EDPC Recommended Character Set that have been remapped into the third and fourth character planes of the revised CNS 11643-1992 are supported in Taiwanese EUC. Other characters that were added to the CNS 11643-1992 standard are not supported.

2.3 Big–5

The Big-5 codeset, denoted as `big5`, is the only codeset that supports the Big-5 character set. The encoding of the Big-5 codeset is similar to that of CNS 11643 in DEC Hanyu. Each Big-5 character is represented by a 2-byte code, which complies with the Big-5 standard. The MSB of the first byte is always set on while that of the second byte can be set on or off.

The Big-5 code range is defined as shown in Table 2–3.

Table 2–3: Big-5 Code Range

Character	Number of Characters	Code Range
Special symbols	408	A140-A3BF
Level 1 characters	5,401	A440-C67E
Level 2 characters	7,652	C940-F9D5

The operating system supports codeset conversion for HKSCS (the Hong Kong Supplementary Character Set) and uses Big-5 encoding for HKSCS representation. HKSCS characters map to Big-5 in the range of 8840 to FEFE. See Section 2.11.4 for more information on HKSCS codeset conversion.

In addition to the code points for special symbols and Chinese characters shown in Table 2–3, three areas are set aside for user defined spaces. Some vendors in Taiwan support user defined characters in the code ranges shown in Table 2–4.

Table 2–4: Big-5 User Defined Spaces

Character	Number of Characters	Code Range
Level 1 user defined space	785	FA40-FEFE
Level 2 user defined space	2,983	8E40-A0FE
Level 3 user defined space	2,041	8140-8DFE

The valid ranges of the two bytes are:

Byte	Valid Ranges
First byte	81-FE
Second byte	40-7E and A1-FE

Figure 2–6 illustrates the encoding of the Big-5 codeset in a 2-byte code space.

Figure 2–6: Code Space for Big-5
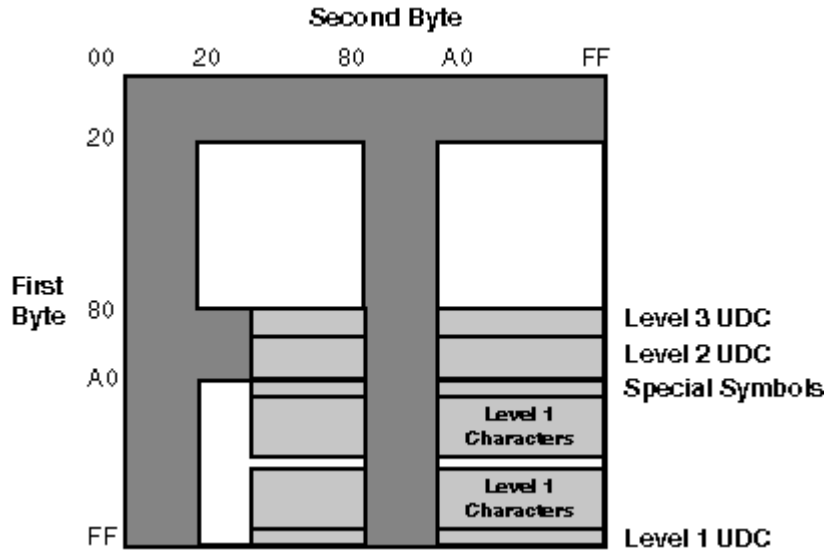
2.4 DEC Hanzi

The ASCII, GB2312-80 and extended GB character sets are combined to form the DEC Hanzi codeset.

DEC Hanzi, or simplified Chinese and denoted as `dechanzi`, uses a 2-byte data representation for symbols and ideographic characters defined in the GB2312-80 character set. To differentiate GB2312-80 codes from ASCII codes, the MSB of the first byte is always set on while that of the second byte is on for GB2312-80 and off for extended GB as shown in Figure 2–7.

Figure 2–7: DEC Hanzi Character Encoding
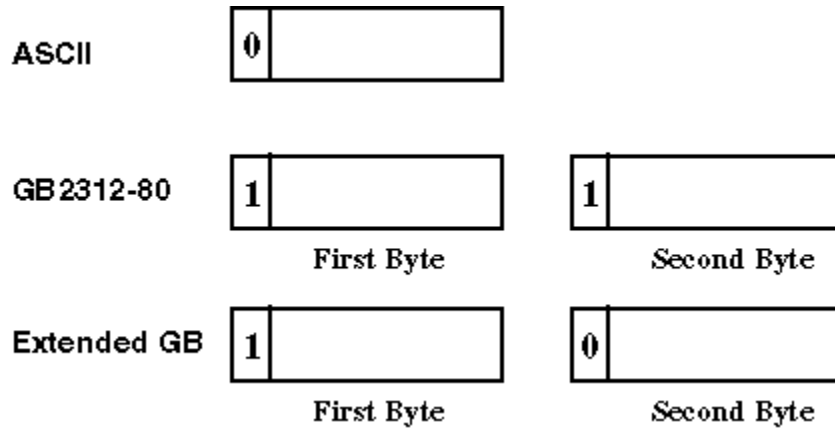
The first byte of a 2-byte code determines its row number, while the second byte determines its column number.

The following formulas illustrate the code of a GB2312-80 character or an extended GB character in relation to its row and column numbers.

GB2312-80 character:

First byte = A0 + row number
Second byte = A0 + column number

Extended GB character:

First byte = A0 + row number
Second byte = 20 + column number

For example, if a character is positioned at the first column of the 16th row on the GB2312-80 code plane, its encoding value is calculated as follows:

First byte = A0 (hex) + 16 = B0 (hex)

Second byte = A0 (hex) + 01 = A1 (hex)

The resulting encoded value is B0A1.

Similarly, if a character is positioned at the first column of the 16th row on the extended GB code plane, its encoding value is calculated as follows:

First byte = A0 (hex) + 16 = B0 (hex)
Second byte = 20 (hex) + 01 = 21 (hex)

The resulting encoded value is B021.

Figure 2–8 illustrates the division of a 2-byte code space and the position of the Chinese character sets.

Figure 2–8: GB2312-80 and Extended GB Code Space
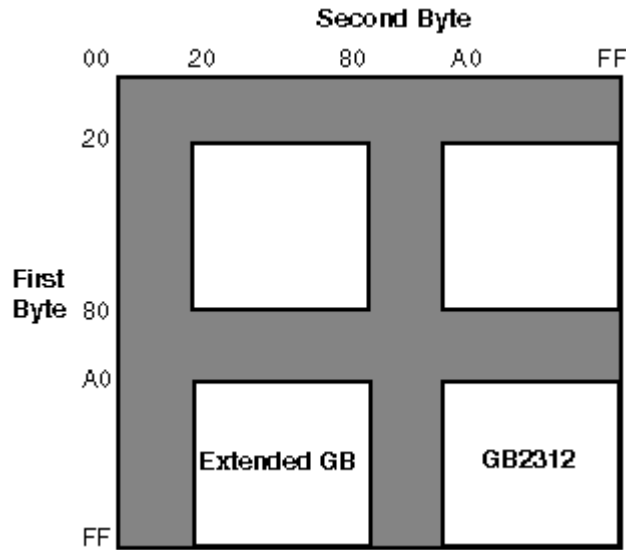
2.5 GBK

The GBK codeset is a character encoding system for simplified Chinese.

The GBK codeset provides a total of 23,940 code points, 21,886 of which are assigned. Each row in the GBK code table consists of 190 characters. ASCII characters, which are single-byte characters, are defined in the range 0x21 to 0x7E. Encoding ranges for 2-byte characters are 0x81 to 0xFE for the first byte and 0x40 to 0x7E and 0x80 to 0xFE for the second byte.

In terms of character-to-code allocation, the sub-range for GB2321-80 characters (0xA1A1-0xFEFE) in GBK is the same encoding range defined for these characters in EUC. GBK is therefore backward compatible with Chinese EUC coding as well as forward compatible with the encoding defined in the ISO 10646 standard.

The GBK codeset is divided into five levels as follows:

Level	Encoding Range	Code Points	Characters
GBK/1	0xA1A1 to 0xA9FE	846	717
GBK/2	0xB0A1 to 0xF7FE	6768	6763
GBK/3	0x8140 to 0xA0FE	6080	6080
GBK/4	0xAA40 to 0xFE40	8160	8160
GBK/5	0xA840 to 0xA9A0	192	166

In addition, the GBK codeset includes the following code points for user defined characters:

Encoding Range	Code Points
0xAAA1 to 0xAFFE	564
0xF8A1 to 0xFEFE	658
0xA140 to 0xA7A0	672

The operating system provides the following codeset converter pairs for converting simplified Chinese characters between GBK and Unicode formats (UTF-16, UCS-4, and UCF-8):

- UTF-16_GBK and GBK_UTF-16

- UCS-4_GBK and GBK_UCS-4

- UTF-8_GBK and GBK_UTF-8

2.6 GB18030

The GB18030 codeset provides 1-byte, 2-byte, and 4-byte encoding with the following structure:

Number of Bytes	Encoding Range	Code Points
1 byte	0x00 to 0x7F	128
2 byte	0x81 to 0xFE	23940
	0x40 to 0xFE (except 0x7F)	
4 byte	0x81 to 0xFE	1587600
	0x30 to 0x39	
	0x81 to 0xFE	
	0x30 to 0x39	

GB18030 1-byte code supports ASCII characters.

GB18030 2-byte code supports all the CJK characters (Chinese, Japanese, Korean) in the Unicode Version 2.1 Standard.

GB18030 4-byte code supports Unicode Version 3.0 additions. The 4-byte code also leaves a large number of unassigned codepoints available for future use.

The operating system provides the following codeset converter pairs for converting simplified Chinese characters between GB18030 and Unicode formats (UTF-16, UCS-4, and UTF-8):

- UTF-16_GB18030 and GB18030_UTF-16

- UCS-4_GB18030 and GB18030_UCS-4

- UTF-8_GB18030 and GB18030_UTF-8

_____ **Note** _____

The GB18030-2000 character set maps the invalid Unicode code points (U+FFFE and U+FFFF) to 4-byte codes. Because these two code points are invalid in UCS, this mapping can cause problems with conversions between the two character sets. Also, the GB18030-2000 character set does no mapping from 4-byte code to the UCS surrogate area (U+D800 to U+DFFF).

_____

2.7 Shift Big–5

The Shift Big-5 codeset, denoted as `sbig5`, is a variant of the Big-5 codeset. The difference between the two is that the second byte of some Big-5 characters is mapped to other values to form Shift Big-5 characters. Table 2–5 illustrates the mappings of Big-5 characters to Shift Big-5 characters.

Table 2–5: Big–5 to Shift Big–5 Character Mappings

Big-5 (Second Byte)	Shift Big-5 (Second Byte)
40	30
5B	31
5C	32
5D	33
5E	34
5F	35
60	36
7B	37
7C	38
7D	39
7E	9F

The Shift Big-5 codeset can be used in codeset conversion and terminal display. See Section 2.11 for details.

2.8 Telecode

The Telecode codeset (called Mitac Telex in earlier versions of the operating system), denoted as `telecode`, consists of two character planes. Each character plane has 8836 character positions. In Plane 1, standard characters occupy positions 0001 to 8045; the remaining 791 positions are for user-defined characters. In Plane 2, standard characters occupy positions 0001 to 8489; the remaining 346 positions are for user-defined characters. Telecode uses 2-byte values to represent characters on both planes.

_____ **Note** _____

For information about the character sets encoded by Telecode, see the *Chinese Code For Data Communication*.

_____

Telecode can be used in codeset conversion and terminal display. See Section 2.11 for further details.

2.8.1 Plane 1 Character Encoding

To differentiate Plane 1 code from Plane 2 code, MSB is set on in both bytes of a Plane 1 character code. You can use the following formula to calculate the value of a Plane 1 character from its position on the plane:

First byte = M + 161

Second byte = N + 161 - M x 94

In this formula, N is the position of the character and M = N / 94.

For example, if a character is at position 2502 on Plane 1, its encoded value is BBDB, which is calculated as follows:

N = 2502, M = 2502/94 = 26

First byte = 26 + 161 = 187 (or, BB (hex))

Second byte = 2502 + 161 - 26 x 94 = 219 (or, DB (hex))

2.8.2 Plane 2 Character Encoding

To differentiate Plane 2 code from Plane 1 code, the MSB of the first byte is set on and that of the second byte is set off for each Plane 2 character code. You can use the following formula to calculate the value of a Plane 2 character from its position:

First byte = M + 161

Second byte = N + 33 - M x 94

In this formula, N is the position of the character on the plane and M = N / 94.

For example, if a character is at position 2502 on Plane 2, its encoded value is BB5B, which is calculated as follows:

N = 2502, M = 2502/94 = 26

First byte = 26 + 161 = 187 (or, BB (hex))

Second byte = 2502 + 33 - 26 x 94 = 91 (or, 5B, hex))

2.9 UCS–4/UTF–16

The UCS codeset is a standard character encoding for the Universal Character Set (UCS) specified in Unicode and ISO/IEC 10646. There are two encoding schemes for UCS. An implementation that parses in 16-bit units (2 octets) is known as UTF-16. This is the canonical Unicode encoding in wide use on personal computers. An implementation that parses in 32-bit units (4 octets) is known as UCS-4. This is the canonical ISO/IEC 10646 encoding that is in use on systems that can support larger data unit size.

On Tru64 UNIX, UTF-16 and UCS-4 can be used in codeset conversion. In addition, UCS-4 is used as an internal process code for some locales. For codeset conversion, see Section 2.11. For locale variants, see Chapter 3.

2.10 UTF–8

The Unicode and ISO/IEC 10646 standards define transformation formats for the UCS. The following UCS transformation formats (UTFs) exist mainly to transform UCS values into sequences of bytes for handling by various byte-oriented protocols:

- UTF-8 is the standard method for transforming UCS-encoded data into a sequence of 8-bit bytes and ensuring interchange transparency for characters in C0 code positions (0 to 31), the SPACE (32) character, and the DEL (127) character.

- UTF-7 is the standard interchange format for environments that strip the eighth bit from each byte.

- UTF-16 is a transformation format that allows systems that are limited to processing of 16-bit units to support the extended character definition space that is included in UCS-4.

The operating system supports UTF-8 and UTF-16. UTF-8 can be used in codeset conversion and in locales. For codeset conversion, see Section 2.11. For locale variants, see Chapter 3.

2.11 Codeset Conversion

You may sometimes need to convert files from one codeset to another. Use the `iconv` utility to convert the encoding of characters in one codeset to another and write the results to standard output. Table 2–6 shows the pairs of Chinese codeset converters that are provided.

Table 2–6: Chinese Codeset Conversion

Conversion _from_to	Conversion _from_to
_from_Big-5_to	_from_Taiwan EUC_to
DEChanyu	DEC Hanyu
Taiwan EUC	Big-5
Shift Big-5	Shift Big-5
Telecode	Telecode
DEC Hanzi	DEC Hanzi
UCS-4	UCS-4
UTF-16	UTF-16
UTF-8	UTF-8
_from_DEC Hanyu_to	_from_DEC Hanzi_to
Taiwan EUC	DEC Hanyu
Big-5	Taiwan EUC
Telecode	Big-5
DEC Hanzi	UCS-4
UCS-4	UTF-16
UTF-16	UTF-8
UTF-8	
_from_UTF-8_to	_from_UTF-16_to
Taiwan EUC	Taiwan EUC
Big-5	Big-5
DEC Hanyu	DEC Hanyu
DEC Hanzi	DEC Hanzi
UCS-4	UCS-4
HKSCS	HKSCS
UTF-16	UTF-8

_from_GBK_to	_from_GB18030_to
UCS-4	UCS-4
UTF-16	UTF-16
UTF-8	UTF-8
_from_Telecode_to	_from_UTF-16_to
Taiwan EUC	Taiwan EUC
Big-5	Big-5
DEC Hanyu	DEC Hanyu
	HKSCS
	UCS-4
_from_Shift Big-5_to	UTF-8
Taiwan EUC	
Big-5	

For example, the following command converts a DEC Hanyu file to Big-5:

```
% iconv -f dechanyu -t big5 <file>
```

Table 2–7 shows the string names you can use as the parameters of the `iconv` utility.

Table 2–7: Codeset Names and Associated Strings

Codeset	String
DEC Hanyu	dechanyu
Taiwanese EUC	eucTW
Big-5	big5
Shift Big-5	sbig5
Telecode	Telecode
DEC Hanzi	dechanzi
GBK	GBK
GB18030	GB18030
Hong Kong Supplementary Character Set	HKSCS
Universal Codeset (4 octet form)	UCS-4
Universal Transfer Format (16-bit)	UTF-16
Universal Transfer Format (8-bit)	UTF-8

2.11.1 Default Conversion String

When converting from one codeset to another, characters in the source codeset that have no corresponding code point in the destination codeset are not converted. By default, the characters that cannot be converted are skipped and have no representation in the converted output.

You can control this behavior by using the ICONV_DEFSTR environment variable to define a default string to replace those unconvertible characters. If you specify a numeric value for this environment variable, the corresponding character value will be used.

The ICONV_DEFSTR environment variable affects all Chinese *iconv* converters. You can also use the "ICONV_DEFSTR_<from_code>_<to_code>" environment variable to control specific codeset conversion. For example, to convert a DEC Hanyu input file to DEC Hanzi with unconvertible characters converted to "?", you would enter the following commands:

```
%setenv ICONV_DEFSTR_dehanyu_dechanzi "?"
%iconv -f dechanyu -t dechanzi hanzi_input > hanyu_output
```

For codeset converters that end in UTF-16, UCS-4, or UTF-8, you can use the "U+XXXX" notation to specify the default character for conversion failure fallback.

_____ **Note** _____

During cut-and-paste operations, those traditional Chinese characters that cannot be converted to simplified Chinese characters are shown as default characters in the applications.

_____

2.11.2 One-to-Many Conversion

When converting from the DEC Hanzi codeset to other Chinese codesets, one simplified Chinese character may be mapped to multiple traditional Chinese characters. By default, the `iconv` utility picks up only the most likely candidate from a list of possible choices. You can control the behavior of the `iconv` utility with the ICONV_ACTION environment variable.

The ICONV_ACTION environment variable determines how the `iconv` utility behaves when there are one-to-many mappings. The possible values are:

- `batch`

  The most likely, or preferred, candidate will be picked up. This is the default. During cut-and-paste operations, the batch mode is always used for one-to-many character mappings.

- `conv_all`

  All possible choices are generated within the brackets "{" and "}" so that you can edit the converted file manually and determine which one should be used.

- `conv_all_nosym`

  All characters except symbols (for example, punctuation marks) are handled in the same manner as `conv_all`.

  _____ **Note** _____

  The ICONV_ACTION environment variable applies only to conversions of simplified Chinese to traditional Chinese and has no effect on UCS-4 and UTF-8 converters.
  _____

2.11.3 User Defined Character Mappings

Some user defined characters in the Big-5 codeset have predefined mappings to user defined spaces in DEC Hanyu. Table 2–8 shows this mapping.

Table 2–8: Mapping Between Big–5 and DEC Hanyu User Defined Characters

DEC Hanyu	Big-5	Code Size
F321 - FB41	FA40 - FEFE	785
FB42 - FEFE	8E40 - 905C	343
AAA1 - C1FE	905D - 9EB8	2256

These predefined user defined character mappings are supported by both the `iconv` methods and the terminal driver.

Some user defined characters do not have predefined mappings. You should use only those user defined characters that have predefined mappings.

2.11.4 Hong Kong Supplementary Character Set

HKSCS, the Hong Kong Supplementary Character Set, was developed by the Chinese government, in collaboration with the Chinese Language Interface Advisory Committee, to provide Chinese characters needed in computing in Hong Kong. The characters contained in HKSCS are only for computer use and can be represented as either Big-5 or Unicode.

While HKSCS is not a character set name, the operating system uses it as the name for extended Big-5 encoding that contains HKSCS characters. Currently, HKSCS support is limited to codeset conversion between HKSCS and Unicode.

See HKSCS.5 for more information.

2.12 Codeset for Peripheral Devices

The operating system provides a mechanism for you to use to configure your system to run applications with peripherals, such as terminals and printers, that support different codesets. You can specify the codesets for the applications, terminals, and printers independently, as described in Table 2–9. The operating system software automatically does the necessary codeset conversion.

Table 2–9: Feasible Chinese Codesets for Applications, Terminals, and Printers

Application Code	Terminal Code	Printer Code
DEC Hanyu	DEC Hanyu	DEC Hanyu
Taiwanese EUC	Taiwanese EUC	Taiwanese EUC
Big-5	Big-5	Big-5
DEC Hanzi	DEC Hanzi	DEC Hanzi
none	Shift Big-5	none
none	Telecode	none
UTF-8	UTF-8	none

Chinese DECterm software supports only DEC Hanyu, Big5, or DEC Hanzi as its terminal code. You must activate the stty drive and set tcode to dechanyu when running in a Taiwanese EUC locale. For example:

```
% stty adec tcode dechanyu
```

The dxterm does not support UTF-8 as a terminal code. Use dtterm when UTF-8 is required for a terminal code.

For details about setting up codesets for terminals and printers, see *Using International Software*.

3

Locales

The Tru64 UNIX operating system supports different Chinese locales for different countries and areas. These include Taiwan, People's Republic of China (PRC), and Hong Kong. Table 3–1 shows the valid Chinese locales with different countries, codesets, and collating sequences.

_____ **Note**_____

zh_TW is an alias of zh_TW.eucTW, and zh_CN is an alias of zh_CN.dechanzi.

_____

Table 3–1: Chinese Locales

Codeset	Locale	Collation Sequence
DEC Hanyu	Zh_TW	Internal code
	zh_TW.dechanyu	Internal code
	zh_TW.dechanyu@radical	Radical
	zh_TW.dechanyu@stroke	Stroke
	zh_TW.dechanyu@chuyin	Chuyin (Phonetic)
	zh_HK.dechanyu	Internal code

Table 3-1: (Cont.) Chinese Locales

Codeset	Locale	Collation Sequence
Taiwanese EUC	zh_TW.eucTW zh_TW.eucTW@radical zh_TW.eucTW@stroke zh_TW.eucTW@chuyin zh_HK.eucTW	Internal code Radical Stroke Chuyin (Phonetic) Internal code
Big-5	zh_TW.big5 zh_TW.big5@radical zh_TW.big5@stroke zh_TW.big5@chuyin zh_HK.big5	Internal code Radical Stroke Chuyin (Phonetic) Internal code
DEC Hanzi	zh_CN zh_CN.dechanzi zh_CN.dechanzi@radical zh_CN.dechanzi@stroke zh_CN.dechanzi@pinyin zh_HK.dechanzi	Internal code Internal code Radical Stroke Pinyin (Phonetic) Internal code
GBK	zh_CN.GBK	Internal code (however, all characters in GB 2312 character set are ordered first, followed by GBK extension characters)
GB18030	zh_CN.GB18030	Internal code (however, all characters in GB 2312 character set are ordered first, followed by GBK characters not already ordered, followed by GB18030 characters not already ordered)
UTF-8	zh_CN.UTF-8 zh_TW.UTF-8 zh_HK.UTF-8	zh_CN.GB18030 zh_TW.dechanyu zh_HK.dechanyu

The collation sequence for the GBK locale is further refined so that characters in the GB 2312 character set are ordered first followed by GBK extension characters.

The collation sequence for the GB18030 locale is further refined so that characters in the GB 2312 character set are ordered first, followed by GBK characters not already ordered, followed by GB18030 characters not already ordered.

Locales that support the same country and codeset are basically the same. The radical, stroke, pinyin, and chuyin modifiers after the at (@) sign specify different criterion for collation and sorting. Moreover, the characters defined in character set standards have collating precedence over user-defined characters, which, in turn, have precedence over undefined or reserved characters.

The dechanyu, eucTW, and dechanzi locales support locale names with @ucs4 modifiers. The @ucs4 modifier indicates that UCS-4 is used as the internal processing code. However, these locales are the same as the locales without the @ucs4 modifier. The @ucs4 subset is provided for backward compatibility and may be removed in the future. See l10n_intro.5 for more information.

If you are using DECwindows Motif, you can select the locale through the Language Menu of Session Manager. If you are using CDE, you can select the locale using the language menu on the CDE login screen. If you are superuser, you can use the SYSMAN I18NCONFIG utility to select a locale and define it as the system default.

The applicable locales are shown in Table 3–2.

Table 3–2: Chinese Language Names

Locale	Language Name
zh_TW	Traditional Chinese Taiwan
zh_TW.dechanyu	Traditional Chinese Taiwan (DEC Hanyu)
zh_TW.eucTW	Traditional Chinese Taiwan (EUC)
zh_TW.big5	Traditional Chinese Taiwan (Big5)
zh_TW.UTF-8	Traditional Chinese Taiwan (Unicode)
zh_HK.dechanyu	Traditional Chinese Hong Kong (DEC Hanyu)
zh_HK.dechanzi	Simplified Chinese Hong Kong (DEC Hanzi)
zh_HK.eucTW	Traditional Chinese Hong Kong (EUC)
zh_HK.big5	Traditional Chinese Hong Kong (Big5)
zh_HK.UTF-8	Traditional Chinese Hong Kong (Unicode)
zh_CN	Simplified Chinese PRC
zh_CN.dechanzi	Simplified Chinese PRC (DEC Hanzi)
zh_CN.GBK	Simplified Chinese PRC (GBK)
zh_CN.GB18030	Simplified Chinese PRC (GB18030)
zh_CN.UTF-8	Simplified Chinese PRC (Unicode)

_____ **Note** _____

For DECwindows Motif and CDE, the locale modifier is ignored.
_____

4

Local Language Devices

4.1 Terminals

The Tru64 UNIX operating system supports Hanyu and Hanzi DECterm. Hanyu and Hanzi DECterm are the respective emulation of the traditional and simplified Chinese terminals, which provide compatible functions for running Chinese character-cell terminal applications. For details on Hanyu and Hanzi DECterm, see Chapter 9. You can also use `dtterm` running in a Chinese locale to display Chinese in a character-cell application.

4.2 Printers

The Tru64 UNIX operating system supports the following dot matrix Chinese printers:

- CP382-D   (traditional Chinese)
- LA88-C   (simplified Chinese)
- LA380-CB   (simplified Chinese)

The following PostScript printers can be configured for traditional or simplified Chinese:

- DEClaser 1152
- DEClaser 5100 with font disk (LN09X-HD)
- PrintServer 17
- All PostScript level 2 printers

The print filters listed in Table 4–1 are provided to support these Chinese printers.

Table 4–1: Chinese Print Filters

Filter Name	Printer Name
cp382dof	CP382-D printer
la88cof	LA88-C printer
la380cbof	LA380-CB printer
dl1152wrof	DEClaser 1152
dl5100wrof	DEClaser 5100
lpsof	PrintServer 17
wwpsof	Level 2 PostScript printers

_____ **Note** _____

To use PrintServer 17, PrintServer Software Version 5.0 or later must be installed on the operating system.
_____

For details about setting up Chinese printer queues, see Chapter 8.

5

Fonts

5.1 DECwindows Fonts

The Tru64 UNIX operating system provides Chinese DECwindows fonts in various sizes and typefaces for 75 dpi (dots-per-inch) display devices. Table 5–1 lists the screen fonts for traditional Chinese.

Table 5–1: Traditional Chinese Screen Fonts

Typefaces	Glyph Size	Bounding Box	Remarks
Screen	15 x 16	16 x 18	Mandatory font
	22 x 22	24 x 24	Mandatory font
Sung	22 x 22	24 x 24	Optional font
	30 x 30	32 x 32	Optional font
Hei	15 x 16	16 x 16	Optional font
	22 x 22	24 x 24	Optional font

There are two sets of DECwindows fonts, one for CNS 11643-1986 and one for DTSCS.

Table 5–2 lists the screen fonts for simplified Chinese.

Table 5–2: Simplified Chinese Screen Fonts

Typefaces	Glyph Size	Bounding Box	Remarks
Screen	15 x 16*	16 x 18	Mandatory font, defined in GB5199.1-85
	22 x 22*	24 x 24	Mandatory font
Songti	15 x 16*	16 x 16	Optional font, defined in GB5199.1-85
	22 x 22*	24 x 24	Optional font
	32 x 32*	34 x 34	Optional font, defined in GB6345.1-86
Heiti	15 x 16	16 x 16	Optional font
	22 x 22*	24 x 24	Optional font
	32 x 32*	34 x 34	Optional font, defined in GB12036-89
Fangsongti	22 x 22*	24 x 24	Optional font
	32 x 32*	34 x 34	Optional font, defined in GB12034-89
Kaiti	22 x 22*	24 x 24	Optional font
	32 x 32*	34 x 34	Optional font, defined in GB12035-89

*The fonts marked with an asterisk are supplied by China Standard Technology Development Corporation (CSTDC) of People's Republic of China.

In addition to these Chinese fonts, several miscellaneous screen fonts are provided for use in Hanyu and Hanzi DECterm, and the Motif toolkit. The operating system also provides default TrueType fonts for DEC Hanzi, GBK, and GB18030 codesets as described in Section 5.3.

The mandatory fonts are available after you install the Chinese language support from the worldwide language support software. Other optional fonts are available only if you install the optional Chinese font subsets. If you do not find the optional fonts on your system, contact your system administrator.

No 100 dpi Chinese fonts are provided in the kit. To allow you to use the Chinese fonts on 100 dpi display devices, a font alias file is provided to map the 75 dpi font names to 100 dpi font names.

5.1.1 XLFD Font Names

You must specify the DECwindows font names in X Logical Font Description (XLFD) format in your application programs or in the application resource files. You can specify wildcards (*) for any fields in the font names.

You can use the following font names for both 75 dpi and 100 dpi display devices. If you want to state the display resolution explicitly, you can specify 75 or 100 in the X- and Y-resolution fields, that is, the second and third asterisks in the following XLFD names.

- Traditional Chinese Screen family font names in XLFD format:

  CNS 11643-1986 Fonts

  -ADECW-Screen-Medium-R-Normal--*-180-*-*-M-160-DEC.CNS11643.1986-2
  -ADECW-Screen-Medium-R-Normal--*-240-*-*-M-240-DEC.CNS11643.1986-2

  DTSCS Fonts

  -ADECW-Screen-Medium-R-Normal--*-180-*-*-M-160-DEC.DTSCS.1990-2
  -ADECW-Screen-Medium-R-Normal--*-240-*-*-M-240-DEC.DTSCS.1990-2

- Traditional Chinese Sung family font names in XLFD format:

  CNS 11643-1986 Fonts

  -ADECW-Sung-Medium-R-Normal--*-240-*-*-M-240-DEC.CNS11643.1986-2
  -ADECW-Sung-Medium-R-Normal--*-320-*-*-M-320-DEC.CNS11643.1986-2

   DTSCS Fonts

  -ADECW-Sung-Medium-R-Normal--*-240-*-*-M-240-DEC.DTSCS.1990-2
  -ADECW-Sung-Medium-R-Normal--*-320-*-*-M-320-DEC.DTSCS.1990-2

- Traditional Chinese Hei family font names in XLFD format:

  CNS 11643-1986 Fonts

  -ADECW-Hei-Medium-R-Normal--*-160-*-*-M-160-DEC.CNS11643.1986-2
  -ADECW-Hei-Medium-R-Normal--*-240-*-*-M-240-DEC.CNS11643.1986-2

  DTSCS Fonts

  -ADECW-Hei-Medium-R-Normal--*-160-*-*-M-160-DEC.DTSCS.1990-2
  -ADECW-Hei-Medium-R-Normal--*-240-*-*-M-240-DEC.DTSCS.1990-2

- Simplified Chinese Screen family font names in XLFD format:

  -ADECW-Screen-Medium-R-Normal--*-180-*-*-M-160-GB2312.1980-1
  -ADECW-Screen-Medium-R-Normal--*-240-*-*-M-240-GB2312.1980-1

- Simplified Chinese Songti family font names in XLFD format:

  -ADECW-Songti-Medium-R-Normal--*-160-*-*-M-160-GB2312.1980-1
  -ADECW-Songti-Medium-R-Normal--*-240-*-*-M-240-GB2312.1980-1
  -ADECW-Songti-Medium-R-Normal--*-340-*-*-M-340-GB2312.1980-1

- Simplified Chinese Heiti family font names in XLFD format:

  -ADECW-Heiti-Medium-R-Normal--*-160-*-*-M-160-GB2312.1980-1
  -ADECW-Heiti-Medium-R-Normal--*-240-*-*-M-240-GB2312.1980-1
  -ADECW-Heiti-Medium-R-Normal--*-340-*-*-M-340-GB2312.1980-1

- Simplified Chinese Fangsongti family font names in XLFD format:

  -ADECW-Fangsongti-Medium-R-Normal--*-240-*-*-M-240-GB2312.1980-1
  -ADECW-Fangsongti-Medium-R-Normal--*-340-*-*-M-340-GB2312.1980-1

- Simplified Chinese Kaiti family font names in XLFD format:

  -ADECW-Kaiti-Medium-R-Normal--*-240-*-*-M-240-GB2312.1980-1
  -ADECW-Kaiti-Medium-R-Normal--*-340-*-*-M-340-GB2312.1980-1

Table 5–3 shows the font names, in XLFD format, of several miscellaneous Chinese screen fonts.

Table 5–3: XLFD of Miscellaneous Chinese Screen Fonts

XLFD Font Name	Character Set
-ADECW-Screen-Medium-R-Normal--*-180-*-*-M-80-iso8859-1	ISO Latin-1
-ADECW-Screen-Medium-R-Normal--*-180-*-*-M-80-DEC-DECctrl	DEC Display Control
-ADECW-Screen-Medium-R-Normal--*-180-*-*-M-80-DEC-DECsuppl	DEC Supplemental
-ADECW-Screen-Medium-R-Normal--*-180-*-*-M-80-DEC-DECtech	DEC Technical
-ADECW-Screen-Medium-R-Normal--*-180-*-*-M-80-DEC-DRCS	DEC DRCS
-ADECW-Screen-Medium-R-Normal--*-240-*-*-M-120-iso8859-1	ISO Latin-1
-ADECW-Screen-Medium-R-Normal--*-240-*-*-M-120-DEC-DECctrl	DEC Display Control
-ADECW-Screen-Medium-R-Normal--*-240-*-*-M-120-DEC-DECsuppl	DEC Supplemental
-ADECW-Screen-Medium-R-Normal--*-240-*-*-M-120-DEC-DECtech	DEC Technical
-ADECW-Screen-Medium-R-Normal--*-240-*-*-M-120-DEC-DRCS	DEC DRCS

5.1.2 Bitmap Font Samples

Figure 5–1 through Figure 5–6 illustrate samples of Chinese fonts.

Figure 5–1: Sung Font Sample
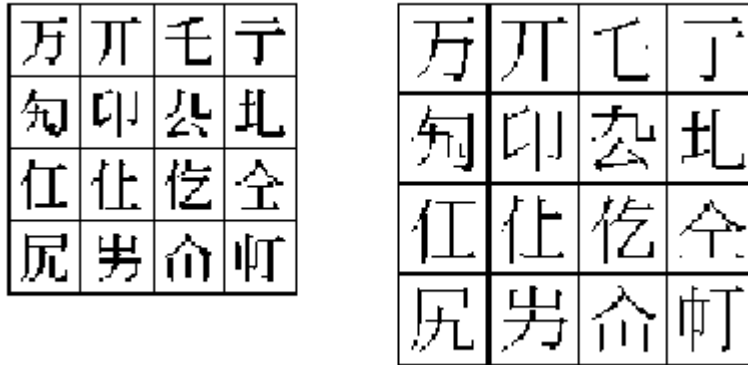
Figure 5–2: Hei Font Sample
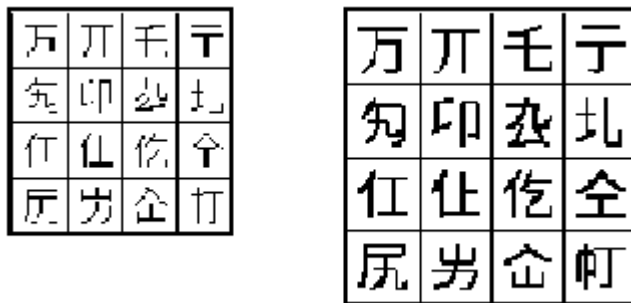
Figure 5–3: Songti Font Sample

啊 阿 埃 挨
氨 安 俺 按
傲 奥 懊 澳
把 耙 坝 霸

啊 阿 埃 挨
氨 安 俺 按
傲 奥 懊 澳
把 耙 坝 霸

啊 阿 埃 挨
氨 安 俺 按
傲 奥 懊 澳
把 耙 坝 霸

Figure 5–4:Heiti Font Sample

啊 阿 埃 挨
氨 安 俺 按
傲 奥 懊 澳
把 耙 坝 霸

啊 阿 埃 挨
氨 安 俺 按
傲 奥 懊 澳
把 耙 坝 霸

啊 阿 埃 挨
氨 安 俺 按
傲 奥 懊 澳
把 耙 坝 霸
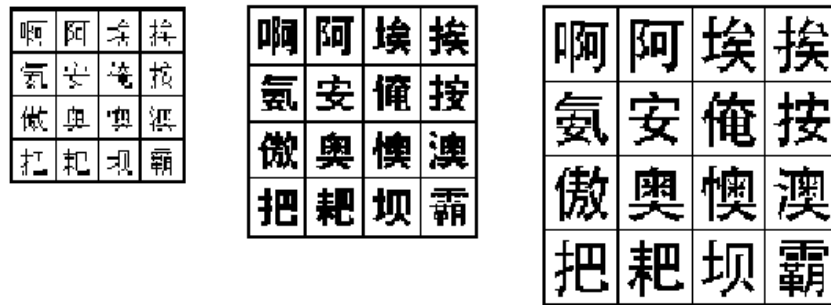
Figure 5–5: Fangsongti Font Sample
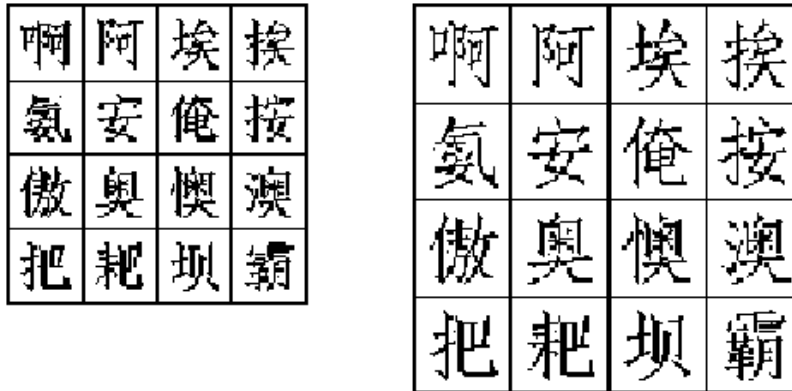
Figure 5–6: Kaiti Font Sample
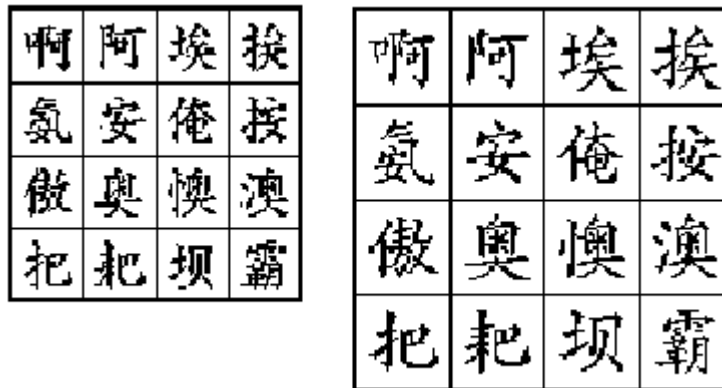
5.1.3 Font Encodings

The X Consortium registers names for font encodings that are used in XLFDs. However, no names currently are registered for CNS 11643 and DTSCS. Therefore, they are currently supported as private encodings as shown in Table 5–4.

Table 5–4: Chinese DECwindows Font Encodings

Character Set	Character Set Registry
CNS 11643-1986	DEC.CNS11643.1986-2
DTSCS	DEC.DTSCS.1990-2

Because the X Window System provides only basic Xlib functions for handling 8-bit and 16-bit characters, the 4-byte data representation of DTSCS is trimmed to remove the two leading bytes, C2 CB, to form a 2-byte encoding. DECwindows applications should either preprocess the 4-byte data and then handle them with the low level Xlib functions or handle Chinese strings with the internationalized text drawing functions provided by X11R6 Xlib or Motif Toolkit.

Figure 5–7 and Figure 5–8 illustrate these two encoding schemes.

Figure 5–7: CNS 11643-1986 Font Encoding Scheme
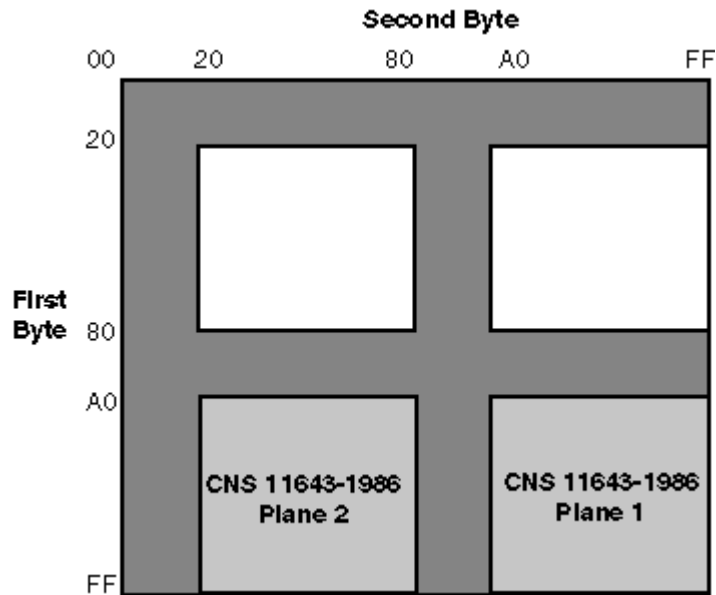
Figure 5–8: DTSCS Font Encoding Scheme
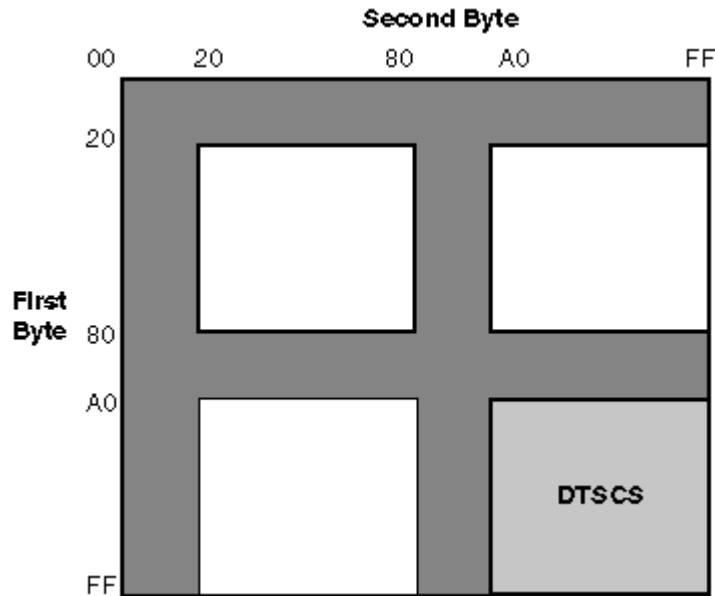
Vendors may adopt different encoding schemes or even different character sets to produce their fonts. The fonts supplied by the operating system are all in the encoding schemes defined in this section. To allow you to run applications on third-party workstations on which different font encodings are installed, the Tru64 UNIX implementation of X11R6 Xlib supports the conversion of encodings during text display.

Table 5–5 shows these encoding conversions.

Table 5–5: Font Encoding Conversion

Character Set	Convert From	Convert To
Taiwanese EUC	euctw-1 (plane 1)	dec.cns11643.1986-2
	euctw-2 (plane 2)	dec.cns11643.1986-2
	euctw-3 (plane 3)	dec.dtscs.1990-2
	euctw-4 (plane 4)	dec.dtscs.1990-2
Big-5	big5-0	dec.cns11643.1986-2

For simplified Chinese, the X Window System defines two encodings for the GB2312-80 character set as shown in Table 5–6. Figure 5–9 illustrates the GB2312–80 encoding scheme.

Table 5–6: Chinese DECwindows Font Encodings

Encoding	Character Set Registry
GL	GB2312.1980-0
GR	GB2312.1980-1

Figure 5–9: GB2312-80 Font Encoding Schemes
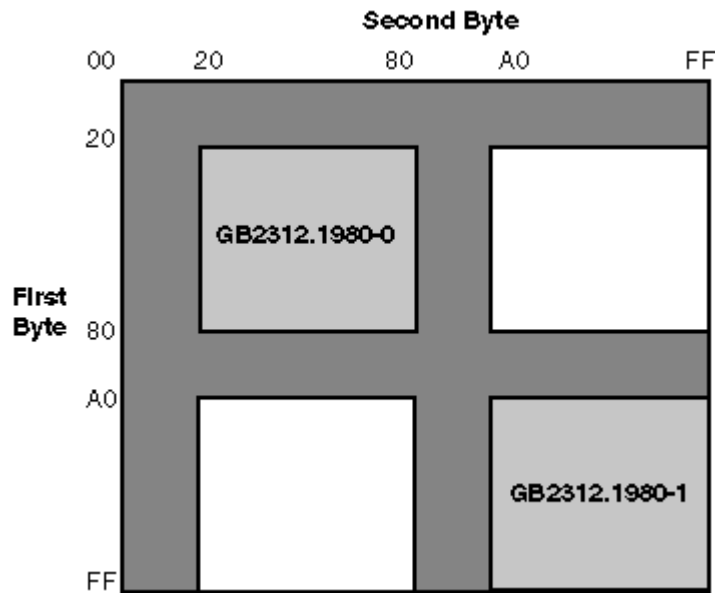
The Chinese DECwindows fonts supplied by the operating system are all in GR encoding. To allow you to run applications on third-party workstations on which only GL-encoded fonts are installed, the Tru64 UNIX implementation of X11R6 Xlib supports the conversion of GR encoding to GL encoding for text drawing and measurement, as shown in Table 5–7.

Table 5–7: GR to GL Font Encoding Conversion

Convert From	Convert To
gb2312.1980-1	gb2312.1980-0

For details, see the *Writing Software for the International Market* manual in the Tru64 UNIX documentation set.

5.1.4 Specifying Fonts in DECwindows Applications

Table 5–8 and Table 5–9 show the default fonts used in the Motif Toolkit.

Table 5–8: Traditional Chinese Default Fonts

XLFD Font Name	Character Set
-ADECW-Screen-Medium-R-Normal--*-180-*-*-M-80-iso8859-1	ISO8859-1
-ADECW-Screen-Medium-R-Normal--*-180-*-*-M-160-DEC.CNS11643.1986-2	DEC.CNS11643.1986-2
-ADECW-Screen-Medium-R-Normal--*-180-*-*-M-160-DEC.CNS11643.1986-2-UDC	DEC.CNS11643.1986-2-UDC
-ADECW-Screen-Medium-R-Normal--*-180-*-*-M-160-DEC.DTSCS.1990-2	DEC.DTSCS.1990-2
-ADECW-Screen-Medium-R-Normal--*-180-*-*-*-*-*	Fontset

Table 5–9: Simplified Chinese Default Fonts

XLFD Font Name	Character Set
-ADECW-Screen-Medium-R-Normal--*-180-*-*-M-80-iso8859-1	ISO8859-1
-ADECW-Screen-Medium-R-Normal--*-180-*-*-M-160-GB2312.1980-1	GB2312.1980-1
-ADECW-Screen-Medium-R-Normal--*-180-*-*-M-160-GB2312.1980-UDC	GB2312.1980-UDC
-ADECW-Screen-Medium-R-Normal--*-180-*-*-*-*-*	Fontset

To override the default fonts of a traditional Chinese DECwindows application, you should specify the ISO Latin-1, DTSCS, and CNS11643 (UDC) fonts as well as the Chinese fontset when creating widget instances. For a simplified Chinese DECwindows application, you should specify the ISO Latin-1, GB2312-80, and extended GB (UDC) fonts as well as the Chinese fontset when creating widget instances. For details, see *Writing Software for the International Market*.

5.2 Outline Fonts

Tru64 UNIX provides the following traditional and simplified Chinese outline fonts for printing on PostScript printers and for display through Level II Display PostScript extension.

For traditional Chinese:

- Sung-Light-CNS11643
- Hei-Light-CNS11643

For simplified Chinese:

- XiSong-GB2312-80
- Hei-GB2312-80

The encoding of these fonts is the same as that illustrated in Figure 5–7 and Figure 5–9.

These Chinese outline fonts have the following uses:

- Printing on PostScript printers. For details see Chapter 8.

- Displaying through the R6 X Windows System Type 1 rasterizer. To use these outline fonts, add the `$I18NPATH/usr/lib/X11/fonts/TChinesePS` and `$I18NPATH/usr/lib/X11/fonts/SChinesePS` directories to your font path with the following command:

  ```
  % xset +fp $I18NPATH/usr/lib/X11/fonts/TChinesePS,
  $I18NPATH/usr/lib/X11/fonts/SchinesePS
  ```

  This is done automatically when the outline fonts are installed.

- Displaying through Display PostScript. You can view PostScript files with Chinese characters using the CDA Viewer or through the Display PostScript extension.

5.2.1 XLFD Font Names of Chinese Outline Fonts

To use the Chinese outline fonts through the Type 1 rasterizer, you can specify the font names in XLFD (X Logical Font Description) format in your application programs or in the application resource files, just like ordinary DECwindows bitmap fonts.

To specify the XLFD font name of an outline font, you should replace the fields currently marked with 0 (zero) with the following information:

- Field 1 — The font height in number of dots. An asterisk (*) usually is entered in this field.

- Field 2 — The font height in point size. For example, you can enter 240 to specify a 24-point font.

- Fields 3 and 4 — The X- and Y-resolution. They usually have the value of 75 or 100.

- Field 5 — The average font width in point size. An asterisk (*) usually is put in this field.

For example, if you want to use a 48-point font of the Sung-Light-CNS11643 family for a 100 dpi display device, you would specify:

```
-dyna-sung-medium-r-normal--*-480-100-100-m-*-CNS11643.1986
```

5.3 TrueType Fonts

The operating system provides both screen and printer TrueType fonts for Chinese characters sets. The following sets of simplified Chinese TrueType fonts are installed as the default fonts for the DEC Hanzi, GBK, and GB18030 codesets.

DEC Hanzi Fonts:

FangSong

-css_dongwen-fangsong-medium-r-normal--0-0-0-0-c-0-gb2312.1980-0
-css_dongwen-fangsong-medium-r-normal--0-0-0-0-c-0-gb2312.1980-1
-css_dongwen-fangsong-medium-r-normal--0-0-0-0-c-0-iso8859-1

HeiTi

-css_dongwen-heiti-medium-r-normal--0-0-0-0-c-0-gb2312.1980-0
-css_dongwen-heiti-medium-r-normal--0-0-0-0-c-0-gb2312.1980-1
-css_dongwen-heiti-medium-r-normal--0-0-0-0-c-0-iso8859-1

KaiTi

-css_dongwen-kaiti-medium-r-normal--0-0-0-0-c-0-gb2312.1980-0
-css_dongwen-kaiti-medium-r-normal--0-0-0-0-c-0-gb2312.1980-1
-css_dongwen-kaiti-medium-r-normal--0-0-0-0-c-0-iso8859-1

SongTi

-css_dongwen-songti-medium-r-normal--0-0-0-0-c-0-gb2312.1980-0
-css_dongwen-songti-medium-r-normal--0-0-0-0-c-0-gb2312.1980-1
-css_dongwen-songti-medium-r-normal--0-0-0-0-c-0-iso8859-1

GBK Fonts:

FangSong

-css_dongwen-fangsong-medium-r-normal--0-0-0-0-c-0-gbk-1
-css_dongwen-fangsong-medium-r-normal--0-0-0-0-c-0-iso8859-1

HeiTi

-css_dongwen-heiti-medium-r-normal--0-0-0-0-c-0-gbk-1
-css_dongwen-heiti-medium-r-normal--0-0-0-0-c-0-iso8859-1

KaiTi

-css_dongwen-kaiti-medium-r-normal--0-0-0-0-c-0-gbk-1
-css_dongwen-kaiti-medium-r-normal--0-0-0-0-c-0-iso8859-1

SongTi

-css_dongwen-songti-medium-r-normal--0-0-0-0-c-0-gbk-1
-css_dongwen-songti-medium-r-normal--0-0-0-0-c-0-iso8859-1

GB18030 Fonts:

FangSong

-css_dongwen-fangsong-medium-r-normal--0-0-0-0-c-0-iso10646-1
-css_dongwen-fangsong-medium-r-normal--0-0-0-0-c-0-iso8859-1

HeiTi

-css_dongwen-heiti-medium-r-normal--0-0-0-0-c-0- iso10646-1
-css_dongwen-heiti-medium-r-normal--0-0-0-0-c-0-iso8859-1

KaiTi

-css_dongwen-kaiti-medium-r-normal--0-0-0-0-c-0- iso10646-1
-css_dongwen-kaiti-medium-r-normal--0-0-0-0-c-0-iso8859-1

SongTi

-css_dongwen-songti-medium-r-normal--0-0-0-0-c-0- iso10646-1
-css_dongwen-songti-medium-r-normal--0-0-0-0-c-0-iso8859-1

6

Keyboards

The Tru64 UNIX operating system supports any standard English PC-style keyboard to enter Chinese. See `keyboard` (5) for more information.

There are some keyboard variants for different countries. For example, in Taiwan, the LK201-D variant has additional symbols printed on the keycaps for different Chinese input methods. The following figures illustrate some of the Chinese keyboard layouts. You can find online copies of these figures at the location specified. These figures are in .ddif format.

Figure 6–1: LK201-D Keyboard Layout
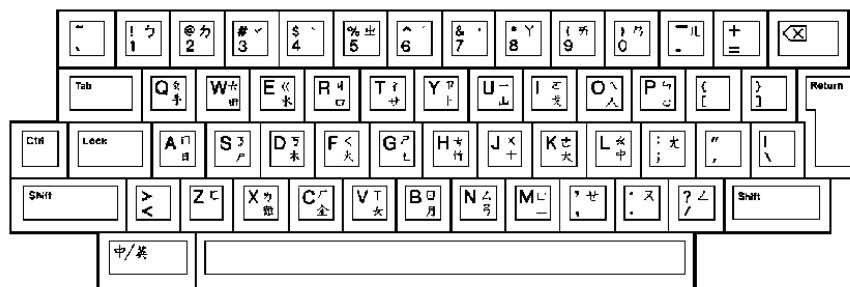
Required Keymap: us_lk201re

Location of file: /usr/lib/cda/hanyu-lk201d-100.ddif

Figure 6–2: LK401-D Keyboard Layout
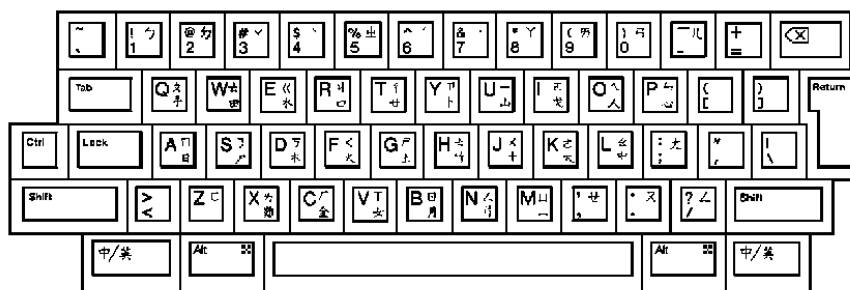
Required Keymap us_lk401aa

Location of file: /usr/lib/cda/hanyu-lk401d-100.ddif

Figure 6–3: LK201-C Keyboard Layout
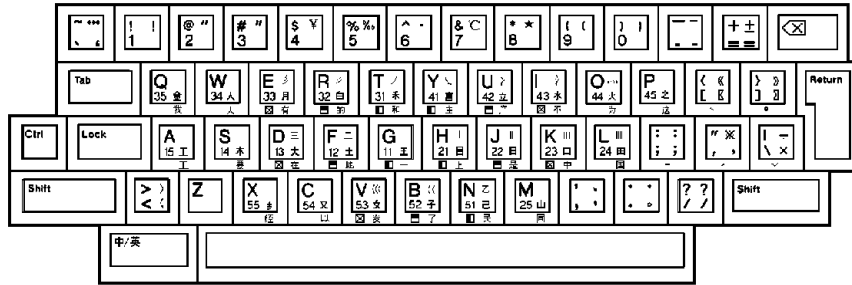
Required Keymap: us_lk201re

Location of file: /usr/lib/cda/hanzi-lk201c-100.ddif

Figure 6–4: LK401-C Keyboard Layout
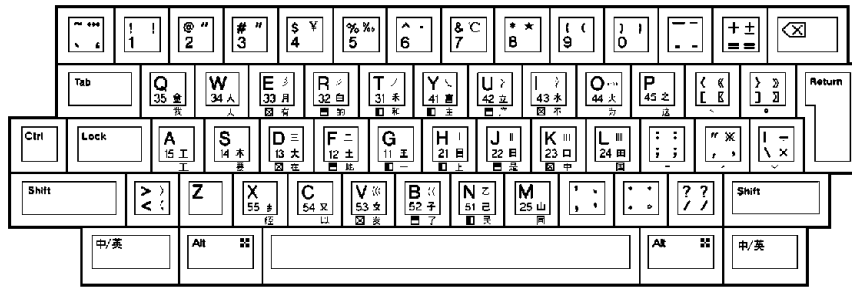
Required Keymap   us_lk401aa

Location of file: /usr/lib/cda/hanzi-lk401c-100.ddif

Figure 6–5: Numeric Keypad for 5-Stroke Input Method
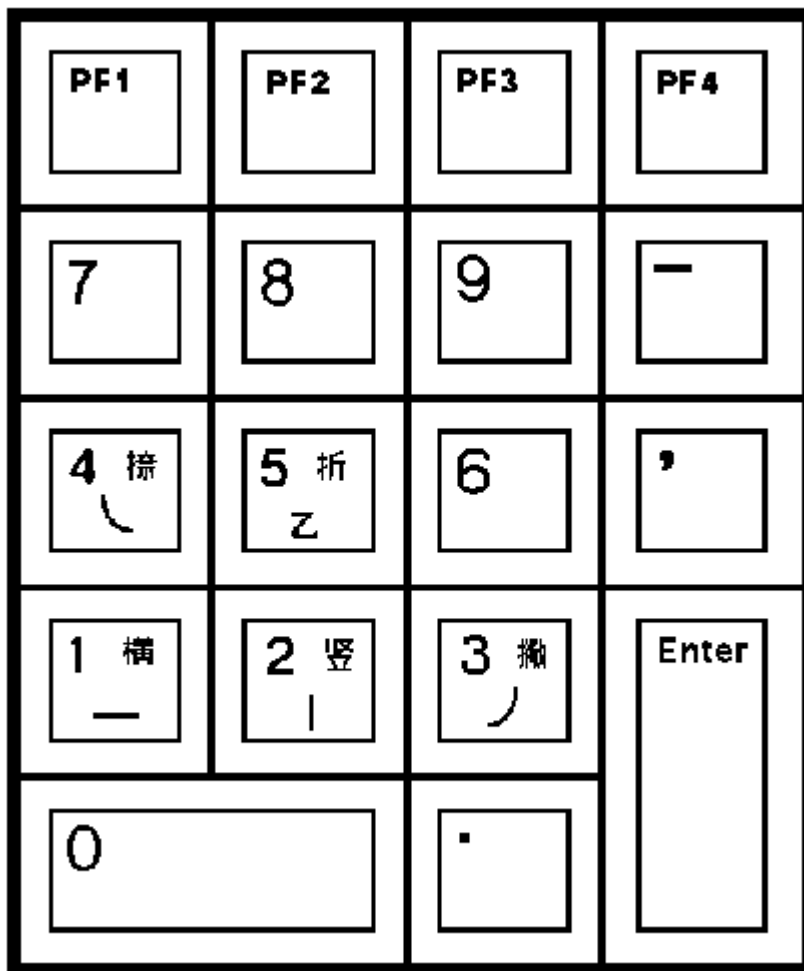
Location of file: /usr/lib/cda/hanzi-keypad-100.ddif

7

Input Methods

There are two groups of input methods for entering Chinese characters on Tru64 UNIX. One group of input methods supports traditional Chinese and the other group supports simplified Chinese.

Traditional Chinese has the following input methods:

- Full-form alphabets
- Tsang-Chi （倉頡）
- Quick Tsang-Chi, also known as Easy （速成）
- Phonetic （注音）
- Internal Code （內碼）
- Phrase input method （片語）
- Symbol Input （符號）

Simplified Chinese has the following input methods:

- Full-form alphabets
- 5-Stroke 〔五笔画〕
- 5-Shape 〔五笔字型〕
- Pin-Yin, or Phonetic 〔拼音〕
- Qu-Wei or Row-Column in GB2312-80 〔国标区位码〕
- Telex Code 〔电报码〕
- Phrase input method〔片语〕
- ItABC or Intelligent ABC

In addition to these input methods, the operating system provides the dxim, dxhanziim, and dxhanyuim input servers. The dxhanziim input server supports the simplified Chinese input methods. The dxhanyuim input server supports the traditional Chinese input methods. dxim is a menu-driven input server that allows you to configure, manage, and use both the traditional and simplified Chinese input methods. dxim also provides a Phrase input method that is separate from the operating system Phrase Utility.

This chapter describes:

- Each of the input methods listed above.

- How to activate, deactivate, and customize Chinese input methods and input method servers.

- How to switch between input methods.

- How to work with input methods under the DECwindows Motif interface.

7.1 Activating Chinese Input Methods

You must activate traditional and simplified Chinese input methods for response to the requirements of applications that are internationalized for the Chinese language. The steps you follow to activate Chinese input methods differ for character cell terminals, DECwindows Motif and CDE environments, and the dxim input server. The following sections describe these activation steps.

7.1.1 Chinese Input Methods and Character–Cell Terminal Applications

Traditional Chinese input methods for character-cell terminal applications are incorporated in Hanyu DECterm terminal emulation software. Simplified Chinese input methods for character-cell terminal applications are incorporated in the Hanzi DECterm terminal emulation software. Applications do not need to provide their own support for Chinese input. They can rely on the terminal or emulation software to provide the input method services.

Hanyu DECterm and Hanzi DECterm are considered DECwindows Motif applications and are activated and deactivated using the same methods as similar applications. See Section 7.1.2 for more information.

On the LK201-C, LK201-D, LK401-C or LK401-D keyboards, select input mode by using the [Compose] key, which is labeled 中/英. See Chapter 6 for more information on keyboards.

Once the Chinese input mode is activated, the firmware of the terminal or the input methods incorporated in DECterm automatically compose Chinese characters and return the input data as appropriate.

7.1.2 The dxhanziim and dxhanyuim Input Servers and DECwindows Motif Applications

For DECwindows Motif applications, Chinese input methods are implemented in the form of independent processes called input servers. These Chinese input servers are X client processes that can work on a standard X server provided the X server has the required Chinese fonts installed. This means that the Chinese input server can run on any system which can access your X display device, including the device itself.

The operating system provides a traditional Chinese input server (`dxhanyuim`) that is interoperable with all existing DECwindows Motif /Hanyu platforms, including OpenVMS DECwindows Motif /Hanyu and UWS/Hanyu. The operating system also provides a simplified Chinese input server (`dxhanziim`) that is interoperable with all existing DECwindows Motif /Hanzi platforms, including OpenVMS DECwindows Motif /Hanzi and UWS/Hanzi. Both input servers provide input method services to the R6 X library (Xlib) supported by Tru64 UNIX. You can write internationalized applications using the standard R6 application programming interface and communicate with these input servers. For details about developing internationalized software with X11R6, see *Writing Software for the International Market*.

Although the codesets returned by traditional and simplified Chinese input servers are fixed, Tru64 UNIX allows you to connect applications to the input servers using any valid Chinese locale. The operating system provides the required codeset conversion.

Before you can input Chinese data, you must start the appropriate Chinese input server on your workstation or any system on your network that can be accessed by your workstation.

English and Chinese user interfaces are provided, so be sure to set the correct session language before starting the input server. There are several ways to start the `dxhanziim` or `dxhanyuim` Chinese input server:

- Using the Session Manager

  After you log into a session, start the Chinese input server as you would any other DECwindows application and select the Hanyu or Hanzi input method from the Applications menu of the Session Manager.

- Automatic Startup of the Input Server

  If you start up your session in one of the traditional Chinese locales, the Hanyu input method menu item is added to the Session Manager's Automatic Startup list by default. Likewise, if you start up your session in one of the simplified Chinese locales, the Hanzi input method item is added to the Session Manager's Automatic Startup list by default. When you log in, the appropriate input server starts automatically. If you do not want to auto-start the input server, use the Session Manager's Customize menu to remove this item from the Automatic Startup list.

- Using a Command

    To start the input server on a workstation, enter one of the following commands:

    ```
    % /usr/bin/X11/dxhanyuim &
    % /usr/bin/X11/dxhanziim &
    ```

    To start the input server for the Hanyu input method on a remote system, enter the following command on that system:

    ```
    % setenv DISPLAY <display-name>:0
    % /usr/bin/X11/dxhanyuim &
    ```

    To start the input server for the Hanzi input method on a remote system, enter the following command on that system:

    ```
    % setenv DISPLAY <display-name>:0
    % /usr/bin/X11/dxhanziim &
    ```

    In these examples, `<display-name>` is the display name of your workstation.

After you invoke the Chinese input server, DECwindows Motif applications that have been internationalized for Chinese can communicate with it to obtain input method services.

_____ **Note** _____

Applications that are started before the `dxhanziim` or `dxhanyuim` input method servers cannot connect to the input server. Therefore, the input method server should be the first item on the Automatic Startup list. However, if the application contains an `XmText` or `XmTextField` widget with the reconnectable resource set to True, the application is able to establish a connection with the input server if the application starts before the input server or when the application is running and the input server stops and restarts. For more information, see `XmText(3X)` and `XmTextField(3X)`.

_____

7.1.3 The dxhanziim and dxhanyuim Input Servers and CDE Applications

For CDE applications, Chinese input methods are implemented by input servers. Before you can input Chinese data, you must start the appropriate Chinese input server on your workstation or on any system that can be accessed by your workstation through the network.

Both English and Chinese user interfaces are provided. Make sure you set up the correct session language before you start the input server.

There are two ways to start the Chinese input servers in CDE:

- Automatic Startup of the Input Server

    If a traditional or simplified Chinese language is selected on the CDE login menu, the appropriate Chinese input server starts automatically. When you log in, the following script runs:

    ```
    /usr/dt/config/Xsession.d/0020.dtims
    ```

    The value of the DTSTARTIMS environment variable determines whether the script will automatically start the specified Chinese input server.

- Using a Command

    Enter one of the following commands to start the input server on a workstation you are using:

    ```
    % /usr/bin/X11/dxhanyuim &
    ```

    or

    ```
    % /usr/bin/X11/dxhanziim &
    ```

    Enter the following commands on a remote system to start the Hanyu input method server on that system:

    ```
    % setenv DISPLAY <display-name>:0
    % /usr/bin/X11/dxhanyuim &
    ```

    Enter the following commands on a remote system to start the Hanzi input method server on that system:

    ```
    % setenv DISPLAY <display-name>:0
    % /usr/bin/X11/dxhanziim &
    ```

    In these examples, <display-name> is the display name of your workstation.

After you invoke the Chinese input server, the CDE applications that are internationalized to support that specified Chinese language can communicate with it to provide input method services.

_____ **Note** _____

Applications that are started before the `dxhanziim` or `dxhanyuim` input method
servers cannot connect to the input server. Therefore, the Hanyu or Hanzi input method
should be the first item on the Automatic Startup list. However, if the application
contains an `XmText` or `XmTextField` widget with the reconnectable resource set to
True, the application is able to establish a connection with the input server if the
application starts before the input server or when the application is running and the input
server stops and restarts. For more information, see `XmText(3X)` and
`XmTextField(3X)`.

_____

7.1.4 Setting Default Input Method with the i18nconfig Utility

The Configure International Software utility (`i18nconfig`) is an option on the SysMan
menu that you can use to select a default input method for locales, such as simplified or
traditional Chinese, that support multiple input methods. To set a default input method with
`i18nconfig`, you must be logged in as superuser and use the following steps:

- Start the SysMan menu from the CDE front panel or from the command line.

- From the SysMan menu, select *Software*. From the list of tasks under Software, select
  Configure International Software. From the list of tasks under Configure International
  Software, select *Managing Locales*.

- If you select a locale from the Locales Supported On… display that supports multiple
  input methods, the Select Input Method button is active. Click on the Select Input
  Method button to display a list of input methods appropriate for the selected locale.

- Double click on an input method to assign it as the system-wide default for the selected
  locale.

To have the change take effect, log off the system then log back in.

7.1.5 DXIM Input Server

Use the `dxim` input server to activate a traditional or simplified Chinese input method,
which converts alphabetic keyboard entry to the appropriate Chinese characters and sends
them to an X client application.

As with the `dxhanziim` and `dxhanyuim` input servers, the `dxim` input server is an X
client process that runs on a standard X server. The `dxim` input server can run on any system
that can access your X display device.

To run successfully, the dxim input server requires that you perform the following prerequisite steps:

- Install WLS (Worldwide Language Support) options on the system.

- Install Chinese fonts on the system. The fonts should be those associated with the Chinese input methods you intend to use.

- Set the current locale set to traditional or simplified Chinese.

- Stop any other input server process running on the system. Keep in mind that if you use the CDE login language option to set the locale, an associated input server process starts. You must stop this process before starting dxim.

- Set the preediting style, or interaction style, to either Off-the-Spot or Root as described in Section 7.3.2.

Start the dxim input server as a background process with the following command:

```
% /usr/bin/X11/dxim &
```

If you want to start dxim from a remote system, log on to the remote system and enter the following command:

```
% setenv DISPLAY <display-name>:0
% /usr/bin/X11/dxim &
```

In the *<display-name>* field, enter the display name for your workstation.

Section 7.1.5.1 describes how to use the dxim menu interface to select and activate traditional and simplified Chinese input methods and establish a default input method for Chinese applications.

After you activate a Chinese input method, applications that have been internationalized to support that input method can communicate with the server to obtain input method services. You must start these applications after the input server starts.

_____ **Note**_____

Applications that are started before dxim cannot connect to the input server. However, if the application contains an XmText or XmTextField widget with the reconnectable resource set to True, the application is able to establish a connection with the input server if the application starts before the input server or when the application is running and the input server stops and restarts. See XmText(3X) and XmTextField(3X).

_____

7.1.5.1 Using DXIM to Select and Activate Input Methods

The `dxim` input server is a multilingual input server because it gives you the means to use and manage input methods for Korean, as well as traditional and simplified Chinese. This manual focuses on `dxim` and the Chinese input methods. For additional information on the input server, see the `dxim` online help.

The `dxim` input server menu is a single menu system divided into two functional parts; Customizing Input Method Classes and Methods and Customizing Input Method Window. The Customizing Input Method Classes and Methods menus are described in Section 7.5.1 and allow you to do the following:

- Select a class of input methods that is appropriate to the locale of the client application. For an application internationalized for the Chinese language, you select and activate one or more of the following classes: traditional Chinese, simplified Chinese, or Phrase.

- Select and activate one or more input methods within a class. With the exception of the Phrase input method, the traditional and simplified Chinese classes under `dxim` support the same set of input methods as `dxhanziim` and `dxhanyuim`. The Phrase input method is a separate class under `dxim` and uses a different database than that used by the operating system Phrase utility.

- Establish an input method class as the default.

- Establish an input method as the default for its class.

- Customize the simplified Chinese 5-Shape and Intelligent ABC input method classes.

- Customize error bell volume and set the input method invocation key.

The Customizing Input Method Window menus are described in Section 7.5.1.2 and allow you to do the following:

- Increase or decrease the root input window font size.

- Set the root input window foreground and background color.

- Set the root input window line spacing.

The `dxim` input server can support multiple clients working under different locales. When a client application connects to `dxim`, the input server determines the client's locale and, if compatible, uses the default input method. If the client locale is not compatible with the default, `dxim` searches for an active input method that is compatible. The input server uses the first compatible input method it finds.

For example, when you connect a zh_CH.dechanzi client application to dxim, the input server checks the application's locale and determines that either the simplified or traditional Chinese input method is available for use with the application. You then enter the invocation key for the Chinese input method and any keyboard input, which dxim converts and transmits to the application. If you attempt to enter the invocation key for Korean, which is inappropriate for this application, dxim ignores it.

Systemwide default definitions for the dxim input server are stored in the resource file, /usr/lib/X11/app-defaults/DXim, and in the configuration file, /usr/i18n/share/dxim/im_modules.conf. As superuser, the system administrator can modify these files to make systemwide changes. For example, the seek order that dxim uses in its search for an input method compatible with a client's locale is defined in /usr/i18n/share/dxim/im_modules.conf and can be modified by the system administrator.

Depending on the locale, the dxim input server does not restrict an application to a single input method. For example, an application under a Unicode locale, such as zh_CN.UTF-8 can be supported by all of the Korean and Chinese input methods.

7.2 Switching Input Method

By default, a client application starts in English language input mode and, depending on which input server is active on your system (dxhanziim, dxhanyuim, or dxim), you use different key combinations to switch from English language input mode to Chinese or Phrase input mode.

The key sequences in this section can be customized by means of the dxhanziim and dxhanyuim Options menu or the dxim Customize Input Method menu.

If the traditional or simplified Chinese Input Server (dxhanziim or dxhanyuim) is active on your system, select the Chinese input mode using the key sequences shown in Table 7–1.

Table 7–1: Key Sequences that Invoke the traditional or simplified Chinese Input Methods

Terminal or Keyboard Type	Default Key Sequence
DECwindows Motif	
LK201	[Compose/Space]
LK401	[Compose]
PCXAL	[Alt/Space] for traditional Chinese
	[Ctrl/Space] for simplified Chinese

Table 7–2 and Table 7–3 describe the key sequences for selecting a specific traditional Chinese or simplified Chinese input method once you are in Chinese input mode.

Table 7–2: Key Sequences Used to Select dxhanyuim Traditional Chinese Input Methods

Input Method	Default Key Sequence
Full-Form Alphabets	[Shift/Space]
Tsang—Chi（倉頡）	[F6]
Quick Tsang—Chi（倉頡）	[F7]
Internal Code（內碼）	[F8]
Phrase（片語）	[F9]
Phonetic（注音）	[F10]
Symbol Input（符號）	[Z] (in Tsang—Chi or Quick Tsang—Chi mode)

Table 7–3: Key Sequences Used to Select dxhanziim Simplified Chinese Input Methods

Input Method	Default Key Sequence
Intelligent ABC	[F4]
Phrase Input 〔片语〕	[F5]
5-Stroke 〔五笔画〕	[F6]
Qu-Wei 〔国标区位码〕	[F7]
Pin-Yin 〔拼音〕	[F8]
Telex Code 〔电报码〕	[F9]
5-Shape 〔五笔字型〕	[F10]

If the dxim input server is active on your system, use the [Ctrl/Space] key sequence to shift from English language input to non-English language input. When you use [Ctrl/Space] to shift to non-English mode, keyboard character input is transmitted to dxim, which converts the input using a selected input method and transmits the converted input to the client application.

To select a traditional or simplified Chinese input method under dxim, use the appropriate key sequence from Table 7–4. Keep in mind that you do not repeat the invocation key for the input method class if you are switching input methods within that class. For example, use the [F1][F10] key sequence to invoke the 5-Shape input method under simplified Chinese, but to switch from 5-Shape to 5-Stroke within simplified Chinese, use only [F6].

Table 7–4: Key Sequences Used to Select dxim Input Methods

Input Method Class	Input Method	Default Key Sequence
Simplified		[F1]
	Intelligent ABC	[F5]
	Symbol Input （符號）	[F11]
	5-Stroke 〔五笔画〕	[F6]
	Qu-Wei 〔国标区位码〕	[F7]
	Pin-Yin 〔拼音〕	[F8]
	Telex Code 〔电报码〕	[F9]
	5-Shape 〔五笔字型〕	[F10]
Traditional		[F2]
	Tsang—Chi （倉頡）	[F5]
	Quick Tsang—Chi （倉頡）	[F6]
	Symbol Input （符號）	[F8]
	Internal Code （內碼）	[F9]
	Phonetic （注音）	[F7]

To select the Phrase input method, use the [F12] key.

_____ **Note** _____

In standard Motif, the function key [F10] is defined as the accelerator of the pull-down menu bar. In DECwindows Motif, the default menu accelerator is [Ctrl/F10] and [F10] is used to invoke the Phonetic input method. To make your DECwindows Motif applications Motif compliant (and change [F10] from Phonetic input method invocation to the menu accelerator), insert the following line in your $HOME/.Xdefaults file:

```
*menuAccelerator: Ctrl <Key> F10:
```

If you change [F10] to the menu accelerator, you must change the Phonetic input method invocation key to another key sequence to invoke that input method.

_____

7.3 Motif Interface Input Method

You can interact with the Chinese input server through a Motif-style user interface. This interface allows an input method to provide feedback about the data being edited, to help you compose a character, list choices for selection, provide options for customizing the input server, and so on.

7.3.1 Input Areas

The X Input Method specification defines the three input areas shown in Table 7–5.

Table 7–5: Window Input Areas

Region	Description
Auxiliary area	An option menu that helps you customize the Chinese input methods and the input method window.
Status area	Displays critical information about the internal state of the Chinese input methods.
Preedit area	Displays the intermediate text that is being composed. Also displays a list of valid candidates for the input key sequences.

7.3.2 Interaction Styles

The use of the input areas depends on the interaction style (or preedit style) selected for the application. The Chinese input method servers enable you to type multiple keystrokes to compose Asian characters. The process of composing characters from keystrokes is called preediting and the type of preediting used is called the interaction style.

The Chinese input server supports two interaction styles:

- Root window

- Off-the-Spot

You can use the `VendorShell` resource `XmNpreeditType` to specify the priority of the interaction styles for DECwindows Motif applications. By default, the resource value priority order is set to "overthespot,offthespot,root,onthespot". The first style in the list is used in an input method. If that interaction style is not available, the second style in the list is use, and so on.

To set a preferred interaction style, you can do one of the following:

- Use the command line option `-xrm` to specify the resource value when you start an application. For example, the following command starts Cardfiler with the Root window interaction style:

  ```
  % dxcardfiler -xrm '*preeditType: root' &
  ```

- Use the Session Manager's Options menu in XDM.

  From the Session Manager's Options menu, select Input Method...

  In the popup Input Style Options window, click on one of the preedit styles

  The `XmNpreeditType` resource is set to a priority list beginning with the preedit style that you have chosen.

- Use the `dtimsstart` utility.

  From the command line, enter the following command:

  `% /usr/dt/bin/dtimsstart`

  From the `dtimsstart` dialog box, click on an interaction style.

  Click on the Default Input Method radio button.

  Click on OK. The preferences you set remain in effect until you change them.

After you choose your preferred interaction style, the applications you invoke start up with the new setting.

Keep in mind that the `dxim` input server supports only Root window and Off-the-Spot interaction styles. If you use an unsupported interaction style, client applications will be unable to connect to the `dxim` input server.

_____ **Note** _____

Some applications, such as DECterm, may provide their own user interface to handle interaction styles. Those mechanisms may override the methods described here.
_____

7.3.2.1 Root Window Interaction

Choose the Root window interaction style (Figure 7–1) if you want to display the preedit data in an input window that is separate from the application window. You can scale and move the input window to meet your preferences. If you want to free up more screen space, you can iconize the input method window. You can also choose to display preedit data in vertical or horizontal layout.

Figure 7–1: Chinese Root Window Interaction Style
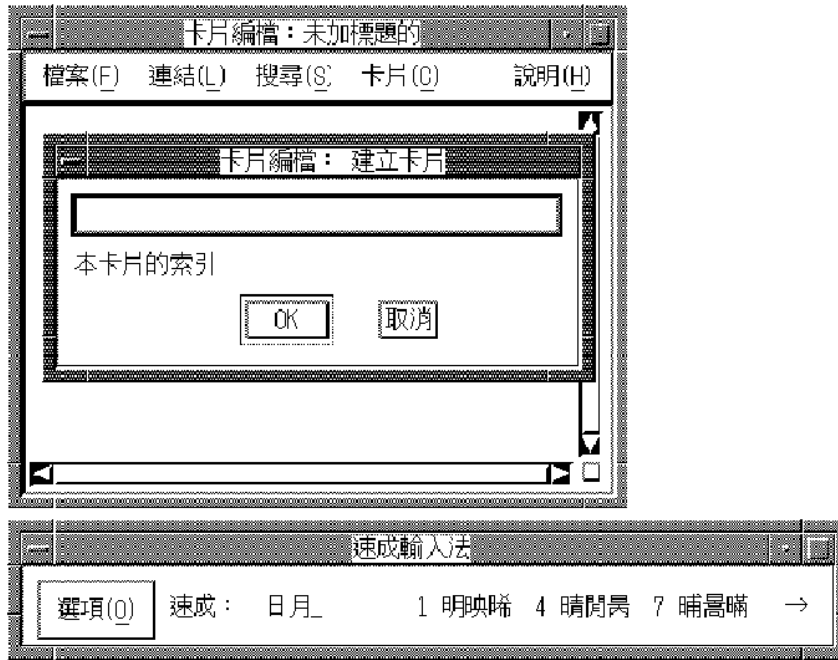
Figure 7–2: Chinese Input Window Icon
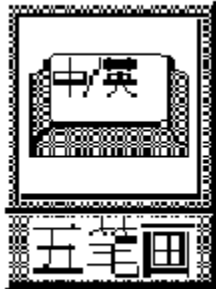
You can continue to input Chinese characters through a Chinese application window when the input window is iconized (Figure 7–2). The input state is displayed on the icon title, which is updated according to the input mode and the input focus. If you want to see the preedit data, you can double click the icon to redisplay the input window.

7.3.2.2 Off-the-Spot Interaction

To display the preedit data in a fixed location of the application window, choose the Off-the-Spot interaction style (Figure 7–3). With this interaction style, the Chinese input server creates the input window at the bottom of the application window. You need not refer to the Root window and you can iconize it to save screen area.

Figure 7–3: Off-the-Spot Interaction Style
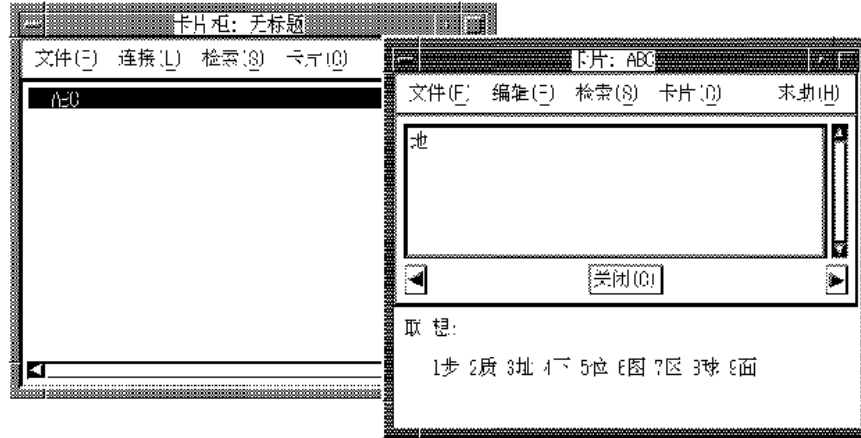
7.4 Operations of the Chinese Input Servers

When you start a Chinese input server, no application is connected to it and the title bar displays "No Connection" in the status area.

If an application is internationalized, started in a Chinese locale, and brought to the foreground, the input state displayed in the status area and the title bar are updated accordingly. If you bring a noninternationalized application window to the foreground, the title bar of the input window changes to indicate there is no connection. (In some environments, bringing an application to the foreground is known as bringing it into focus.)

The input server can maintain an individual state of composition for different input contexts or application windows.

In addition, under the dxhanziim and dxhanyuim input servers running the Root window interaction style, each application window can be associated with its own attributes, such as font size, font style, layout, input window size, and position. You can set the input focus to an application window and then compose a Chinese character or customize the input window.

The input server retains the composing state and input window attributes. Thus, when you bring an application window into focus (that is, from the background to the foreground), the input server restores its original composing state and attributes. For example, if you are using an input window under the 5-Stroke input method, then shift focus to an input window under the Pin-Yin input method, when you bring the original window back in focus, the 5-Stroke input method will be active.

7.5 Customizing the Chinese Input Servers

The `dxim`, `dxhanziim`, and `dxhanyuim` input servers have different levels and different methods of customization.

Under the `dxim` input server, the input window provides an options menu containing the following selections:

- Customize Class that displays a menu interface you use to customize the input methods

- Customize Window that displays a menu interface you use to customize the `dxim` input server window

- Help that displays online help for the `dxim` input server

- Exit

Under the `dxhanziim` and `dxhanyuim` input servers, the auxiliary area of the input window provides an options menu where you can customize the input server. You click on the Options button to view the customization pulldown menu.

The following sections describe the `dxim` and `dxhanziim`/`dxhanyuim` customization menus. Keep in mind that each of these input servers has an extensive online help system that can provide you with details and examples of each option.

7.5.1 Customizing the dxim Input Server

If you select Customize Class from the `dxim` options menu, it displays the `dxim` Input Method Class and Module Customization menu. This series of menus enables you to customize the following attributes:

- Activate or deactivate one or more of the available input servers. The input servers you can activate include `dxim`, `dxhanziim`, `dxhanyuim`, Phrase, and Korean. The Korean input server is described in the Tru64 UNIX *Technical Reference for Using Korean Features*. The other input servers are discussed in this chapter.

- For each active input server, activate or deactivate one or more input methods. The Chinese input servers support multiple input methods. For the Phrase input server, the phrase input method is the only selectable input method.

- For each active input server and input server method, you can establish a default input server for the system. You can also establish a default input server method that becomes active when its associated input server starts.

- If the selected simplified Chinese input server method is 5-Stroke, Pin-Yin, Qu-Wei, or Telex, you can click on the Customize button to activate the 5-Shape Radicals display. With the 5-Shape Radicals display active, you can enter a non-5-Shape character and view the correct 5-Shape Radical character required to produce the non-5-Shape character. That is, you can use the 5-Shape Radical display as a training tool.

- If the selected simplified Chinese input server method is Intelligent ABC (ItABC), you can click on the Customize button to activate an ItABC Options dialog box. The dialog box allows you to select options for repetitive input or defining and saving words and phrases.

- For each active input server and input server method, you can redefine the default invocation key.

If you select Customize Window from the `dxim` options menu, it displays the `dxim` Input Method Window Customization menu. This series of menus enables you to customize the following attributes:

- Set font size for root preedit data.

- Set the foreground and background color.

- Set the line spacing for the preedit text lines.

If you log in as superuser to use the `dxim` input server customization menus, the changes you make are written to `/.dxim/DXim` and apply to root. If you log in as a non-root user, the changes you make are written to `~user/.dxim/DXim` and apply to that user ID.

The following sections provide an overview of these menus. For a complete description of the menu options, see the `dxim` online help system.

7.5.1.1 Input Method Class and Module Customization

The Input Method Class and Module Customization menus enable you to activate, deactivate, and customize the simplified and traditional Chinese and Phrase input method classes and input methods within those classes. Keep in mind that you must use either the OK or Apply button in conjunction with the Save Current Settings pull-down menu option to save a customized setting for use by the `dxim` input server.

The Customize Class menu is the primary input method customization menu. The following table briefly describes each of the Customize Class menu options. For a full description of the menu options, see the `dxim` input server online help.

Menu Field	Description
Active Class	Displays the input method classes that are currently active on the system. Select one or more of these classes to customize or delete.
Invocation Key	Displays the key sequence used to invoke the selected input method class. Use the checkboxes under the display to change the invocation key.
Bell Volume Slider	A bell rings when an error is made during character composition. Drag the slider to adjust the volume of the bell.
Options pull-down menu	Select Get System Default to reset the Customize Class menu.
	Select Save Current Settings to save values for use by the dxim input server.
Add…	Opens a dialog box that allows you to select and activate a currently inactive input method class.
Delete	Deactivates a selected input method class. The class and its methods are made inactive and not available to applications.
Customize…	Opens a dialog box that allows you to activate, deactivate, and customize the input method modules associated with the simplified Chinese input methods. The Customize Input Method dialog box selections are similar to the Customize Class menu selections described in this table.
Set Default	Establishes the selected input method class as the default.
OK	Saves the customizations to become effective the next time you start dxim. Closes the menu.
Apply	Same as OK, except that the menu does not close.
Cancel	Closes the menu without making any changes.
Help	Displays help on this menu.

7.5.1.2 Input Method Window Customization

The Input Method Window Customization menus enable you to modify the `dxim` root input window characteristics. Keep in mind that you must use either the OK or Apply button in conjunction with the Save Settings as Defaults pull-down menu option to save a customized setting for use by the `dxim` input server.

The Customize Window menu is the primary input window customization menu. The following table briefly describes each of the Customize Window menu options. For a full description of the menu options, see the `dxim` input server online help.

Menu Field	Description
Big Font/Small Font	Select the font size for displaying preedit data in the `dxim` root input window.
Foreground Color/Background Color	Select either foreground or background. A color mixing dialog box opens that enables you to customize the foreground or background colors of the root input window.
Line Spacing Slider	Drag the slider to specify the line spacing, in pixels, between the preedit data text lines.
Options pull-down menu	Select Restore System Defaults to reset the Customize Window menu.
	Select Save Settings as Defaults to save values for use by the `dxim` input server.
OK	Saves the customizations to become effective the next time you start `dxim`. Closes the menu.
Apply	Same as OK, except that the menu does not close.
Cancel	Closes the menu without making any changes.
Help	Displays help on this menu.

7.5.2 Customizing the dxhanziim and dxhanyuim Input Servers

The `dxhanziim` and `dxhanyuim` input server options menu enables you to customize the following attributes:

- Vertical Layout (Section 7.5.2.1)

- Horizontal Layout (Section 7.5.2.2)

- Select Phrase Input Class (Section 7.5.2.3)

- User Phrase Database (Section 7.5.2.4)

- System Phrase Database (Section 7.5.2.5)

- Current Window (Section 7.5.2.6)

- Input Method Customization (Section 7.5.2.7)

- Help (Section 7.5.2.8)

- Quit (Section 7.5.2.9)

The following sections describe each of these menu options.

7.5.2.1 Vertical Layout

You can choose the Vertical Layout option only if the current layout is horizontal. When you choose this option, the input window and the layout of its contents immediately display in a vertical orientation. The vertical input window remains at the same origin.

7.5.2.2 Horizontal Layout

You can choose this option only if the current layout is vertical. When you choose this option, the input window and the layout of its contents immediately display in a horizontal orientation. The horizontal input window remains at the same origin.

7.5.2.3 Select Phrase Input Class

You can use the Select Phrase option to customize the Phrase Input Mode. DECwindows Motif shares the phrase databases that are created and managed by the base operating system Phrase Utility. After you create a phrase database and define your phrases, both character-cell terminal applications and DECwindows Motif applications can use the data for phrase input.

In order to use the phrase databases, the LANG environment variable must be set to reflect the required codeset (for example, `zh_TW.dechanyu`). For details about the Phrase Utility and phrase definition file, see *Using International Software* in the operating system documentation set.

The Select Phrase Input Class option allows you to focus on a particular class of phrases during phrase input. When you choose this option, a dialog box pops up and you can select the phrase class that you want to use.

To select all classes, you choose the * (wildcard) option. If you choose this option, the phrase input method searches all classes of phrase definitions for the phrase code that you entered.

When there is more than one phrase that matches the phrase definition, a list box is displayed. To choose the right phrase definition, select the appropriate item in the list box.

The dxim input server uses a different Phrase input method and phrase database than the operating system Phrase Utility and phrase databases described in this section. See Section 7.11.1 for a description of the dxim Phrase input method.

_____

7.5.2.4 User Phrase Database

Use the User Phrase Database option to access your private user phrase database.

The Phrase input method allows you to access two phrase definition databases: the system phrase database and the private user phrase database. You create and maintain your own private phrase database for storing your frequently used phrases. This is called the user phrase database.

For details about creating a phrase database, see *Using International Software*.

7.5.2.5 System Phrase Database

The system phrase database is for public access by all users using your system. It should be created and modified by your system administrator. Use the System Phrase Database option to access the system phrase database.

_____ **Note**_____

The databases that you can access are the ones available on the system on which you start your Chinese input server.

_____

7.5.2.6 Current Window

The Current Window option allows you to customize the attributes of a specific application window.

_____ **Note**_____

The Current Window option is available only if you chose Root window as your interaction style and an internationalized application input area is in the foreground. Otherwise, this option is dimmed out. If you choose the Off-the-Spot interaction style, the application determines the attributes.

_____

When you choose this option, a dialog box pops up and the following options are displayed:

- Font Size

  Choose the font size for displaying preedit data. Click on either the Big Font or Small Font toggle buttons.

- Font Typeface

  Choose the font typeface to be used in the input window. To choose the font typeface in a traditional Chinese input server, click on one of the following toggle buttons:

    - Hei

    - Sung

    - Screen

  To choose the font typeface in a simplified Chinese input server, click on one of the following toggle buttons:

    - Heiti

    - Songti

    - Kaiti

    - FangSongti

    - Screen

  You can define a typeface that does not exist in the options list in the Chinese input server resource file. This typeface will be displayed beside the Other: label in the customization window.

- Line Spacing

  The Chinese input server can display preedit data on more than one text line. Usually, this happens when a list of items is displayed for your selection. You use the Line Spacing option to specify the spacing between text lines in pixels. To adjust the line spacing, drag the Line Spacing slider or move the pointer to the desired position on the slider and click MB1.

- Foreground and Background Colors

  You can customize the foreground and background colors of the input window.

  For monochrome display, the following options are provided:

    - Dark Text, Light Background

    - Light Text, Dark Background

For color display, you can choose from a palette of colors to design a visually pleasing input window. To customize the foreground or background color, first select the color that you want to change by clicking one of the following toggle buttons:

- Input Window Foreground Color

- Input Window Background Color

A color mixing window pops up in which you can mix the color using the three sliders, which represent the intensities of the primary colors. The modified color is displayed in the right half of the color box while the left half displays the original color.

7.5.2.7 Input Method Customization

When you choose the Input Method Customization option, a dialog box displays in which you can customize the following attributes.

In the traditional Chinese input server, the customizable attributes are:

- Default Input Method

- Bell Volume

- EDPC support

- Invocation Key

In the simplified Chinese input server, the customizable attributes are:

- Default Input Method

- Active Input Method

- Bell Volume

- Display 5-Shape Radicals

- Word Frequency and Bixing Input – Intelligent ABC input method only

- Invocation Key

The following list describes each of the customizable attributes:

- Default Input Method

  Use the Default Input Method option to choose the Chinese input method invoked by default when you activate the Chinese input server. The input methods you can choose in the traditional Chinese input server (dxhanyuim) are:

  - Alphabetic

  - Tsang–chi

  - Quick Tsang–Chi

- Phonetic

- Internal code

- Phrase

The input methods you can choose in the simplified Chinese input server (`dxhanziim`) are:

- Alphabetic

- Phrase

- 5-Stroke

- Qu-Wei (Row-Column)

- Pin-Yin

- Telex Code

- 5-Shape

- Intelligent ABC

- Active Input Method

  The Active Input Method option allows you to decide which input methods can be activated at run time. By default, all input methods, except for the simplified Chinese Telex input method, are active.

  You use the Active Input Method list to add or remove an input method on the Active list. If you remove an input method from the Active list, you can no longer select that input method with an invocation key.

- Bell Volume

  When you make an error while composing a Chinese character, the bell rings to alert you. To adjust the bell volume, drag the Bell Volume slider or move the pointer to the desired position on the slider and click MB1.

- EDPC support (in `dxhanyuim` only)

  The traditional Chinese input server supports the input of both CNS 11643 and DTSCS (that is, EDPC) characters. However, you can choose to disable the input of EDPC characters, thus ensuring that the data that you enter contains only CNS 11643 characters. This option is useful if you need to prepare Chinese data and interchange it with systems supporting only CNS 11643.

  To enable or disable the input of EDPC characters, click on the EDPC Characters Input button.

- Display 5-Shape Radicals (in `dxhanziim` only)

  If you choose the Input Method Customization option while running the 5-Stroke, Pin-Yin, Qu-Wei, or Telex input methods on the simplified Chinese input server, a dialog box is displayed in which you can choose the Display 5-Shape Radicals option.

  When you choose the Display 5-Shape Radicals option, the character candidates displayed in the 5-Stroke, Pin-Yin, Qu-Wei, or Telex input methods are associated with the 5-Shape key sequences required to enter those characters in the 5-Shape input method.

  You can use this option as an instructional tool. For example, if this option is active and you enter a character using the Pin-Yin input method, the resulting display provides instruction on the 5-Shape Radical sequence required to produce that same Pin-Yin character in the 5-Shape input method.

- Word Frequency and Bixing Input – Intelligent ABC input method only

  If you choose the Input Method Customization option while running the Intelligent ABC input method on the simplified Chinese input server, a pull-down menu is displayed in which you can automatically adjust word frequency in repeated selections and specify bixing input in repeated selections.

  These property settings are described in 7.19.3.

- Invocation Key

  The key sequences for invoking and switching Chinese input methods are set by default. You can change these default key sequences to meet your personal preference or working style. This option allows you to customize the following key sequences for both `dxhanyuim` and `dxhanziim`:

  - Start Input Method

  - End Input Method

  - Phrase Input

  - Invoke Next Input Method

  - Full and Half Form Alphabetic input method

  The following choices are for `dxhanyuim` only:

  - Tsang-chi

  - Quick Tsang-chi

  - Internal Code

  - Phonetic

The following choices are for `dxhanziim` only:

- 5-Stroke

- Qu-Wei (Row-Column)

- Pin-Yin

- Telex Code

- 5-Shape

- Intelligent ABC

You customize key sequences using the interface in the bottom part of the dialog box. Select a trigger key and toggle the on/off state of the Ctrl, Alt, and Shift modifiers. The trigger keys that you can choose include NoSymbol, [F1] through [F20], [Space], [Return], [Compose] and [A] through [Z]. If you choose NoSymbol, no invocation sequence will be provided for the selected action.

For each modifier key, you can select the on/off state with the toggle buttons described in Table 7–6.

Table 7–6: Modifier State Customization

Modifier	On State	Off State
Ctrl	Ctrl	~Ctrl
Shift	Shift	~Shift
Alt	Alt	~Alt

The tilde (~) sign means that you should not press that modifier key when invoking the action. In addition to the on/off state, you can also deselect both of the states for a modifier key so that neither state is selected. To do this, click the toggle button that is currently set on. If you deselect a modifier, the input server will accept the invocation key with or without holding the modifier key.

When you select an invocation key sequence, the state of the toggle switches and the trigger key displayed at the bottom of the dialog box is updated to reflect the current value and the label at the bottom left-hand side changes.

Figure 7–4: Customization of Invocation Key Sequences in
`dxhanyuim`

Figure 7–5: Customization of Invocation Key Sequences in
`dxhanziim`
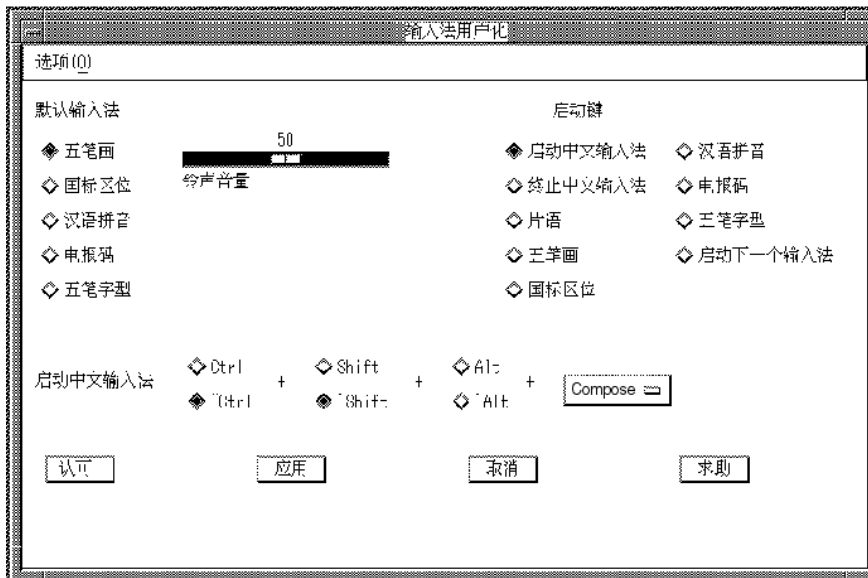
For example, if you want to change the End Input Method key sequence to [Ctrl/Space], select the Ctrl, ~Alt and ~Shift buttons.

To reduce the number of keys required for selecting input methods, set the hot keys for all the input methods to "NoSymbol", and define a key sequence for the Choose Next Input Method option. This method releases the [F6] to [F10] function keys for use by other DECwindows applications. To switch the input method, press the key sequence for Choose Next Input Method and the Chinese input server will cycle through all supported input methods. For example, if you use the Tsang–Chi input method and you want to switch to the Internal Code input method, press the hot key twice.

7.5.2.8 Help

The Help option provides the following menu items that you use to display help for the Chinese input server:

- Context-Sensitive Help

- Overview

- Using Help

- Product Information

7.5.2.9 Quit

Use the Quit option to terminate the input server. If you select this option, a dialog box pops up and asks if you really want to exit.

7.5.2.10 Saving Your New Settings

All attributes that you customize with the Current Window and Input Method Customization menus can be saved into a resource file in your login directory. Each customization window provides the following options:

- Save Settings as Defaults

  Saves all current attributes as default values. These attributes are saved to a private resource file in your login directory; `.DXhanziim` if you are using `dxhanziim`, and `.DXhanyuim` if you are using `dxhanyuim`.

- Restore system setting

  Restores all system default attributes

7.6 Alphabetic Input Methods

There are two alphabetic input methods available under the English language mode; Half Form Alphabet and Full Form Alphabet.

The Half Form Alphabet input method allows you to enter uppercase and lowercase English characters, numerals, and symbols marked on the keyboard. Full Form Alphabet input method allows you to enter 2-byte alphabets, numerals, and symbols defined in the Chinese character sets.

To invoke the Full Form Alphabet input method, press [Shift/Space]. The string " 全 形" (full form) is displayed in the status area, as shown in Figure 7–6. Once the prompt appears, all characters that you enter at the keyboard are sent as 2-byte characters.

To exit the Full Form Alphabet input method, press [Shift/Space] again.

Figure 7–6: Full Form Alphabet Input Method

ＡＢＡＢ＿
〔全形〕

7.7 Tsang–Chi Input Method

To understand the Tsang–Chi input method, you must understand the concepts of Tsang–Chi root radicals, auxiliary forms, and character-splitting.

7.7.1 Tsang–Chi Root Radicals

The Tsang–Chi input method is based on the concept of root radicals. The input method requires a Chinese character to be broken down into various root radicals according to the character's shape.

Altogether 24 Tsang–Chi root radicals have been defined from which almost all existing Chinese characters can be composed. The root radicals are divided into four groups and assigned to the alphabet keys [A] - [Y] on the main keyboard (the [X] and [Z] keys are not assigned). The following tables illustrate the classification of the root radicals, their corresponding English keys, their auxiliary forms, and the way that they are derived.

- Philosophical classification – Table 7-7a

- Stroke classification – Table 7-7b

- Human Body classification – Table 7-7c

- Form classification – Table 7-7d

Table 7–8 is a quick reference to all of the root radical classification.

Table 7–7: Tsang–Chi root Radicals Classification

a.　Philosophical:

類別	字根 (口訣)	英文鍵	輔助字形	聯想方式	例子
哲	日	A	囗、日	取日的變形。	巴、眉、冒、是
	月	B	冂、冖 爪、夕	取月的外形。 取月的變形。	同、巾、冗、罕 采、受、炙、然
	金	C	⺍ 八、儿	取金的部分形。 取⺍的變形。	弟、丫、曾 兇、父、匹
理	木	D	寸 扌朩 朮	取木的主形。 取寸的變形。 取木的主形。	寸、才、干、子 也、虔、五、夫 虎
	水	E	氵 又 氺	取水做偏旁之形。 取水的左右重合之形。 取水的變形。	江、河、海、溪 支、叟、冗、叔 柔、泰、隸、米
類	火	F	灬 ⺌ 小	取火做字尾之形。 取火的上半形。 取⺌的變形。	杰、然、無、為 尚、肖、半、少 糸、不、少、紀
	土	G	士 士	取土做偏旁之形。 取土的變形。	坎、坷、坦、地 吉、吾、志、壬

b. Stroke:

類別	字根(口訣)	英文鍵	輔助字形	聯想方式	例子
筆 畫 類	丿(斜)	H	丿 厂	斜的定義為丿。取丿的變形。	幺、禾、天、八、反、派
	丶(點)	I	丶 厶、广	點的定義為丶。取丶的變形。	犬、之、雨、大、台、公、店、庫
	十(交)	J	亠	取十的變形。	言、完、案、安、實、黃、賣、奮
	大(文)	K	乂 尸 疒	乂的定義義乂。取大的部分形。取疒的變形。	爻、刈、艾、交、尤、友、有、左、尤、痛、痕、深、戎、痛
	中(縱)	L	串 市	縱的定義為丨，即由上而下。丨亦為縱向，意義相近。取丨的變形。	川、片、介、索、書、庚、畫、杉、初、褲、被
	一(橫)	M	二、厂 厂、乀	皆取一的變形。	承、皿、瓜、原、无、石、刁
	弓(鈎)	N	亅 →、丷 乁、乙	鈎的定義為亅。取亅的變形。取→的變形。取乙的變形。	了、丁、小、可、乙、足、业、子、角、凰、久、朵、乞、几、風、兀、朵

c. Human Body:

類別	字根(口訣)	英文鍵	輔助字形	聯想方式	例子
人 身 類	人	O	亻 厶、入 八、乁 乀、乚	取人做偏旁之形。取人做字首之形。取人的部分形。取人的變形。	仁、什、仃、佰、秦、歐、佥、内、八、禾、子、之家、余、兆、率
	心	P	忄 小、勹 匕、乚	取心做偏旁之形。取心做字尾之形。取心的變形。取匕的變形。	忙、怪、恰、恭、慈、添、勻、官、比、旬、勺、氏、屯、切、弋、民
	手	Q	扌 丰 扌、牛	取手做偏旁之形。取手的主形。取丰的變形。	扣、扎、扒、把、牛、毛、氈、牟、夫、朵、年、牮
	口	R			閏、吧、邑、員

d. Form:

類別	字根 (口訣)	英文鍵	輔助字形	聯 想 方 式	例 子
字形系	尸 (側)	S	ㄱ、匚 ㄱ、ㅌ ㅏ	側的定義為左右向開口，所以為ㄱ、匚。取ㄱ、匚的變形。取日的部份形。	尹、弓、臣、巨 區、西、臣、珏 力、習、長、非、聰
	廿 (並)	T	廿 卅、艹 屮	並的定義為兩形相並，所以取廿。取廿的變形。	革、卉、苗、曲 昔、共、前、羊 冊、盧、皿、並
	山 (仰)	U	凵 屮 乚、乀	仰之定義為向上開口，故為凵。取凵的變形。取凵的變形。	凼、函、屬、匃 退、舅、乚、凱、氏、氏
	女 (紐)	V	〈 乚、乀	紐的定義為曲紐，所以為〈。取〈的變形。	幻、巡、臘、玄 亡、衣、長
	囝 (方)	W	囗 凸	方的定義為一方框，內含其他形，所以為囗。取囗的變形。	困、因、因、國 四、頁、西、莒 母、冊、毋、貫
	卜 (卜)	Y	一、卜 ＝、辶 ⼍	取卜的變形。取一的變形。取⼀的變形。	市、下、上、卓 卒、斗、斗、迷 冷、測

7–32 Tru64 UNIX Technical Reference for Using Chinese Features

Table 7–8: Quick Reference Table of the Tsang–Chi Root Radicals

哲 理 類			筆 畫 類			人 身 類			字 形 類		
日	四	曰	竹	ノ	厂	人	﹑	㇏	尸	㇈	匸	㇆
										巨	㇄	
	A		(斜)	H		㇒人	O		(側)	S		
月	門	宀	戈	﹑	厶	心	匕	弋	廿	卄	丷	屮
	冖	夕		广			七	勹		屮	业	屮
	B		(點)	I			P		(延)	T		
金	⼆	八	十	宀		手	丰	丰	山	凵	屮	
	儿						扌			𠃊	﹨	
	C		(文)	J		扌	Q		(仰)	U		
木	十	㇆	大	乂	宀	口			女	く	𠃊	
	朮			广						㐄		
	D		(乂)	K			R		(紐)	V		
水	又	氺	中	丨	丰				田	口	昷	
				礻								
⺡	E		(縱)	L					(方)	W		
火	屮	小	一	工	厂				卜	亠	㇉	辶
				𠃌	㇗					﹑	⺀	
灬	F		(橫)	M						Y		
土	士		弓	㇕	㇇	屮						
				𠃌	乙							
⼟	G		(鉤)	N								

7.7.2 Tsang–Chi Code Generation

To input a Chinese character using the Tsang–Chi input method, you must generate the character's Tsang–Chi code based on character decomposition. Most Chinese characters can be divided into two categories: the composite form (Figure 7–9) and the connected form (Figure 7–10). The composite form can be split into a character head and a character body while the connected form cannot.

Table 7–9: Composition Form Characters

Composite Form	Examples
Left-right form	針、 憶、 轉、 謝
Top-bottom form	哲、 靈
Inclusion form	國、 圓

Table 7–10: Connected Form Characters

Connected Form	Examples
Connected form	亞、 兩、 爾

7.7.2.1 General Rules

The general rules for generating Tsang–Chi codes are:

* The character category must be composite or connected.

* The code according to the writing order is usually one of the following:

    — From outside to inside

    — From top to bottom

    — From left to right

* For composite character decomposition, the maximum number of radicals is five. The character head can be decomposed into a maximum of two radicals and the character body can be decomposed into a maximum of three radicals.

    For connected character decomposition, the maximum number of radicals is four.

* If more than one Tsang–Chi code exists for a character, enter the one with fewer radicals.

- If several Tsang–Chi codes have the same number of radicals, enter the one that better represents the character.

7.7.2.2 Connected Characters

Connected characters are those which cannot be split due to the existence of crossed or connected strokes, such as 武, 面, 貞. Each character can be input by entering at most four radicals. If more than four radicals can be derived, enter the first three radicals and the last to generate the Tsang–Chi code. Table 7–11 illustrates some examples of decomposing connected characters.

Table 7–11: Examples of Connected Character Decomposition

字例	拆字	取碼	字首	取碼	字身	取碼
七	卜女	YV	卜	Y	女	V
也	心木	PD	心	P	木	D
凹	尸尸山	SSU	尸	S	尸山	SU
勿	心竹竹	PHH	心	P	竹竹	HH
斤	竹一中	HML	竹	H	一中	ML
斥	竹一卜	HMY	竹	H	一卜	MY
商	卜金月口	YCBR	卜	Y	金月口	CBR
世	心廿	PT	心	P	廿	T
某	廿一一女	TMMV	廿	T	一一女	MMV
弟	金弓中竹	CNLH	金	C	弓中竹	NLH
靈	一口口土	MRRG	一	M	口口土	RRG
岡	月金口	BCR	月	B	金口	CR

7.7.2.3 Composite Characters

Composite characters are those that can be split from top to bottom, left to right, and outside to inside, such as 思, 珠, 圓. You can decompose the character head into one to two radicals. If more than two radicals are generated, enter the first and the last radicals.

The character body can be decomposed into one to three radicals. If it is made up of three or fewer radicals, you should enter all the radicals. If it is made up of more than three radicals that are connected, enter the first two radicals and the last.

If the character body is itself a composite character, you can further decompose the character body into subhead and subbody. Enter the first and last radicals of the subhead, and the last radical of the subbody.

Table 7–12 illustrates examples of decomposing composite characters. In the "Shape" column, a solid square represents a character head while a square represents a radical of the character body.

Table 7–12: Composite Character Decomposition

外觀	字例	字首	取碼	字身	取碼
■□	明計	日金	A C	月十	B J
■▯	卦讀	土上卜口	GG YR	卜土由金	Y GWC
▮▯	頎	人人	OO	一月金	MBC
■■□	黥觸	女木凸月	VD LB	大尸弓中	KS NL
▪▪□	剔	卜人	YO	中弓	LN
■□□	川棚	中手	L Q	中中廿山月	LL TUB
▪□□□	汌	水	E	中中中	LLL
■（上）▯	昌占	日卜	A Y	日口	A R
■■（上）▯	裁祭	土戈竹人	GI HO	口日	R A
■■■（上）▯	壘	女女	VV	田木	WD
▪（上）□	查奎	土大	G K	口廿土土	RT GG
▪▪（上）▯	叄	弓人	HO	一口廿	NRT
回 / 回	國醫區	田尸尸	W S S	戈中一 一口含 口口口	IRM LMC RRR

(左側分類：並列形、上下形、包含形)

7.7.2.4 Exceptional Characters

Approximately 95 percent of Chinese characters can be decomposed according to the rules described in Section 7.7.2.1. The remaining 5 percent are exceptional characters that need to be entered in different ways. The exceptional characters can be divided into the following groups:

• Compound Characters

  The Tsang–Chi input method has defined nine compound characters. A compound character can be a connected character, or the character head or body of a composite character. In any case, compound characters must be represented by their first and last radicals. Table 7–13 lists compound characters.

Table 7–13: Compound Characters

複合字	拆字	取碼	舉例	拆　字	取碼	備　　　註
門	日弓	AN	開 們	日弓尸口十 人日弓	ANSRJ OAN	門當字首使用，兩碼。 門當字身使用只須兩碼
目	月山	BU	睛 盲	月山手一月 卜女月山	BUQMB YVBJ	目當字首使用。 目當字身使用。
鬼	竹戈	HI	魁 塊	竹戈卜十 竹日竹戈	HIYJ HAHI	鬼當字首使用。 鬼當字身，只須取二碼。
鬥	中弓	LN	鬧 鬨	中弓卜中月 中弓廿金	LNYLB LNTC	鬥當字首使用，取首字形"丨"及尾字形" "。
阝	弓中	NL	降 那	弓中竹水手 一世弓中	NLHEQ MTNL	阝當字首使用。 阝當字身，也是取兩碼
隹	人土	OG	集 雕	人土木 心山人土	OGD PUOG	隹當字首使用。 隹當字身只取首尾兩碼
虎	卜心	YP	虐 虛	卜心尸一 口卜心一	YPSM RYPM	虎不分割，當字首使用取'卜'及"七"。
幾	女戈	VI	幾 機	女戈田 木女戈人	VIW DVIO	"幾"字形，視作與下面字形分離，當字首。
羸	卜弓	YN	贏 羸	卜弓女 卜弓廿手	YNV YNTQ	"羸"字形，整個當字首使用。

- Difficult Characters

Difficult Characters are those which are difficult to decompose in the Tsang–Chi input method. Usually, these characters are composed of some special root radicals that are neither the Tsang–Chi root radicals nor their auxiliary forms. Using the Tsang–Chi input method you can press the [X] key (which is labeled with 難 and therefore will be referred to as the [難] in this document) to access the special root radicals and use them to compose difficult characters. Table 7–14 lists difficult characters.

The rules of decomposing difficult characters are:

— If it is easy to identify the first and the last radicals, enter the first radical, the [難] key, and the last radical.

For example, " 身" can be decomposed into " 竹難竹" (HXH).

— If the first radical is easy to identify while the others are difficult, enter the first radical and then press the [難] key for the rest.

For example, " 齊" can be decomposed into " 卜難" (YX), and " 臼" can be decomposed into " 竹難" (HX).

— Never use the [難] key for the first radical.

Table 7–14: Difficult Characters

舉例	拆字	取碼	舉例	拆字	取碼
身	竹 難 竹	HXH	鷯	弓 火 竹 難 竹	NFHXH
鴈	戈 難 火	IXF	�	廿 戈 難 火	TIXF
慶	戈 難 水	IXE	纊	廿 手 戈 難 水	TQIXE
鹿	戈 難 心	IXP	姊	女 中 難 竹	VLXH
龜	弓 難 山	NXU	淵	水 中 難 中	ELXL
兼	廿 難 金	TXC	圙	中 弓 弓 難 山	LNNXU
黽	口 難 山	RXU	謙	卜 口 廿 難 金	YRTXC
臼	竹 難	HX	蠅	中 戈 口 難 山	LIRXU
肅	中 難	LX	兒	竹 難 竹 山	HXHU
齊	卜 難	YX	蕭	廿 中 難	TLX
卍	弓 難	NX	濟	水 卜 難	EYX

註：凡是字的外形與臼字相近者，其拆字方式均相同，即只
取臼左上部的 ㇇ (竹)，其餘均以〈難〉鍵代替。例如：

興 ----→	竹 難 月 金	HXBC	與 ----→	竹 難 卜 金	HXYC
興 ----→	竹 難 十 金	HXJC	叟 ----→	竹 難 中 水	HXLE
鹽 ----→	竹 難 月 廿	HXBT	肅 ----→	一 十 竹 難	MJHX

- Special Characters

  Some special characters are composed by superimposing the root radicals " 木 ," " 大, and " 火" on other strokes or radicals. To keep the decomposed radicals as simple as possible, enter the root radicals first, before the rest of the character is entered. Table 7–15 lists special characters.

Table 7–15: Special Characters

特殊字	拆字	取　碼	取形順序	説　　　明
東	木田	DW	木田	先取木，再取相疊的田
束	木中	DL	木中	"中"字形較口完整。
柬	木田火	DWF	木口小	先取木，再取口、小。
東	木月	DB	木冂	取木，再把冂重疊上去
末	木十	DJ	木十	先取木，再取重疊的十
來	木人人	DOO	木人人	先取木，再取人、人。
桒	竹木中	HDL	ノ木中	棄烏連體字。
桑	竹木中心	HDLP	ノ木丨匕	ノ取完，再取木之主形
奧	大大大大	KKKK	大乂乂乂	大主形先取，再取乂。
拳	火人手	FOQ	火人手	火主形先取，再取人。
脊	火金月	FCB	火八月	火主形先取，八再補上

7.7.3 Invoking Tsang–Chi Input Method

When you invoke the Tsang–Chi input method, the Chinese string " 倉頡" is displayed in the status area, as shown in Figure 7–7.

Figure 7–7: Invocation of the Tsang–Chi Input Method

〔倉頡〕：

The radicals that you enter with the Tsang-Chi input method are displayed in the preedit area, as shown in Figure 7–8. To correct the data, do one of the following:

- Press the Delete key and reenter the correct radical
- Press the 倉頡 key (that is, F6 on a standard LK201 or LK401 keyboard) to erase all radicals in the preedit buffer

To signal end of input, press the Return key or Space bar.

Figure 7–8: Entering a Tsang–Chi Radical

〔倉頡〕：木

7.7.4 Tsang-Chi Multiple Candidates

If there is exactly one character represented by a Tsang–Chi code, the character is sent directly to the application. Sometimes, multiple candidates for a Tsang–Chi code are available for selection when the code represents more than one Chinese character. In this case, the candidates are displayed in the preedit area in the following order:

1. Most frequently used

2. Less frequently used

3. Seldom used

The preedit area can display up to nine candidates at a time, as shown in Figure 7–9.

Figure 7–9: Multiple Candidates

1 X X X   4 X X X   7 X X X   →

The numbers 1, 4, and 7 divide the nine characters into three groups so that you can easily select the desired candidate. To select a character that is displayed in the preedit area, press the corresponding numeric key on the main keyboard.

When there are more than nine candidates for selection, the indicators, " → ," " ← → " and " ← " are displayed in the preedit area. Table 7–16 lists the indicators and their definitions.

Table 7–16: Meaning of Arrow Characters

Indicator	Definition
→	The current row is the first row and you can press [Space] or [⇒] to move to the next row.
← →	The current row is somewhere between the first and the last row. You can press: • [Space] or [⇒] — move to the next row • [⇐] — move to the previous row • [⇑] — move to the first row
←	The current row is the last row and you can press [⇐] to move to the previous row or [⇑] to the first row.

If you enter another Tsang–Chi code without selecting a candidate, the first candidate in the list is sent to the application.

If you do not want to select any candidate, but want to clear the Tsang–Chi code, press the Return key or the 倉頡 key (that is, F6).

7.7.5 Tsang-Chi Repeat Character Input

If you want to repeat input of the same character, press the equals (=) key.

7.7.6 Tsang-Chi Error Handling

If you input incorrect data, the bell will ring. If no character is generated after you enter a Tsang–Chi code, this indicates that there is no character for the code. The radicals already entered remain in the preedit buffer. To handle the error, do one of the following:

1. Press the Delete key to erase the radicals, one at a time, and then reenter the correct radicals.

2. Press the Return key or the 倉頡 key to erase all radicals in the preedit buffer and then reenter the correct radicals.

3. Enter new radicals without pressing the Return key. The radicals in the preedit buffer are replaced by the newly entered radicals.

7.8 Quick Tsang–Chi Input Method

The Quick Tsang–Chi input method, also known as the Quick input method or the Easy input method, is a variant of the Tsang–Chi input method and follows the same principles and rules for decomposing characters into radicals as described in Section 7.7.2.1. However, the process for entering radicals is simplified and requires only the first and the last radicals.

For example, the character "商" is decomposed into the following:

• Under the Tsang–Chi input method: "卜," "金," "月," "口

• Under the Quick Tsang–Chi input method: "卜" and "口

This section discusses the mechanism of the Quick Tsang–Chi input method. For details about character decomposition, see Section 7.7.2.

7.8.1 Quick Tsang–Chi Code Generation

As in the Tsang–Chi input method, the character decomposition in Quick Tsang-Chi is based on whether a character is of the composite form or the connected form. However, the Quick Tsang–Chi input method requires only the first and the last radicals regardless of the number of radicals obtained.

7.8.2 Invoking Quick Tsang–Chi Input Method

When you invoke the Quick Tsang–Chi input method, the Chinese string " 速成 " is displayed in the status area, as shown in Figure 7–10.

Figure 7–10: The Quick Tsang–Chi Input Method

〔速成〕 :

7.8.3 Entering Quick Tsang–Chi Code

The radical that you enter is displayed in the preedit area, as shown in Figure 7–11. To correct the data, press the Delete key and reenter the correct radical. Alternatively, you can press the 速成 key (that is, F7 on a standard LK201 or LK401 keyboard) to erase all radicals in the preedit buffer. If only one radical is required to input a character, press the Return key or Space bar to signal the end of input.

Figure 7–11: Entering a Quick Tsang–Chi Code

〔速成〕 :月

7.8.4 Quick Tsang-Chi Multiple Candidates

If exactly one character is represented by a Quick Tsang–Chi code, the character is sent directly to the application. Frequently, multiple candidates for a Quick Tsang–Chi code are available for selection when the Quick Tsang–Chi code represents more than one Chinese character. In this case, the candidates are displayed in the preedit area.

To clear the Tsang-Chi code and not select any candidate, press the Return or the 速成 key (that is, F7).

7.8.5 Quick Tsang-Chi Repeat Character Input

If you want to repeat the input of the same character, press the equals (=) key.

7.8.6 Quick Tsang-Chi Error Handling

If you input incorrect data, the bell will ring. If no character is generated after you enter a Quick Tsang–Chi code, the bell indicates that there is no character for the code. The radicals you entered remain in the preedit buffer. To handle the error, you can do one of the following:

- Press the Delete key to erase the radicals, one at a time, and then reenter the correct radicals.

- Press the Return key or the 速成 key to erase all radicals in the preedit buffer, and then reenter the correct radicals.

- Enter new radicals without pressing the Return key. The radicals in the preedit buffer are replaced by the newly entered radicals.

7.9 Phonetic Input Method

The Phonetic input method is based on Chinese phonetic symbols (bopomofo) that represent the pronunciation of Chinese characters.

7.9.1 Phonetic Symbol Categories

Phonetic symbols can be divided into three categories: consonants, vowels, and tone marks. There are 21 consonants, 16 vowels, and 5 tone marks. The 5 tone marks for Chinese pronunciation are the first, the second, the third, the fourth and the light tones. Chinese phonetic symbols are assigned to the alphanumeric keys on the main keyboard. Table 7–17 is a summary of all consonants, vowels, and tone marks.

Table 7–17: Phonetic Symbols

聲母	英文鍵
ㄅ	1
ㄆ	Q
ㄇ	A
ㄈ	Z
ㄉ	2
ㄊ	W
ㄋ	S
ㄌ	X
ㄍ	E
ㄎ	D
ㄏ	C
ㄐ	R
ㄑ	F
ㄒ	V
ㄓ	5
ㄔ	T
ㄕ	G
ㄖ	B
ㄗ	Y
ㄘ	H
ㄙ	N

韻母	英文鍵	
ㄚ	8	
ㄛ	I	
ㄜ	K	
ㄝ	,	
ㄞ	9	
ㄟ	O	
ㄠ	L	
ㄡ	.	
ㄢ	0	
ㄣ	P	
ㄤ	;	
ㄥ	/	
ㄦ	–	
ㄧ	U	[註]
ㄨ	J	[註]
ㄩ	M	[註]

聲調	英文鍵
一聲 ( )	Space bar
二聲 (´)	6
三聲 (ˇ)	3
四聲 (`)	4
輕聲 (·)	7

[註] : ㄧ、ㄨ、ㄩ亦屬介音。

_____ **Note** _____

The vowels ㄧ, ㄨ, and ㄩ are also called semi-vowels.
_____

7.9.2 Phonetic Code Generation

The pronunciation of a Chinese character is composed of consonants, vowels, and tone marks. Therefore, a phonetic code can be generated according to the following rules:

- The phonetic symbols are entered in the following order:

  1. Consonant
  2. Vowel
  3. Tone marks

- A phonetic representation must have at least one consonant or one vowel.

- The end of input must be indicated with a tone mark or a Return.

Table 7–18: Examples of Phonetic Input

Code Format	Phonetic Symbols	Characters
Consonant + vowel + tone mark	ㄕㄟˊ	誰
Consonant + [vowel] + tone mark	ㄓ	知, 織
Vowel + tone mark	ㄢ	安, 庵
Consonant + semivowel + vowel + tone mark	ㄒㄧㄠˇ	曉, 小
Semivowel + vowel + tone mark	ㄨㄛˇ	我

7.9.3 Invoking Phonetic Input Method

When you invoke the Phonetic input method, the Chinese string "注音" is displayed in the status area, as shown in Figure 7–12.

The example in Figure 7–12 shows how to input the character "吉" by entering the phonetic symbols ㄐㄧˊ (RU6) at the main keyboard.

Figure 7–12: Entering the Phonetic Symbols for "吉"

    [注音] : ㄐㄧˊ   1 及吉吃   4 即急級   7 集極擊   →

7.9.4 Entering Phonetic Code

The phonetic symbols that you enter are displayed in the preedit area. To correct the data, press the Delete key and reenter the correct symbol. Alternatively, you can press the [注音] key (that is, F10 on a standard LK201 or LK401 keyboard) to erase all phonetic symbols in the preedit buffer.

You can press various termination keys to signal that you are done entering the phonetic symbols. Table 7–19 shows the results of entering phonetic symbols with different termination keys.

Table 7–19: Phonetic Symbols with Different Termination Keys

Tone	Key	Description	Example
1st	Space	Lists characters with the same consonant or vowel in the order of the first, second, third, fourth, and light tone marks.	Type 业, then press the Space bar, "业、 之、 支...,直、 姪、 值..., 止、只、 旨..., 至、 志、 ..." will be listed.
2nd	6	Lists characters of the second tone.	Type 3ㄧㄢˊ, "年、 黏、 哖..." will be listed.
3rd	3	Lists characters of the third tone.	Type ㄌㄧㄠˇ, "了、 瞭、 暸..." will be listed.
4th	4	Lists characters of the fourth tone.	Type ㄌㄧㄝˋ, "列、 劣、 冽..." will be listed.
Light	7	Lists characters of the light tone.	Type ㄕˊ, "匙..." will be listed
None	Return	Characters corresponding to the phonetic symbols are displayed according to the order of tone marks. If only a consonant is entered before pressing Return, characters corresponding to any valid combinations of this consonant and other vowels are also displayed.	Type 业,, then press the Return key, characters corresponding to the order "业、 业ˊ、 业ˇ、 业ˋ、 业ㄚ、 业ㄚˊ、 ...,业ㄜ、 业ㄜˊ、 ...,业ㄨㄚ..." will be listed.

7.9.5 Phonetic Symbol Multiple Candidates

If a phonetic string matches exactly one character, the character is sent directly to the application. Frequently, a phonetic string matches multiple Chinese characters. In this case, the candidates are displayed in the preedit area.

If you do not want to select any candidate, but want to clear the phonetic code, press the Return key or the 注音 key (that is, F10).

7.9.6 Phonetic Symbol Repeat Character Input

If you want to repeat the input of the same character, press the equals (=) key.

7.9.7 Phonetic Symbol Error Handling

If you input incorrect data, the bell will ring. If no character is generated after you enter a phonetic code, this indicates that there is no character for the code. The phonetic symbols you entered remain in the preedit buffer. To handle the error, you can do one of the following:

- Press the Delete key to erase the phonetic symbols, one at a time, and then re-enter the correct symbols.

- Press the Return key or the 注音 key to erase all phonetic symbols in the preedit buffer, and then reenter the correct phonetic symbols.

- Enter new phonetic symbols without pressing the Return key. The phonetic symbols in the preedit buffer are replaced by the newly entered symbols.

7.10 Internal Code Input Method

Each character in DEC Hanyu has been assigned a unique internal code, just like the ID number of a company employee. For a complete list of the characters and their internal codes, see *The DEC Chinese Code Book* (Part Number EK-VT38D-CB-001).

_____ **Note** _____

In this Tru64 UNIX release, the Internal Code input method supports only the DEC Hanyu internal code. The internal codes for Taiwanese EUC and Big-5 are not supported. Even if you set the locale to one of the Taiwanese EUC or Big-5 locales, this input method still requires you to specify the DEC Hanyu internal code.

_____

7.10.1 Internal Code Input Procedure

When you invoke the Internal Code input method, the Chinese string " 內碼 " is displayed in the status area.

To enter an internal code, you can optionally enter a character set number to be followed by a 4-digit hexadecimal number that specifies the position of the character with respect to the character set. The character set number can be 1 for the CNS 11643-1986 character set, or 2 for the DTSCS character set.

If you omit the character set number and enter only the 4-digit hexadecimal code, you must press the Return key or the Space bar to signal the end of input. If you enter the character set number and the 4-digit hexadecimal code, the respective character is sent automatically without pressing the Return key. Figure 7–13 shows the input of "例" using the Internal Code input method.

Figure 7–13: Input of "例" Using the Internal Code Input Method

租賃契約條_

〔內碼〕: 1CBF

The internal code that you enter is displayed in the preedit area. To correct the data, press the Delete key and re-enter the correct code. Alternatively, you can press the 內碼 key (that is, F8 on a standard LK201 or LK401 keyboard) to erase all characters in the preedit buffer.

Since internal codes are unique for any symbols or Chinese character in DEC Hanyu, there are no multiple candidates for an internal code.

7.10.2 Internal Code Repeat Character Input

If you want to repeat the input of the same character, press the equals (=) key.

7.10.3 Internal Code Error Handling

If you input incorrect data, the bell will ring. If no character is generated after you enter an internal code, this indicates that there is no character for the code. The internal code you entered remains in the preedit buffer. To handle the error, you can do one of the following:

- Press the Delete key to erase the characters representing the internal code, one at a time, and then reenter the correct internal code.

- Press the Return key or the 內碼 key to erase all characters in the preedit buffer, and then reenter the correct internal code.

- Enter a new internal code without pressing the Return key. The characters in the preedit buffer are replaced by the newly entered internal code.

7.11 Phrase Input Method

The Phrase input method enables you to define a database of frequently used phrases and store that database in a user or systemwide directory. Each frequently used phrase is identified by a phrase code. To input a phrase, enter its phrase code. The Phrase input method converts the code to the respective phrase.

There are two Phrase input methods available on the Tru64 UNIX operating system; a Phrase input method available by means of the dxim input server and a Phrase input method available by means of the dxhanziim and dxhanyuim input servers. These two Phrase input methods use different databases and different user interfaces, as described in the following sections.

7.11.1 dxim Phrase Input Method

The dxim Phrase input method allows you to define a database of frequently used phrases and store that database in a user or systemwide directory.

You associate each of the phrases in the phrase database with a unique phrase code. The phrase code can be any alphabetic label. The associated phrase can be any string of characters that is appropriate to the application's locale. You create the phrase database as a simple text file containing phrase codes and phrases and store it in a user directory or a system directory on a path that is locale dependent.

For example, if you create a phrase database file for use with applications running in the zh_CN.dechanzi locale, you copy the file to one of the following directories:

- For a systemwide database file that will be used with simplified Chinese applications using the Hanzi codeset, copy the phrase database to /usr/i18n/share/dxim/zh_CN.dechanzi/phrase_data.txt.

- For a personal database file that will be used with simplified Chinese applications using the Hanzi codeset, copy the phrase database to $HOME/.dxim/zh_CN.dechanzi/phrase_data.txt.

Use the following rules when populating the phrase database:

- Use a text editor to create and populate the Phrase input method database files.

- Enter the phrase code and its associated phrase on the same line. Separate the code and phrase with spaces. However, the phrase code itself cannot contain spaces. The associated phrase can contain spaces, but not in the first or last character position.

- A phrase code consists of alphabetic characters and underscores. The associated phrase can contain any character that is legal in the target locale.

To invoke the Phrase input method, you must use dxim to activate the input method and use the default key sequence (or an invocation key sequence that you have defined). For example, if the Phrase input method is active and the client application locale supports it, press the default invocation key, [F12], or a redefined invocation key to invoke the Phrase input method.

When you use the Phrase input method and enter a phrase code as input to the application, dxim searches your personal database file, if it exists, and the systemwide database file along the locale-dependent path. If dxim finds the same phrase in both databases, it uses the definition in your personal database.

When you enter a phrase code, it appears in the preedit area. To correct erroneous input, use the [Delete] key and re-enter the code. Press the [Return] key to signal the end of phrase code input. If you enter an incorrect code, the error bell rings. The message, "Requested phrase does not exist", may also display. To correct the problem, make sure that the phrase is in the database file and that the file path and application locale are consistent.

For additional information on the dxim Phrase input method, see the online help for dxim.

7.11.2 dxhanziim and dxhanyuim Phrase Input Method

The Tru64 UNIX operating system provides a Phrase Utility that you use to create phrase databases for use under dxhanziim and dxhanyuim. For details, see *Using International Software* in the Tru64 UNIX documentation set.

In addition, the firmware of the VT382-D traditional Chinese terminal is designed to allow you to download phrases into its built-in memory. For DECwindows Motif applications, you do not need to download phrase definitions because the Chinese input servers can directly access the phrase databases. In addition, you can select the phrase database being used.

When you invoke the Phrase input method, the Chinese string "片語" is displayed in the status area of dxhanyuim, and "片语" in the status area of dxhanziim. Figure 7–14 shows an example of entering a phrase in dxhanyuim by specifying its phrase code. Figure 7–15 shows an example of converting the phrase code to the phrase.

Figure 7–14: Entering a Phrase Code

國際觀光旅館名錄

名稱：國賓大飯店/AMBASSADOR

地址：台北市中山北路二段６３號

63, Chung Shan N. Rd., Sec. 2, Taipei

名稱：_

〔片語〕：ASIA

Figure 7–15: Converting a Phrase Code to a Phrase

國際觀光旅館名錄

　名稱：　國賓大飯店/AMBASSADOR

　地址：　台北市中山北路二段６３號

　63, Chung Shan N. Rd., Sec. 2, Taipei

　名稱：　環亞大飯店/ASIA WORLD PLAZA_ [倉頡]：

The phrase code that you enter is displayed in the preedit area, as shown in Figure 7-14. To correct the data, press the Delete key and reenter the correct code. Alternatively, you can press the 片語 key (that is, F9 on a standard LK201 or LK401 keyboard) to erase all characters in the preedit buffer.

The phrase code can consist of at most eight characters. If it has fewer than eight characters, press the Return key or Space bar to signal the end of input. If it has exactly eight characters, the respective phrase is sent automatically without having to press the Return key after you enter the last character.

The Phrase input method is different from other input methods in the sense that once a phrase is entered, the input mode switches back to the original input mode from which the Phrase input method was invoked. To input another phrase, press the 片語 key again and then enter another phrase code.

7.11.3 Phrase Error Handling

If the data that you input is incorrect, the bell will ring. Errors can occur if the requested phrase can not be found or if there is a problem with the phrase definition file. The following sections discuss these error conditions.

7.11.3.1 Requested Phrase Does Not Exist

If the message "此片語不存在" (Requested phrase does not exist) is displayed when you enter a phrase code, you can check the following:

- Is the phrase in the phrase definition file.

- Are you using a VT382-D terminal:

  - Check to be sure that the phrase definition file containing the phrase definition has been downloaded to the VT382-D Chinese terminal.

  - Check to see if the terminal power supply has been interrupted since you last loaded the phrase definition file.

- Check to see if you have reset the VT382-D terminal. The Recall and Default operations provided by the VT382-D setup menu clear the phrase definitions that have been downloaded.

- Are you using a Chinese input server:

    - Check to be sure that the phrase definition file containing the phrase definition has been selected by the Chinese input server.

    - Check to see if the LANG environment variable has been set up correctly before starting the Chinese input server.

If the phrase has been defined but it does not exist in the in-memory phrase database, reload or reselect the phrase definition file.

7.11.3.2 Errors in the Phrase Definition File

Errors can occur in the phrase definition file. If you cannot solve the problem using the procedures in Section 7.11.3.1, you can do one of the following:

- Check the syntax of the definition statements in the phrase definition file. A phrase cannot be downloaded if the syntax of its definition is incorrect.

- If you are using a VT382-D terminal, a maximum of 100 phrases can be downloaded. If the number of phrases in the phrase definition file exceeds this limit, the definitions beyond the limit are not downloaded.

7.12 Symbol Input in traditional Chinese

The Symbol input method is an intuitive way to input 2-byte symbols, such as punctuation marks, table and mathematical symbols, foreign characters, phonetic symbols, and traditional Chinese control characters.

You can invoke the Symbol input method only from the Tsang–Chi or the Quick Tsang–Chi input method. You can use it to input more than 600 symbols, including the following:

- All full form alphabets (A -Z, a-z)

- Full form numerals (0-9)

- Full form symbols (such as ＃, ＄, ％...)

- Phonetic elements (such as ㄅ, ㄆ, ㄇ, ㄈ)

- Chinese radicals and special symbols (such as 「, 」, ※)

7.12.1 Invoking Symbol Input Mode

To enter the Symbol input mode, press the Z key when invoking Tsang–Chi or the Quick Tsang–Chi input method. When you invoke the Symbol input mode, the Chinese string "符號" is displayed in the status area.

7.12.2 Rules of Symbol Input

Use the following rules when entering symbols:

- If a symbol exists on the keyboard, input it the same way you enter single-byte characters.

Symbol	Input Code	Key Sequence
!	!	<Shift/1>
#	#	<Shift/3>
$	$	<Shift/4>
A	A	[A]
a	A	[a]

- Symbols of similar form or meaning have the same key sequence.

Symbol	Input Code	Description
(	(	vertical parenthesis
)	)	vertical parenthesis
⌒	(	horizontal parenthesis
⌣	)	horizontal parenthesis

- If a symbol does not exist on the keyboard, decompose it according to its shape.

Symbol	Key Sequence
≠	[=][/]
♀	[O][+]
§	[S][S]
kg	[K][G]

- For graphical symbols, press the hyphen (-) key and a number within the range 1-8 to specify horizontal symbols or press the vertical bar key (|) and a number within the range 1-7 to specify vertical symbols.

- To enter symbols for constructing a table or diagram, press the T key and an alphabet key specifying the direction.

Symbol	Key Sequence
└	[T][Z]
┴	[T][X]
┘	[T][C]
├	[T][A]
	[T][S]
┼	

┌ Q	┬ W	┐ E
├ A	┼ S	┤ D
└ Z	┴ X	┘ C

- To enter arrows of various directions, press the A key and an alphabet key specifying the direction.

Symbol	Key Sequence
↙	[A][Z]
↓	[A][X]
↘	[A][C]
←	[A][A]

↖ Q	↑ W	↗ E
← A		→ D
↙ Z	↓ X	↘ C

- Press the P key to indicate the input of phonetic symbols. To specify the required symbol, press the key with the respective phonetic symbol marked on the keyboard. See Chapter 6 for the keymap of phonetic symbols.

Symbol	Key Sequence
ㄅ	[P][1]
ㄆ	[P][Q]
ㄇ	[P][A]
ㄈ	[P][Z]
ㄉ	[P][2]

- Press the G key (or g for the lowercase) to input Greek letters and then enter the first two letters of the Greek character's name. Use uppercase or lowercase letters as appropriate.

Symbol	Key Sequence
Α	[G][A][L]
Β	[G][B][E]
α	[g][a][l]
β	[g][b][e]
Ε	[G][E][P]

7.12.3 Entering Symbol Code

You can input a symbol by entering its symbol code and then pressing the Return key or Space bar to signal the end of input.

The following example describes how to enter the full form alphabet "Ａ":

1. Press the Z key to invoke the Symbol input mode.

   〔符號〕：

2. Press the A key to enter the first alphabet of the Symbol code.

   〔符號〕：Ａ

7.12.4 Symbol Code Multiple Candidates

If a symbol code matches exactly one character, the character is sent directly to the application. However, sometimes a symbol code matches multiple candidates. In this case, the candidates are displayed in the preedit area. To continue the preceding example:

1. In the preceding example, press the Return key to signal the end of input.

〔符號〕：A 　　　1 Ａ日曰

2. Press the 1 key to select " Ａ "

Ａ_

〔倉頡〕：

_____ **Note**_____

After the symbol is input, the input mode switches back to the original input mode.
_____

7.12.5 Symbol Code Error Handling

If no symbol is input after you enter a symbol code, the symbol code is invalid and the input mode switches back to the original input mode. You can press the Z key to invoke the Symbol input mode again and reenter the correct symbol code.

7.13 Input of User–Defined Characters in DECwindows Motif

Tru64 UNIX supports the input of User-Defined Characters (UDC) through the Tsang–Chi and Quick Tsang–Chi input methods. When you invoke the Chinese input server, the Tsang–Chi dictionaries are loaded. When you enter an input key sequence, the Tsang-Chi and Quick Tsang-Chi input methods search these dictionaries for a match in the following order.

1. The private UDC Tsang–Chi dictionary that you create

2. The system wide UDC Tsang–Chi dictionary

3. The system default Tsang–Chi dictionary

You can use the `cedit` utility to invoke the UDC Manager Utility and define the input sequences of user-defined characters. You then use the `cgen` command with the `-iks` option to produce a UDC Tsang–Chi dictionary for the Chinese input server. You should also define the corresponding font glyphs (both 18 X 16 and 24 X 24 size fonts). After you define the user-defined characters and their corresponding information in the system or user database, restart the input server to reread the database. For details, see `cgen(1)` and `cedit(5)`.

_____ **Note** _____

Use the dxterm terminal emulator to run the UDC Manager Utility. The dtterm terminal emulator does not support `cedit` functions and the UDC Manager Utility may hang under dtterm.

_____

By default, the Chinese input server puts your private UDC Tsang–Chi dictionary in the `~/.iks/dwimdb-dechanyu` directory. You can use the USER_UCD_DICT environment variable to override this default.

The system wide UDC Tsang–Chi dictionary is, by default, placed in the `/var/i18n/iks/dwimdb-dechanyu` directory. If the dictionary is located elsewhere on your system, you use the SYSTEM_UDC_DICT variable to override the default location. If you are a system administrator, you can install a UDC Tsang–Chi dictionary by placing it at the default location and setting proper access mode for system-wide use.

The system default Tsang–Chi dictionary, which is shipped with Tru64 UNIX, is placed in the `/usr/i18n/hanyu/hanyu_tsangchi.dic` directory. If it is installed at a different location, you can use the HANYU_SYSTEM_DICT environment variable to specify the location.

The Chinese input server supports UDC input key sequences only in the DEC Hanyu codeset.

To display UDC in DECwindows Motif applications or the Chinese input server, UDC fonts should be available on your system. For instance, the Chinese input server will display UDC using one of the following fonts:

- -adecw-screen-*--18-*-cns11643.1986-udc

- -adecw-screen-*--24-*-cns11643.1986-udc.

If you define your own UDC fonts, you can override the system UDC fonts by adding the directory in which the UDC fonts are located to the font path. For example, if your private UDC fonts are located in the ~/fonts directory, enter the following command to update the font path:

```
% xset +fp ~/fonts
```

Use the UDC Manager Utility (`cedit`) to define UDC font glyphs (both 18 X 16 and 24 X 24 size fonts) and use the `cgen` utility to generate fonts for use in DECwindows Motif. For details, see *Writing Software for the International Market*.

7.14 5–Stroke Input Method

The 5-Stroke input method makes use of the basic strokes to construct the Chinese characters. A stroke is a segment of continuous line or curve that constitutes a Chinese character. Table 7–20 describes the five categories of strokes.

Table 7–20: Stroke Categories

Category	Description
Horizontal strokes or 横 〔一〕	Horizontal lines and left-to-right ticks 〔提〕
Vertical strokes or 竖 〔丨〕	Vertical lines and right-to-left hooks 〔竖左钩〕
Slash or 撇 〔丿〕	Slanting lines and curves drawn towards lower-left
Backslash or 捺 〔丶〕	Dots 〔点〕, slanting lines, and curves drawn towards lower-right
Zip-zap curves or 折 〔乙〕	Including different types of joints or corners which can be drawn in single continuous strokes

Using the 5-Stroke input method you can input single Chinese characters and Chinese terms. Approximately 5,000 terms can be input using this method.

7.14.1 5-Stroke Input Mechanism

To input a Chinese character using the 5-Stroke input method, enter its 5-Stroke code through the numeric keypad according to the writing order. Table 7–21 shows the codes representing the five categories of strokes.

Table 7–21: 5-Stroke Code

Stroke	Code	Key
一	1	[KP1]
丨	2	[KP2]
丿	3	[KP3]
丶	4	[KP4]
乙	5	[KP5]

Figure 6–5 shows the numeric keypad layout for entering 5-Stroke codes.

The following are the general rules of writing order for Chinese characters:

1. Write from top to bottom.

2. Write from left to right.

3. Write from outside to inside.

4. Write the inside radical before drawing the last stroke of the outside radical (for example, the last stroke of " 围 " is " 一 ").

7.14.1.1 Single Character Input

If a Chinese character is composed of exactly five strokes, enter the strokes according to the writing order. If the character is composed of fewer than five strokes, press the [KP0] key to signal the end of input. If the character is composed of more than five strokes, enter the first four strokes and the last stroke. Table 7–22 shows some examples of using the 5-Stroke input method.

Table 7–22: Input of Single Characters with the 5-Stroke Input Method

Character	No. of Strokes	Write Order	5-Stroke Code	Key Sequence
用	5	ノ, 乙, 一,一, 丨	35112	[KP3][KP5][KP1][KP1][KP2]
五	4	一, 丨, 乙, 一	12510	[KP1][KP2][KP5][KP1][KP0]
总	9	丶, ノ, 丨, 乙, 丶	43254	[KP4][KP3][KP2][KP5][KP4]

If you are uncertain about the type of strokes or the writing order of strokes, press the wildcard key [KP6] in place of the strokes. For details, see Section 7.14.5**.**

7.14.1.2 Input of Terms

The 5-Stroke input method can be used to input terms. Eight strokes are required for entering terms. Press the [KP7] key before inputting strokes. The [KP7] key signals the system that subsequent key strokes are for composing a term instead of a character.

The number of strokes input for each character in a term depends on the number of characters composing the term as shown in Table 7–23.

Table 7–23: Input Terms with the 5-Stroke Input Method

Number of Characters in Term	Strokes to be Input
2	First 4 strokes of each character
3	First 2 strokes of the first two characters and the first 4 strokes of the last character
4	First 2 strokes of each character
> 4	First 2 strokes of the first three characters and the last character

If a character is composed of fewer strokes than required, press the [KP6] key for the outstanding strokes. If this character is the last character, press the [KP0] key to signal the end of input.

Term	Number of Characters	Characters	No. of Strokes	Req'd Strokes	5-Stroke Code	Key Sequence
世界	2	世	5	4	1221	[KP7][KP1][KP2]
		界	9	4	2512	[KP2][KP1][KP2]
						[KP5][KP1][KP2]
工人	2	工	3	4	121(6)	[KP7][KP1][KP2]
		人	2	4	34(0)	[KP1][KP6][KP3]
						[KP4][KP0]
计算机	3	计	4	2	45	[KP7][KP4][KP5]
		算	14	2	31	[KP3][KP1][KP1]
		机	6	4	1234	[KP2][KP3][KP4]
星期一	3	星	9	2	25	[KP7][KP2][KP5]
		期	12	2	12	[KP1][KP2][KP1]
		一	1	4	1(0)	[KP0]
百货公司	4	百	6	2	13	[KP7][KP1][KP3]
		货	8	2	32	]KP3][KP2][KP3]
		公	4	2	34	[KP4][KP5][KP1]
		司	5	2	51	
一分为二	4	一	1	2	1(6)	[KP7][KP1][KP6]
		分	4	2	34	[KP3][KP4][KP4]
		为	4	2	43	[KP3][KP1][KP1]
		二	2	2	11	
中华人民共和国	7	中	4	2	25	[KP7][KP2][KP5]
		华	6	2	32	[KP3][KP2][KP3]
		人	2	2	34	[KP4][KP2][KP5]
		国	8	2	25	

7.14.2 5-Stroke Input Procedure

When you invoke the 5-Stroke input method, the string "五笔画" is displayed in the status area. When you input the 5-Stroke code, all the Chinese characters that match the code sequence are displayed in the 26th line of the terminal or the preedit area of the display. Each character is associated with a number and you can select the character immediately using the corresponding numeric key on the main keyboard.

For example, if you press the [KP1] key, all characters starting with a horizontal stroke are displayed in the preedit area.

五笔画:1 1有DEF 2地FBN 3来GOI 4于GFK 5下GHI   +

If you press the [KP2] key, all Chinese characters with starting 5-Stroke code 12 are displayed in the preedit area.

五笔画:12 1地FBN 2下GHI 3可SKD 4机SMN 5本SGD   +

You can use the Delete key to correct the 5-Stroke code. When you delete digits in the 5-Stroke code, the characters displayed in the preedit area are changed to match the remaining 5-Stroke code.

For example, press the Delete key once:

五笔画:1 1有DEF 2地FBN 3来GOI 4于GFK 5下GHI   +

Press the Delete key once more:

五笔画:

Alternatively, you can press the Return key without selecting any character. The preedit area is refreshed and you can enter a new 5-Stroke code.

7.14.3 5-Stroke Multiple Candidates

Because the same 5-Stroke code can represent more than one Chinese character, multiple candidates are sometimes available for selection when you enter a 5-Stroke code. For example:

将所有区域_
五笔画: 12345 1机SMN 2榜SUPY 3概SUCQ 4棍SJXX 5杭SYMN +

To select a character that is displayed in the preedit area, press the corresponding numeric key on the main keyboard. In the example, press numeric key 3 to select "概". If the desired character is displayed at the first position (in this case, the "机" character), you do not need to press the 1 key to select "机". Enter another 5-Stroke code and the character will automatically be selected. This character is called the default character.

If there are too many candidates for selection, the candidates are displayed in two or more rows and a plus (+) sign is displayed to signal this. Press the Space bar or [KP9] key to display the next row of characters. To display the previous row of characters, press [Shift/Space] or the [KP8] key.

7.14.4 5-Stroke and the Association Mode

To enhance input efficiency, the Chinese input method maintains a list of built-in phrases in memory. If you select a character from the candidate list (see Section 7.14.3) and there is a built-in phrase whose first character matches this character, Association Mode is automatically activated. For example:

将所有区域划_

联 想 → 1– 分

The string "联想→" is displayed in the status area. You can select one of the built-in phrases in Association Mode. After you select the phrase, you automatically exit Association Mode.

If there are many built-in phrases, they are displayed in two or more rows and a plus (+) sign is displayed to signal this. Press the Space bar or [KP9] key to view the next row of phrases. To move backwards, press [Shift/Space] or the [KP8] key.

If you do not want to use any of the associated phrases, enter another 5-Stroke code to automatically exit the Association Mode.

_____ **Notes** _____

Association Mode is only active during single character input. It cannot be activated during term input.

Association Mode is activated only if you press a numeric key to select a single character. However, the 1 key, which is one of the ways to enter a default character, is a special case. If you do not press the 1 key to select the default character, the Association Mode is not activated though the default character is entered.

_____

7.14.5 5-Stroke Wildcard Key

The [KP6] key is a wildcard key, which can replace any strokes about which you are uncertain.

For example:

"字" -" 丶", " ?", " 乙", " 乙", " 一"

This example assumes you are uncertain about the second stroke, and so press [KP4], [KP6], [KP5], [KP5], [KP1].

Another example:

" 化" – " ノ" , " ｜" , " ?" , " ?"

This example assumes you are uncertain about the third and the fourth strokes, and so press [KP3], [KP2], [KP6], [KP6], [KP0].

7.15 5–Shape Input Method

The 5-Shape input method is a high performance input method that composes root radicals 〔字根〕 to form Chinese characters. Using the 5-Shape input method, you can enter one to four root radicals to input thousands of Chinese characters and terms.

In addition to character input, the 5-Shape input method also supports term input. The 5-Shape input method supports approximately 5,000 terms, each of which can be defined by four radicals.

_____ **Note** _____

When using a VT382-C terminal with SoftODL enabled, you can use a user-defined key sequence in the 5-Shape input method to input a user-defined character (UDC). This assumes the key sequence does not apply to other existing characters. Use `cedit` to create UDCs and their corresponding input key sequences.

_____

7.15.1 Distribution of 5-Shape Radicals

There are 130 root radicals defined on 25 keys on the main keyboard. (See Figure 6–3 and Figure 6–4 for the layout of the LK201-C and LK401-C keyboards.) The 25 keys are divided into five groups, each containing five keys, as shown in Table 7–24.

Table 7–24: Shape Code

	Key 1	Key 2	Key 3	Key 4	Key 5
Group 1	[11]/[G]	[12]/[F]	[13]/[D]	[14]/[S]	[15]/[A]
Group 2	[21]/[H]	[22]/[J]	[23]/[K]	[24]/[L]	[25]/[M]
Group 3	[31]/[T]	[32]/[R]	[33]/[E]	[34]/[W]	[35]/[Q]
Group 4	[41]/[Y]	[42]/[U]	[43]/[I]	[44]/[O]	[45]/[P]
Group 5	[51]/[N]	[52]/[B]	[53]/[V]	[54]/[C]	[55]/[X]

Figure 7–16: Distribution of Radicals

Figure 7–17: 5-Shape Radical Keys

五 笔 字 型 键 位 字 根

字母	键号	笔画	键名	基本字根
G	11	一	王	土戋五一
F	12	二	土	士二干十中寸雨
D	13	三	大	犬三手尹长古石厂厂ナ广
S	14		木	丁西
A	15		工	戈艹廿丗匚七弋
H	21	丨	目	且上止龰卜卜广疒
J	22	刂	日	日曰早刂刂刂虫
K	23	刂刂	口	川
L	24	刂刂刂	田	甲口四皿皿车力
M	25		山	由贝门凸几
T	31	丿	禾	禾竹丿⺀彳攵夂
R	32	彡	白	手扌⺕⺀⺊厂斤斤
E	33	彡	月	日舟彡⺍⺈乃用豕豕⺜⻑
W	34		人	亻八⺹⺯
Q	35		金	钅⺈角⺈乂儿灬⺊⺈夕⺹乚
Y	41	丶	言	讠文方广丶亠⺀主
U	42	冫	立	辛冫⺕⺀丷六立门广
I	43	氵	水	氺⺌水氺氵⺍业小⺍⺍
O	44		火	业⺌灬米
P	45		之	辶廴宀冖⺄
N	51	乙	己	巳己コ尸尸乙心忄⺍羽
B	52	巜	子	孑耳阝卩凵了也凵
V	53	巛	女	刀九彐彐巛
C	54		又	厶マ巴马ム
X	55		纟	纟弓匕匕幺⺍

Figure 7–16 and Figure 7–17 show that the distribution of radicals has three characteristics:

1. The root radicals and the Chinese key name are similar in shape. For example:

   The key name of Key [11] is " 王 " while its radicals include " 五 ", and so on.

2. The code for the first stroke of a root radical is the same as the group number, while the code for the second stroke is the same as the key number. For example:

   The first stroke of the radicals on key [41] such as " 文 "," 方 " and " 广 " is " 丶 ", which is the same as the group number. The second stroke is " 一 ", which is the same as the key number.

3. The key number denotes the number of strokes for constructing the root radicals. That is, the key number equals the number of strokes of the radicals. For example:

   The group number of " 一 " is 1. Therefore, " 一 " is located on key [11], " 二 " is located on key [12], and " 三 " is located on key [13].

7.15.2 5-Shape Decomposition of Chinese Characters

The 5-Shape input method decomposes Chinese characters according to their root radicals. The decomposition can be categorized as follows:

- Left-Right Type (Type 1)

  The character can be split into a left half and a right half, such as " 汉 "," 对 ", and " 湘 ".

- Upper-Lower Type (Type 2)

  The character can be split into an upper half and a lower half, such as " 华 "," 字 ", and " 想 ".

- Miscellaneous Type (Type 3)

  The character does not have a clear separation of left-right or upper-lower. Examples include " 本 "," 同 ", and " 这 ".

  The miscellaneous type can be further divided into individual type and inclusion type. Examples of individual type include " 本 "," 天 ", " 册 "," 司 "," 头 "," 进 ", and " 厢 ". Inclusion type refers to those characters with one of their root radicals totally or partially included in another root radical, such as " 司 "," 连 "," 凶 ", and " 回 ".

7.15.3 5-Shape Distinction Code

If a character is composed of fewer than four root radicals, you must enter a code that represents the decomposition type and the type of the last stroke, which is called the distinction code. Figure 7–18 illustrates the 5-shape input method distinction code.

Figure 7–18: Distinction Code for the 5-Shape Input Method

字 型 / 末 笔 形	左右型 ▮ 1	上下型 ▬ 2	杂合型 ⊠ 3
横　1	11(一)	12(二)	13(三)
竖　2	21(l)	22(ll)	23(lll)
撇　3	31(╱)	32(⸝)	33(⸜)
捺　4	41(╲)	42(⺀)	43(⺀)
折　5	51(乙)	52(《)	53(《《)

For example, the codes for "汀" , "洒" ,and "沐" are all [43][14] . To distinguish among them, you must enter the distinction code.

Character	Last Stroke	Decomposition	Distinction Code
汀	l (2)	Left-Right (1)	[21]
洒	一 (1)	Left-Right (1)	[11]
沐	丶 (4)	Left-Right (1)	[41]

The root radicals available on the main keyboard do not require a distinction code. Although the characters which are key names (such as " 大 ") and radical names (such as " 上 ") are composed of fewer than four radicals, no distinction code is required. For example," 厂 " can be decomposed into " 一 " and " 丿 " .

For those characters that do not have a unique writing order, such as " 刀 "," 切 "," 力 "," 九 ", and " 七 " , you can use zip-zap stroke code (5) for the last stroke. For example," 仇 " can be decomposed into " 亻 " , and " 九 ", and you may add the distinction code [51].

For those inclusion type characters, such as " 国 "," 围 "," 团 "," 进 ", and " 这 " , you should take the last stroke of the interior radical. For example, the distinction code for " 进 " is [23] because the last stroke is [2].

Single dots located near a radical, such as the point in " 义 " and " 太 " , belong to decomposition type 3. For example, the distinction code of " 为 " is [43].

7.15.4 Principles of 5-Shape Character Decomposition

The following four principles can be used for decomposing Chinese characters:

1. Writing order.

   For example, " 远 " should be decomposed into " 二 ", " 儿 " , and " 辶 " instead of " 辶 "," 二 "," 儿 " .

2. Priority for larger radicals.

   Among the different ways of decomposition, choose the one that results in larger radicals. For example," 尺 " should be decomposed into " 尸 " and " 丶 " instead of " 〕 " and " 人 " .

   Another interpretation of this principle is that you should choose the decomposition method that creat fewer radicals.

3. Separated structure.

   If a character can be decomposed as a separated structure or a cross structure, choose the separated structure. For example," 于 " should be decomposed as " 一 " and " 十 " instead of " 二 " , and " 丨 " .

4. Intuition.

   If you cannot choose a character decomposition method based strictly on one of the previous principals, apply a reasonable approach based on a combination of your experience and the previous three principals.

7.15.5 5-Shape Input Mechanism

The 5-Shape input method allows for single-character input and term input.

7.15.5.1 Single Character Input

Each Chinese character can be composed of one to four root radicals. If the character is available on the keycap as a keyname, press the key four times. For example:

" 言" - press [41][41][41][41]

If the character is a root radical, press the key for the radical and enter the first, second, and last strokes. If the character is defined with fewer than four codes, press the Space bar to signal the end of input. For example:

Character	Decomposed Radicals	Number of Radicals	Key Sequence
西	西, 一, 丨, 一	4	[14][11][21][11]
厂	厂, 一, 丿	3	[12][11][31][Space]

There is one exception for the five basic strokes. The characters" 一" ," 丨 " ," 丿 " , " 丶 " , and " 乙" are single-stroke characters and they require two codes for input. To avoid multiple candidates for the same code, the 5-Shape input method requires adding two more codes ([24] [24]) to the two required codes. Table 7–25 shows the key sequences for entering these basic strokes.

Table 7–25: Entering Basic Strokes

Basic Stroke	Key Sequence
一	[11][11][24][24]
丨	[21][21][24][24]
丿	[31][31][24][24]
丶	[41][41][24][24]
乙	[51][51][24][24]

If the character does not appear on the keycap, follow the decomposition principles described in Section 7.15.4. If the character is composed of exactly four root radicals, enter the root radicals according to the writing order. If it is composed of more than four root radicals, enter the first, second, third, and last root radicals. If it is composed of less than four radicals, enter the radicals together with the distinction code. If the number of codes is still less than four, press the Space bar to signal the end of input.

For example:

Character	Decomposed Radicals	Distinction Code	Key Sequence
到	一，厶，土，刂	-	[11][54][12][22]
攀	木，X，X，手	-	[14][35][35][32]
。汉	氵，又	[41]	[43][54][41][Space]

7.15.5.2 Term Input

Approximately 5,000 built-in terms are defined in the Chinese input methods. You can use the 5-Shape input method to input multiple character terms by entering four root radicals as shown in Table 7–26.

Table 7–26: Input of Terms with the 5-Shape Input Method

Number of Characters in Term	Root Radicals to be Entered
2	The first two root radicals of each character
3	The first root radicals of the first two characters and the first two root radicals of the last character
4	The first root radical of each character
> 4	The first root radicals of the first three characters and the last character

Consider the following:

Term	Number of Characters	Character	Number. of Req'd Root Radicals	Root Radical	5-Shape Code
经济	2	经济	2 2	纟，又 氵，文	[55][54][43][41]
计算机	3	计算机	1 1 2	讠 竹 木，几	[41][31][14][25]
艰苦奋斗	4	艰苦奋斗	1 1 1 1	又 艹 大 丷	[54][51][13][42]
电子计算机	5	电子计机	1 1 1 1	曰 子 讠 木	[22][52][41][14]

7.15.6 5-Shape Input Procedure

When the 5-Shape input method is invoked, the string "五笔字型" is displayed in the status area. You can enter a 5-Shape code through the main keyboard and the data is displayed in the preedit area. When a complete 5-Shape code is entered, the character that matches the code is sent. For example:

吊_
五笔字型：kmhj

To correct the input data, press the Delete or Return key to erase the data.

If there is no valid character for the 5-Shape code, the bell rings to signal an error and you can enter another code.

7.15.7 5-Shape Multiple Candidates

If input codes match multiple candidates, valid candidates are displayed in the preedit area.

五笔字型：bnh_ 1了BNH 2阝BNH 3阝BNH 4凵BNH

The mechanism for selecting candidates is similar to that for the 5-Stroke input method. For details, see Section 7.14.

7.15.8 5-Shape and Association Mode

As with the 5-Stroke input method, Association Mode is activated automatically when you select a character from the candidate list. For details, see Section 7.14.

7.15.9 5-Shape Simple Code Characters

There are three levels of simple code characters; Frequently-Used Characters (Level 1), Level 2, and Level 3.

7.15.9.1 Frequently-Used Characters (Level 1 Simple Code)

Among the 6,000 Chinese characters defined in GB2312-80, there are 25 very frequently used characters. These are defined on the keys [11] to [55]. To input these characters, press the key associated with the character (such as [11], [25]) and then press the Space bar.

The 25 frequently used characters are:

	[*1]	[*2]	[*3]	[*4]	[*5]
[1*]	一	地	在	要	工
[2*]	上	是	中	国	同
[3*]	和	的	有	人	我
[4*]	主	产	不	为	这
[5*]	民	了	发	以	经

7.15.9.2 Level 2 Simple Code

The 5-Shape input method has about 600 Level 2 Simple Code Characters. To input a Level 2 Simple Code Character, enter the first two radicals and then press the Space bar. For example, the 5-Shape code for "帮" is [13][31][52][21] (that is, DTBH). But since it is a Level 2 Simple Code Character, you can press [13][31] and the Space bar to input the character.

The Level 2 Simple Code Characters include:

啊阿爱安暗芭吧百瓣半办帮包保报北

本比笔必避陛边变表宾冰并伯泊不步

部才财采菜餐参惭灿册查产昌长偿吵

炒车晨忱陈称城成呈承持池弛耻赤炽

充抽出处春磁此从粗村达答打大呆代

胆淡当档刀导到得灯邓迪地第帝电佃

甸盯钉锭定订东度断队对多夺朵儿二

罚法凡反贩芳方防妨纺放肥分坟粉丰

风烽冯凤夫服负妇肝敢肛纲高革格蛤

个各给耿攻功公宫贡共垢构估孤姑骨

顾怪关官观管光归轨辊果过害汉好恨

虹红后呼胡虎互画划化怀换蝗煌晃灰

毁会婚伙或圾基机肌吉极及级几季寂

计记际继纪家加尖间艰检渐江匠降胶

交角较叫节杰结介届紧进近睛经景久

九旧舅就居具决军开楷苛科可克客肯

空扣枯宽昆困扩拉来乐累肋类楞离理

李历立力联脸良量辽料列林灵另刘六

龙娄搂卢卤录吕屡率绿罗妈马嘛慢么

玫煤没们眯迷米秘绵面秒民明名末牟

姆睦哪奶南男难内能尼你年宁怕炮朋

批皮平普妻七岂钱前欠强悄峭切且亲

沁轻顷庆秋区曲取权全劝然让认扔仍

如入闰弱洒三扫色砂纱闪商少社审生

失时实史式示世事氏收手守曙术甩霜

双水睡顺说思肆寺四诉虽孙所他它台

太膛提啼天条铁厅听烃瞳同屯驼妥拓

外晚宛汪为委胃卫闻无五物务析吸习

戏细瞎下仙嫌显现限线相向宵小肖协

械懈心信兴行凶胸休秀旭轩喧眩学雪

旬寻巡训押呀牙烟炎眼燕央杨洋阳样

遥药要也业叶衣姨矣亿忆义因阴引隐

蝇哟用由邮友右于愉与玉遇原员约匀

允晕杂灾载早澡灶则增曾赠粘斩崭占

站张涨找折这珍贞针阵争芝支知脂之

职直止只旨志炙中肿轴宙珠朱烛主注

驻妆浊籽子字棕综最昨左作

7.15.9.3 Level 3 Simple Code

Approximately 4,000 Chinese characters defined in GB2312-80 are classified as Level 3 Simple Code Characters. To input these characters, enter the first three radicals and then press the Space bar.

7.15.10 5-Shape Wildcard Key

The Z key on the main keyboard is the wildcard key for the 5-Shape input method. When inputting a single character, the Z key can replace any root radicals about which you are uncertain.

For example, if you are not sure whether "律" should be decomposed into "彳"," ヨ"," 二", and " 丨" (TVFH) or " 彳"," ヨ" ," 丨 ", and " 二" (TVHF), you can enter [31][53][Z][Z] and all characters with starting root radicals [31] and [53] will be displayed for your selection. For example:

五笔字型：tvzz_ 1很 TVEY 2律TVFH 3笋TVTR 4魏TVRC　　　　+

7.15.11 Entering 5–Shape Code Through the Numeric Keypad

The 25 alphabetic keys on the main keyboard (all except the wildcard "Z" key) are used to enter 5-Shape code. You can also enter 5-Shape codes through the numeric keypad.

For instance, the 5-Shape code of "丙" is [11][[25][34][43] (GMWI). To enter "丙", you can press [KP1][KP1][KP2][KP5][KP3][KP4][KP4][KP3] on the numeric keypad. The function of the [KP0] key is similar to the function of the Space bar.

To enter the wildcard code through the numeric keypad, press the [KP6] key twice.

7.16 Pin–Yin Input Method

The 6,000 Chinese characters defined in GB2312-80 are sorted according to the Pin-Yin (phonetic) representation. Each character has its own Pin-Yin representation that you can use to input the character. To use this input method, you should have basic knowledge of Putonghua (Mandarin) and Chinese pronunciation.

The Pin-Yin input method requires each character to be entered by using one to six Roman characters and one tone mark (-, /, v, \, or o).

Only single characters can be entered using the Pin-Yin input method. Term input is not supported. But you can activate Association Mode when selecting a character.

7.16.1 Pin-Yin Input Mechanism

Each Chinese character has a phonetic representation that can be specified in a sequence of syllables. To enter the pronunciation, enter the syllables through the main keyboard (keys A - Z). To enter the tone marks, press the key with preassigned tone marks as defined in Table 7–27.

Table 7–27: Pin-Yin Tone Marks

Tone Mark	Symbol	Key Label
First tone	-	;
Second tone	/	,
Third tone	V	\
Fourth tone	\	[
Light tone	O	]

For example, the phonetic representation of "日" is "ri". Press [R][I].

7.16.2 Pin-Yin Input Procedure

When you invoke the Pin-Yin input method, the string "汉语拼音" is displayed in the status area. When you enter the phonetic representation, the data is displayed in the preedit area. All characters that match the representation are displayed.

For example, when you press the A key, all characters with phonetic representations starting with "a" are displayed in the preedit area:

汉语拼音：a＿ 1啊KSKG 2吖KUHH 3啊KBSK 4锕QBSK ＋

When you press the N key, all characters with phonetic representations starting with "an" are displayed:

汉语拼音：an＿ 1鞍AFPV 2桉SPVG 3厂DGT 4鹌DJNG   ＋

If the phonetic representation is invalid, the bell rings. You can correct the data by using the Delete key or press Return to clear the data.

_____ **Note** _____

> If the phonetic representation has only one character, you must still press the 1 key to select the character.

_____

7.16.3 Pin-Yin Multiple Candidates

If a Pin-Yin string matches multiple candidates, the candidates are displayed in the preedit area. To select one, press the corresponding numeric key. For details, see Section 7.14.

To reduce the number of candidates, you can enter a tone mark after the phonetic representation. The input method displays only those characters with the specified tone.

For example, the phonetic representation of "牛" is "niu". Eight characters are displayed for selection after you press [N][I][U]:

汉语拼音：niu＿ 1妞VNFG 2牛RHK 3扭RNFG 4钮QNFG ＋

If you press the / key after [N][I][U], there will be only one candidate for selection:

汉语拼音：niu2＿ 1牛RHK

_____ **Note** _____

> If you enter a tone mark after entering a phonetic representation, you do not need to press the [1] key to select the character in the first position. The first candidate is selected automatically when you enter another phonetic representation.

_____

7.16.4 Pin-Yin and Association Mode

Similar to the 5-Stroke input method, Association Mode is activated automatically when you select a character from the candidate list. For details, see Section 7.14.

7.16.5 Pin-Yin Multiple Phonetic Representations

It is possible to have multiple phonetic representations for the same character. For example, "调" can be pronounced as "diao" and "tiao". Therefore, you can enter either [D][I][A][O] or [T][I][A][O] for "调".

7.16.6 Pin-Yin Radical Characters

Approximately 20 radical characters are defined in GB2312-80. To input those characters using the Pin-Yin input method, input the phonetic representation of the first character of the radical name. For example, the radical name of "艹" is "cao zi tou" 〔草字头〕. Thus, you would enter [C][A][O].

7.17 Qu–Wei Input Method

The Qu-Wei code is the representation of Qu (row) and Wei (column) in the DEC Hanzi codeset. For example, "啊" is placed in the first column of the sixteenth row. Its Qu-Wei code is 1601. To input "啊", press [1][6][0][1].

Use the Qu-Wei input method to input any characters from the GB18030 and Unicode standards.

Because each Qu-Wei, Hexadecimal, or Unicode code or entity name in the GB18030-2000 standard unambiguously defines one Chinese character, the Qu-Wei input method does not support term input or Association Mode.

The Qu-Wei input method accepts the following input formats:

- 1601 – Qu-Wei codes in the DEC GB2312 character set: input four digits.

- 11601 – Qu-Wei codes in the extended GB character set: input five digits starting with the number 1.

- x7a or X7a – Hexadecimal value: input two, four, or eight digits.

- u5de5 or U5de5 – Unicode values U+5de5: input four or six digits.

- &yen – The entity string name for the yen character.

When you invoke the Qu-Wei input method, the string "国标区位" is displayed in the status area. The input methods for the various codes are:

- To input a Qu-Wei code, enter the code by using the numeric keys (0 - 9) on the main keyboard.

- To input a Hexadecimal code, enter x or X followed by the Hexadecimal value.

- To input a Unicode value, enter u or U followed by the Unicode value. See Section 7.17.3 for a table of Entity Code and Unicode Value equivalents.

- To input an entity, enter & followed by the entity string name.

7.17.1 DEC GB2312 Characters

The character, " 啊 " is defined in the DEC GB2312 character set and its Qu-Wei code is
1601. To input " 啊 ", enter [0][1][6][0][1] or [1][6][0][1] from the main keyboard, and
press the Space bar to complete the input.

7.17.2 Extended GB Characters

The Qu-Wei codes of characters in the extended GB character set start with "1". For
example, if you press [1][1][6][0][1], the user-defined character in the first column of the
sixteenth row is entered.

If you make a mistake during input, you can correct the data by pressing the Delete key to
erase the last key stroke, or the Return key to clear the entire code.

7.17.3 Entity Names and Unicode Values

The Qu-Wei input method allows you to input Unicode values. The following table describes
character entity names and their corresponding Unicode values, as defined by the World
Wide Web Consortium, that may be helpful to you.

Entity Name	Unicode Value	Entity Name	Unicode Value	Entity Name	Unicode Value
nbsp	U+00A0	ocirc	U+00F4	rArr	U+21D2
iexcl	U+00A1	otilde	U+00F5	dArr	U+21D3
cent	U+00A2	ouml	U+00F6	hArr	U+21D4
pound	U+00A3	divide	U+00F7	forall	U+2200
curren	U+00A4	oslash	U+00F8	part	U+2202
yen	U+00A5	ugrave	U+00F9	exist	U+2203
brvbar	U+00A6	uacute	U+00FA	empty	U+2205
sect	U+00A7	ucirc	U+00FB	nabla	U+2207
uml	U+00A8	uuml	U+00FC	isin	U+2208
copy	U+00A9	yacute	U+00FD	notin	U+2209
ordf	U+00AA	thorn	U+00FE	ni	U+220B
laquo	U+00AB	yuml	U+00FF	prod	U+220F
not	U+00AC	fnof	U+0192	sum	U+2211
shy	U+00AD	Alpha	U+0391	minus	U+2212
reg	U+00AE	Beta	U+0392	lowast	U+2217
macr	U+00AF	Gamma	U+0393	radic	U+221A
deg	U+00B0	Delta	U+0394	prop	U+221D
plusmn	U+00B1	Epsilon	U+0395	infin	U+221E
sup2	U+00B2	Zeta	U+0396	ang	U+2220
sup3	U+00B3	Eta	U+0397	and	U+2227
acute	U+00B4	Theta	U+0398	or	U+2228
micro	U+00B5	Iota	U+0399	cap	U+2229
para	U+00B6	Kappa	U+039A	cup	U+222A
middot	U+00B7	Lambda	U+039B	int	U+222B
cedil	U+00B8	Mu	U+039C	there4	U+2234
sup1	U+00B9	Nu	U+039D	sim	U+223C
ordm	U+00BA	Xi	U+039E	cong	U+2245
raquo	U+00BB	Omicron	U+039F	asymp	U+2248
frac14	U+00BC	Pi	U+03A0	ne	U+2260
frac12	U+00BD	Rho	U+03A1	equiv	U+2261
frac34	U+00BE	Sigma	U+03A3	le	U+2264

Entity Name	Unicode Value	Entity Name	Unicode Value	Entity Name	Unicode Value
iquest	U+00BF	Tau	U+03A4	ge	U+2265
Agrave	U+00C0	Upsilon	U+03A5	sub	U+2282
Aacute	U+00C1	Phi	U+03A6	sup	U+2283
Acirc	U+00C2	Chi	U+03A7	nsub	U+2284
Atilde	U+00C3	Psi	U+03A8	sube	U+2286
Auml	U+00C4	Omega	U+03A9	supe	U+2287
Aring	U+00C5	alpha	U+03B1	oplus	U+2295
AElig	U+00C6	beta	U+03B2	otimes	U+2297
Ccedil	U+00C7	gamma	U+03B3	perp	U+22A5
Egrave	U+00C8	delta	U+03B4	sdot	U+22C5
Eacute	U+00C9	epsilon	U+03B5	lceil	U+2308
Ecirc	U+00CA	zeta	U+03B6	rceil	U+2309
Euml	U+00CB	eta	U+03B7	lfloor	U+230A
Igrave	U+00CC	theta	U+03B8	rfloor	U+230B
Iacute	U+00CD	iota	U+03B9	lang	U+2329
Icirc	U+00CE	kappa	U+03BA	rang	U+232A
Iuml	U+00CF	lambda	U+03BB	loz	U+25CA
ETH	U+00D0	mu	U+03BC	spades	U+2660
Ntilde	U+00D1	nu	U+03BD	clubs	U+2663
Ograve	U+00D2	xi	U+03BE	hearts	U+2665
Oacute	U+00D3	omicron	U+03BF	diams	U+2666
Ocirc	U+00D4	pi	U+03C0	quot	U+0022
Otilde	U+00D5	rho	U+03C1	amp	U+0026
Ouml	U+00D6	sigmaf	U+03C2	lt	U+003C
times	U+00D7	sigma	U+03C3	gt	U+003E
Oslash	U+00D8	tau	U+03C4	OElig	U+0152
Ugrave	U+00D9	upsilon	U+03C5	oelig	U+0153
Uacute	U+00DA	phi	U+03C6	Scaron	U+0160
Ucirc	U+00DB	chi	U+03C7	scaron	U+0161
Uuml	U+00DC	psi	U+03C8	Yuml	U+0178
Yacute	U+00DD	omega	U+03C9	circ	U+02C6
THORN	U+00DE	thetasym	U+03D1	tilde	U+02DC
szlig	U+00DF	upsih	U+03D2	ensp	U+2002

Entity Name	Unicode Value	Entity Name	Unicode Value	Entity Name	Unicode Value
agrave	U+00E0	piv	U+03D6	emsp	U+2003
aacute	U+00E1	bull	U+2022	thinsp	U+2009
acirc	U+00E2	hellip	U+2026	zwnj	U+200C
atilde	U+00E3	prime	U+2032	zwj	U+200D
auml	U+00E4	Prime	U+2033	lrm	U+200E
aring	U+00E5	oline	U+203E	rlm	U+200F
aelig	U+00E6	frasl	U+2044	ndash	U+2013
ccedil	U+00E7	weierp	U+2118	mdash	U+2014
egrave	U+00E8	image	U+2111	lsquo	U+2018
eacute	U+00E9	real	U+211C	rsquo	U+2019
ecirc	U+00EA	trade	U+2122	sbquo	U+201A
euml	U+00EB	alefsym	U+2135	ldquo	U+201C
igrave	U+00EC	larr	U+2190	rdquo	U+201D
iacute	U+00ED	uarr	U+2191	bdquo	U+201E
icirc	U+00EE	rarr	U+2192	dagger	U+2020
iuml	U+00EF	darr	U+2193	Dagger	U+2021
eth	U+00F0	harr	U+2194	permil	U+2030
ntilde	U+00F1	crarr	U+21B5	lsaquo	U+2039
ograve	U+00F2	lArr	U+21D0	rsaquo	U+203A
oacute	U+00F3	uArr	U+21D1	euro	U+20AC

7.18 Telex Code Input Method

Telex Code is a four-digit code used to define:

- Single Chinese characters

- Terms

- Non-Chinese symbols

Telex Code is based on the Standard Telex Code (标准电码本) published by the People's Republic of China. The Telex Code input method is simple and straightforward. You can input characters, terms, or symbols by entering their corresponding Telex Code from the keyboard.

When you invoke the Telex Code input method, the string "电报码" is displayed in the status area. Use the numeric keys 0 - 9 on the main keyboard to enter the codes for characters, terms, or symbols:

Character/String	Type	Telex Code	Key Sequence
乎	Character	0039	[0][0][3][9]
十六日	Term	9916	[9][9][1][6]
Ⅸ	Symbol	9949	[9][9][4][9]

7.19 Intelligent ABC Input Method

The Intelligent ABC (ItABC) input method is one of the most popular Chinese input methods because it is easy to learn and usually faster than other input methods. Intelligent ABC supports full and mixed input of words, phrases, and short sentences (Quanpin and Jianpin), coordinated input of Chinese characters and other marks (Hunpin), character shape input (Bixing), and combined input of pronunciation elements and Chinese character shapes (Yinxinghunhe).

The following sections describe the various types of input, the input process, and useful features of Intelligent ABC. For a full description of Intelligent ABC and examples of how best to use this and other input methods of the simplified Chinese Input Server, see the online help for the dxhanziim Input Server.

7.19.1 Basic Input Methods

The Intelligent ABC input method offers you five types of input; Quanpin, Jianpin, Hunpin, Bixing, and Yinxinghunhe.

- Quanpin Input

    If you are familiar with Chinese Pin-Yin, use Quanpin input. The Quanpin input process is the same as the Chinese Pin-Yin writing process if you use standard Chinese Pin-Yin input.

    When using Quanpin input, enter phrases separated by spaces or by marks, such as commas or periods. You can also use normal Pin-Yin input. The system generates an audible warning if the input is greater than the allowed number of characters and marks.

    When you use Quanpin input, pay particular attention to phrases without punctuation or syllable-dividing marks.

- Jianpin Input

    If you are not familiar with Chinese Pin-Yin, use Jianpin input.

Under Jianpin input, you use the first letter of each syllable. However, for the syllables zh, ch, and sh, use the first two letters. Also, when you select Jianpin input, make frequent use of syllable-dividing marks.

- Hunpin Input

  Hunpin input is open and comprehensive, allowing a mix of full and simplified input. For example, when a phrase has more than two syllables, some syllables can use Quanpin and some syllables use Jianpin.

  Keep in mind that syllable-dividing marks are very important when you select Hunpin input.

- Bixing input

  With Bixing input, Intelligent ABC follows the Bixing code of eight strokes. The eight strokes are:

  - heng (including ti)
  - shu (including shuwangou)
  - pie
  - dian and na
  - zhe (clockwise)
  - wan (counterclockwise)
  - cha
  - fang

  To determine the Bixing code, follow the order of strokes observed in calligraphy, but remember that you are limited to a maximum of six strokes. For more information on Bixing code and Chinese characters, see the online help for the dxhanziim Input Server.

- Yinxinghunhe input

  If you are familiar with the Intelligent ABC input method, select Yinxinghunhe input. With Yinxinghunhe input, you can combine Pin-Yin, and Bixing input (pronunciation and shape, respectively) to avoid overlapping codes in Quanpin or Jianpin input.

  Yinxinghunhe input has the following format (where "pinyin" can be Quanpin, Jianpin, or Hunpin input):

  (pinyin + [bixing description]) + (pinyin + [bixing description])
  + …… + (pinyin + [bixing description])

You must use "pinyin", pronunciation input, when you input multisyllable words. You can, however, omit the [bixing description] value for shape input. When used, [bixing description] must not be more than two strokes.

When you input single syllable words or characters, you can specify the [bixing description] value alone. Also, you can mix "pinyin" and [bixing description] input to avoid overlapping codes in Quanpin and Jianpin input.

7.19.2 Input Process

The input process you use can take one of the following forms:

- Basic input process

  The basic input process is a three-step process of start, middle, and finish.

  Start – Press a key to enter the Pin-Yin process. This initial key is one of the 26 letters (uppercase or lowercase) from the English alphabet. If the initial key is the letter i, I, u, or v, the input has special meaning (see Section 7.19.4).

  Middle – Input sentences. You can use any characters or marks, including numerals, as sentence input. However, for standard alteration, sentence input should conform to the combination rules.

  Finish – Enter a key signifying end of input. Use a space or punctuation mark to signify the end of word input. Use the Enter key to signify character input. Under special circumstances, use [,] or CTRL+-.

  If there is more than one selection, use the display of candidates on the selection window. If alteration cannot be completed, the division and formation of words is retained in memory.

- Customize the Intelligent ABC input method

  See Section 7.5.2.7 for more information.

- Define New Words

  In the Input Method Customization window, you can click on the customization button to display a pull-down menu. The pull-down menu includes an option labeled, Define New Words by Intelligent ABC input method.

  If you choose this option, a dialog box displays that allows you to add or delete words and phrases. If you add words and phrases, they are saved in the file of user-defined words and phrases. If you input the outer codes of words and phrases after specific alteration of "u" type, the corresponding words and phrases are returned to the application program.

- Keyboard Definition

  See the online help for the dxhanziim Input Server for an example of Input Method Keyboard Definition.

7.19.3 Intelligent ABC Input Method Features

The Intelligent ABC input method provides the following features. The online help for the Simplified Chinese Input Server, dxhanziim, fully describes the use of these features.

- Automatic Dividing and Forming of Words

  Automatic dividing of words refers to the process of dividing the continuous input of a Pin-Yin passage into several simple paragraphs and converting those paragraphs into Chinese phrases according to grammatical rules.

  Forming of words refers to the process of combining several words and word elements into a new word.

- Automatic Memory

  Automatic memory is usually used to retain new words such as the names of people or places that do not exist in the word palette. The process is done automatically or with little user intervention. The words of automatic memory are all standard Pin-Yin words and can be used in the same manner as the words in the basic word palette.

- Forcible Memory

  Use Forcible Memory to define nonstandard Chinese Pin-Yin words. With this function, you can add new words to your word palette directly.

  To retain a new word, you must input both the contents of the word and the code for it. The contents of the word can be Chinese words, expressions, and phrases, and it can consist of Chinese characters and other marks. The code for a word can be the Chinese Pin-Yin, original foreign language words, or any other marks you wish to use.

  The nonstandard words you define must not be greater than 15 characters. The word code you define must not be greater than 9 marks. You can define up to 400 words.

- Obscure Recall

  This function allows you to use incomplete information to recall a word that was recently used. You use the CTRL+- key to operate the Obscure Recall function.

  Obscure Recall is useful when you enter a simple passage containing repetitious content.

- Frequency Adjustment and Memory

  The frequency of a word refers to the frequency of word usage. The word order of homonyms in the Intelligent ABC standard word palette reflects the regular pattern of its usage. But an individual user may need to adjust the default word order. The Frequency Adjustment and Memory function allows you to make these adjustments under the Intelligent ABC input method.

  If you select Word Frequency Adjustment in the property setting, the Frequency Adjustment and Memory function begins automatically.

  Because the system uses the most frequently used word as the recognized alteration result, you can use this feature to adjust this frequency and, thus, the alteration result.

  The Frequency Adjustment and Memory function handles a word length range of 1 to 3 syllables. For words of single syllables, the word frequency is not adjusted until the words are used two times.

- Leading Characters and Succeeding Characters

  The system treats a leading character (or prefix) and a succeeding character (or suffix) automatically in the process of forming new words. Some characters can be used only as leading characters and some words can be used only as succeeding characters. The dxhanziim online help describes these cases.

  Intelligent ABC can distinguish leading and succeeding characters and combines the words automatically in the process of input.

- Word Palette System

  Intelligent ABC has a basic word palette of approximately 60,000 words.

  In addition, Intelligent ABC has an automatic screening function. The vocabulary for Automatic Memory is up to 17,000 words and the vocabulary for Forcible Memory is up to 400 words.

7.19.4 Intelligent ABC Usage Notes

Consider the following when using the Intelligent ABC input method. These usage notes are covered in detail by the dxhanziim online help.

- The Intelligent ABC input method allows you to convert Arabic numerals to Chinese numerals and to simplify input of some common classifiers. See the dxhanziim online help for details on transformation of punctuation marks, letters in numeral input, and normal and emphasized Chinese numerals.

- You can use a multisyllable word to reduce the number of candidates in Chinese character input. The multisyllable word you use can be from the standard palette or from a user-defined palette. The dxhanziim online help provides an example of this.

- If you use the Forcible Memory function to define a word in advance, start your input with the letter u. Likewise, to input figures and marks, in the normal state enter the letter v and the appropriate number (1 through 9) of the GB-2312 set of characters and marks.

- To type an English word or phrase during input of Pin-Yin, enter v as an indication mark, then type the English word or phrase, and terminate with the space key.

- The input of Chinese punctuation marks is the same among different input methods. (See the dxhanziim online help for information on automatic input of shuminghao under Intelligent ABC.)

7.20 Symbol Input in simplified Chinese

When you press [Ctrl/0] in the 5-Stroke, 5-Shape, or Pin-Yin mode, the string "符号" is displayed in the status area. All letters and numerals that you enter are converted to two-byte letters and numerals.

To return to the original input mode, press [Ctrl/0] again.

7.21 Conversion Between Input Servers and DECwindows Motif Applications

If you need to work with both simplified and traditional Chinese data, you can use DECwindows Motif to enter simplified Chinese data with the traditional Chinese input server, and vice versa.

To set up the environment for entering simplified Chinese data with the traditional Chinese input server, you should:

1. Set the session language to Chinese China.

2. Set the XMODIFIERS environment variable to select the Chinese input server:

   ```
   %setenv XMODIFIERS @im=DECTW
   ```

Hereafter, applications start in simplified Chinese and are connected to the traditional Chinese input server. You can enter simplified Chinese data using the input methods provided by the Chinese input server, such as the Tsang–Chi input method.

If you do not set this modifier and both simplified and traditional Chinese input servers are running, a Chinese application connects to the first input server that is started.

The data in the preedit area of the Chinese input server is displayed in simplified Chinese.

Those Chinese characters that cannot be converted to simplified Chinese characters are represented by a reverse question mark. If two or more user-defined characters have the same input key sequence, two or more reverse question marks might appear near the end of the line. In that case, you should try each one to find the character you want.

Use the UDC Manager (cedit) to define the user-defined characters and their corresponding input key sequences. Since a VT382-D terminal is a DEC Hanyu terminal, you can create an input key sequence file only under the DEC Hanyu locale. Generation of an input key sequence file under the Taiwanese EUC and BIG-5 locales is not supported in this release.

_____ **Note** _____

Use the dxterm terminal emulator to run the UDC Manager Utility. The dtterm terminal emulator does not support `cedit` functions and the UDC Manager Utility may hang under dtterm.

_____

To set up the environment for entering traditional Chinese data with the simplified Chinese input server:

1. Set the session language to Chinese Taiwan.

2. Set the XMODIFIERS environment variable to select the simplified Chinese input server:

    ```
    %setenv XMODIFIERS @im=DECCN
    ```

Hereafter, applications start in Chinese and are connected to the simplified Chinese input server. You can enter Chinese data using the input methods provided by the simplified Chinese input server, such as the 5-Shape and the 5-Stroke input methods.

8

Chinese Printing Support

This chapter introduces the Chinese printing support provided by Tru64 UNIX. It describes the supported printers, print file formats, features, and maintenance procedures for supporting Chinese printing.

8.1 Supported Printers

Tru64 UNIX supports text and PostScript printers.

8.1.1 Text Printers

Tru64 UNIX supports text printers with built-in Chinese fonts.

8.1.2 PostScript Printers

Tru64 UNIX supports Chinese printing on PostScript printers in the following ways:

- Based on the built-in or downloaded fonts installed in printers

- Based on the *font-faulting* mechanism, which is explained in Printing Features

- Based on the font-embedding mechanism

For details about the supported printer types and print filters, see Chapter 4.

8.2 Print File Formats

Tru64 UNIX supports printing mixed ASCII and Chinese characters in the following print file formats:

- Plain text files on text printers and PostScript printers

- Files with the nroff control sequences (for printing with underline, superscript, subscript, and bold attributes) on text printers and PostScript printers

- PostScript files on PostScript printers

The print filters for PostScript printers can automatically detect the format of a print file and convert it to the proper format for printing.

8.3 Printing Features

Tru64 UNIX supports the following printing features:

- Font Embedding
- Font Faulting
- Software On–Demand Font Loading
- Codeset Conversion
- Outline Fonts

8.3.1 Font Embedding

Font embedding refers to a mechanism where all the necessary font data is embedded within the printer input file. It is not necessary for printers to have resident fonts. This mechanism is supported only on PostScript printers that support PostScript level 2 or level 1 with multibyte font extension. Outline fonts are used for the embedding, if available. Otherwise, lower quality bitmap fonts are used.

The font-embedding mechanism is useful since few printers have Chinese resident fonts. However, there must be enough memory inside the printer to hold the font data or the print job might fail. The amount of memory required depends on the nature of the print job. It is recommended that at least 4 MB or more printer memory be available for the embedding mechanism.

Because of the large amount of data transferred to the printer, it is also necessary to use a high-speed link between the host computer and the printer. Using the serial port for printing is too slow for the font-embedding mechanism.

See wwpsof(8) for more information about how to use the font embedding mechanism.

8.3.2 Font Faulting

Font faulting is a mechanism used to handle the large memory required by fonts for some codesets, particularly multibyte codesets for Asian languages. Using font faulting, font information is stored on either:

- The secondary storage of a supporting host machine, called a font-faulting server, or
- An internal font disk

The font information is loaded into the printer on demand, thus conserving printer memory.

Font faulting is often essential for multibyte ideographic fonts because the memory required to store a single font can exceed the memory capacity of many printers. Specialized local language printers, such as Japanese printers, do not require font faulting because the local language fonts reside on the printer. However, other printers require a mechanism to load these fonts as needed for different parts of the same print job.

The font-faulting mechanism is very useful for a desktop printing environment, in which a large number of different single-byte fonts may be required. In this case, simultaneously storing all the fonts in memory reduces the available memory, and therefore speed, of the printer. Also, it is possible that the number of required fonts is so large that they cannot all be in memory at the same time.

Font faulting for multibyte fonts is done on a per character (or per glyph) basis because these fonts support extremely large numbers of characters. Font faulting for single-byte fonts is done on a per font basis. Single-byte fonts are small and relatively simple, so loading the whole font is more efficient.

The font-faulting mechanism can be used with the following printers:

- DEClaser 1152
- DEClaser 5100
- PrintServer 17

See Section 8.5 for information about configuring these printers.

8.3.3 Software On–Demand Font Loading

Software On-Demand Font Loading (SoftODL) is a mechanism through which a terminal or a bitmap printer downloads the relevant bitmap font information for a user-defined character (UDC) at the time the character needs to be displayed or printed. The Chinese bitmap printers that support this feature include:

- CP382D controller (for traditional Chinese)
- LA88-C (for simplified Chinese)
- LA380-CB (for simplified Chinese)

8.3.4 Codeset Conversion

Tru64 UNIX includes a codeset conversion mechanism used to print text files that have a codeset different from the one used by the printer. For printers with built-in or downloaded Chinese fonts, the codeset of the printer should be defined to match the codeset of the built-in fonts. For printers using the font-faulting mechanism, the codeset of the printer should be defined to match the codeset of the font to be loaded. For printers using the font-embedding mechanism, the codeset of the printer should not be defined.

8.3.5 Outline Fonts

Tru64 UNIX provides a large set of outline fonts for printing files in various languages. Depending on how many local language support subsets are installed on your system, more than 150 outline fonts may be available.

There are four sets of Chinese outline fonts, two for traditional Chinese and two for simplified Chinese. These fonts are:

- Sung-Light-CNS11643 (traditional)

- Hei-Light-CNS11643 (traditional)

- Hei-GB2312-80Xi (simplified)

- XiSong-GB2312-80 (simplified)

Those fonts with the CNS11643 extension are traditional Chinese fonts encoded in the DEC Hanyu codeset, with glyphs for plane 1 and plane 2 characters. Those fonts with the GB2312-80 extension are simplified Chinese fonts encoded in DEC Hanzi.

8.4 Printing Commands and Daemons

Before you can utilize the printing features supported by Tru64 UNIX, there are some commands and daemons that you should understand. The following sections discuss these commands and daemons and illustrate how they are used for configuration.

8.4.1 Country-Specific Options to the lpr Command

In addition to the usual options to the `lpr` command, the `-A` option is used to pass country-specific parameters. You can use `ya` to set the parameters to the `-A` option in the `/etc/printcap` file. For example, you can specify the parameters using the `-A` option to the `lpr` command as:

```
% lpr -A "flocale=zh_TW.big5 font=Sung-Light-
CNS11643 plocale=zh_TW.dechanyu" <file>
```

You can define the same set of parameters in the `/etc/printcap`, file as:

```
:ya="flocale=zh_TW.big5 font=Sung-Light-
CNS11643 plocale=zh_TW.dechanyu":\
```

The parameters supplied with the `-A` option to the `lpr` command override the corresponding default values in the `/etc/printcap` file.

The following parameters are applicable to Chinese printing:

- `flocale=<any valid locale>`

  Specifies the locale for the source text file. The printer filters use this locale to validate the characters inside the source text file. If this value is not set properly, the text is interpreted using the current locale. In Chinese printing, this value is particularly important in order for the `lpr` command to correctly interpret the characters. Moreover, if the `plocale` option is also set, the `lpr` command performs codeset conversion for the source text file.

- `plocale=<any valid locale>`

  Specifies the locale for the printer. If the printer has built-in fonts, the `plocale` value should match the codeset of the built-in fonts. If the printer employs the font faulting mechanism, the `plocale` value should match the font used to print the text file.

- `font=<supported outline font>`

  Specifies the font name for printing the source text files on a PostScript printer. This parameter is used for printing text files only, as PostScript files are already tagged with the required font name.

- `odldb=<odl database path>`

  Specifies the path of the SoftODL database files. This parameter is used to override the system default SoftODL database path, thus allowing users to access their own SoftODL database.

- `odlstyle=<odl style and size>`

  Specifies what SoftODL font style and size to use. The value is of the form <style>-<NxN> (that is, normal-24x24). If not specified, the system default SoftODL style and size are used.

- `line=<number of lines>`

  Specifies the number of lines per page. You can use this parameter with the `-w` command to control the font size and orientation of the output.

8.4.2 PostScript Font Management Utility (pfsetup)

A PostScript font management utility, `pfsetup`, is provided to help you set up print queues to use the font-faulting mechanism. This utility has the following syntax:

`pfsetup [-s] [-d] [queue_name...]`

The following options can be used:

-s   setup mode — Allows you to set up lists of fonts to be downloaded

-d   download mode — Downloads fonts to printers according to the lists prepared
       with the -s option

If you do not specify an option, pfsetup displays the information about the print queues that have been set up with this utility. If you do not specify a particular queue name, the pfsetup command processes every applicable queue.

In setup mode, the pfsetup command displays all PostScript printer fonts available on the system and prompts you to select the fonts or font headers to be downloaded onto individual print queues:

```
% /usr/sbin/pfsetup -s

=========================================================
Printer queue: lp1 | 1 | dl1152w

No font has been setup for downloading in queue lp1

These are fonts available in your system for downloading. Fonts
chosen for downloading are marked with *

1 Hei-Light-CNS11643            2 Sung-Light-CNS11643
3 Hei-GB2312-80                 4 XiSong-GB2312-80
5 XiSong-GB2312-80              6 AngsanaUPC-Bold@
7 AngsanaUPC-BoldItalic@        8 AngsanaUPC-Italic
9 AngsanaUPC-Light             10 CordiaUPC-Bold
11 CordiaUPC-BoldItalic        12 CordiaUPC-Italic
13 CordiaUPC-Light             14 EucrosiaUPC-Bold
15 EucrosiaUPC-BoldItalic      16 EucrosiaUPC-Italic
17 EucrosiaUPC-Light           18 FreesiaUPC-Bold
19 FreesiaUPC-BoldItalic       20 FreesiaUPC-Italic
21 FreesiaUPC-Light            22 IrisUPC-Bold
23 IrisUPC-BoldItalic          24 IrisUPC-Italic
25 IrisUPC-Light               26 JasmineUPC-Bold


[C]ontinue | [S]etup | [L]ist fonts | [Q]uit | [N]ext queue <C>
```

The action keys provided by pfsetup have the following meaning:

Key	Action
[C]ontinue	Displays a further font listing
[S]etup	Proceed to setup
[L]ist fonts	List fonts again
[Q]uit	Quit pfsetup
[N]ext queue	Proceed to the next queue

If you choose the s option, another prompt is displayed:

```
[A]dd fonts | [R]emove fonts | [L]ist fonts | [Q]uit |
[N]ext queue <N>
```

In response to the prompt, you can enter a to add fonts to the list for a print queue or enter r to delete fonts from the list for a print queue. The fonts you select are highlighted with an asterisk (*).

_____ **Note** _____

The fonts on this list vary according to the language variants that have been installed on your system.

_____

To download the fonts or font headers selected in setup mode, use the pfsetup command with the -d option.

For fonts of ideographic character sets, the pfsetup utility downloads only their font headers. Data defining their font glyphs is downloaded only on an as-needed basis through the font faulting mechanism. This saves printer memory. For single-byte fonts, the utility downloads the entire font onto printers for efficiency.

8.4.3 Font-Faulting Daemon (ffd)

To handle font-faulting requests from a PostScript printer with the two channel approach, (such as the DEClaser 1152, Section 8.5.2), the font faulting-daemon, ffd, must be running on your system. When the daemon receives a font data request, it extracts the required font glyph data from the specified font and sends it to the printer through the secondary channel.

If you configure a print queue that uses the two channel approach, or modify the secondary channel of a print queue, you must restart the font-faulting daemon.

To restart the font-faulting daemon, log in as a superuser and stop the existing font-faulting daemon with the following command:

```
% /sbin/init.d/ffserver stop
```

To start the font-faulting daemon, you can enter:

```
% /sbin/init.d/ffserver
```

8.4.4 PrintServer Printing Command (wwlpspr)

To fully utilize the features of PrintServer 17 such as two-sided printing, multiple pages per side, and so on, it is necessary to use the lpspr command provided by PrintServer Software Version 5.0 or later for Tru64 UNIX. This command, however, does not provide the features like locales and fonts for printing text files. To support printing files to PrintServer 17, the Tru64 UNIX software provides a unified wwlpspr command.

The `wwlpspr` command is a front-end program that parses the parameters passed by users and calls different commands such as print filter, `lpr`, or `lpspr` with the appropriate command. This provides users with a unified interface.

For details on the `wwlpspr` command, see `wwlpspr(1)`.

8.5 Chinese Printer Setup

This section describes how to set up the following printers to print Chinese characters:

- Dot Matrix Printers

- DEClaser 1152

- DEClaser 5100

- PrintServer 17

- Generic PostScript Printers

8.5.1 Dot Matrix Printers

The only consideration in configuring the CP382-D, LA88-C, and LA380-CB printers is whether ODL printing should be enabled or not, and, if enabled, the default ODL database path and style.

To configure any of the Chinese dot matrix printers:

1. Add a printer using the `lprsetup` command and specify the `cp382d`, `la88c`, or `la380cb` printer.

   The following prompt is displayed:

   ```
   Do you want to enable ODL? [n]
   ```

2. Answer `y` if you want ODL printing, and `n` if you do not.

   If you answer `y`, the default ODL database and style will be used and the following prompt is displayed:

   ```
   Enter symbol name:
   ```

3. To change the default, enter `ya` at the prompt. The following prompt is displayed:

   ```
   Enter a new value for symbol 'ya'?
   ["plocale=zh_TW.dechanyu"]
   ```

4. Enter the following string at the prompt, using the pathname and style of your ODL:

```
"plocale=zh_TW.dechanyu odldb=<path to the default ODL db>
odlstyle=<default ODL style>"
```

Remember the double quotation marks around the parameters, and replace the
`plocale` with the value found in the previous prompt.

You can now use the `lpr` command to send Chinese text files to print queues connected to
these dot matrix printers. You can either set the LANG environment variable or use the
`-A` option to denote the codeset of the text files. For example, the following command prints
the file encoded in the Taiwanese EUC codeset:

```
% lpr -A "flocale=zh_TW.eucTW" my.file1
```

You can override the default ODL path by including the `odldb` and `odlstyle` options
to the `-A` option of the `lpr` command. For example, the following command uses the ODL
database in the `/usr/priv/odl` directory:

```
% lpr -A "odldb=/usr/priv odlstyle=normal-24x24" my.file2
```

8.5.2 DEClaser 1152

The DEClaser 1152 printer can be used to print Chinese characters by using the font faulting
mechanism with two communications channels; one channel is for normal data and the
second channel is for font-faulting data, as shown in Figure 8–1.

**Figure 8–1: Two-Channel Communication of the Font-Faulting
Mechanism**
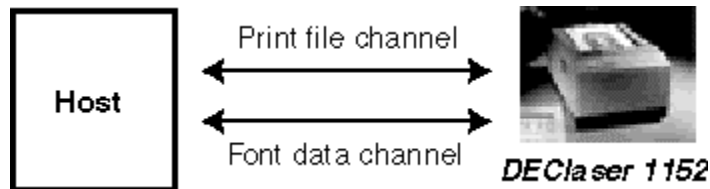
Font faulting requires 4MB of printer memory. If your DEClaser 1152 printer has only 2MB
of memory, you must install the LN07X-UF memory board to provide the additional 2MB of
memory. See the printer manual for information about installing the LN07X-UF memory
board.

You also must establish one and only one system to be the font-faulting server for the printer.
This server sends font information to the printer through a secondary communication
interface, or channel. The printer's secondary channel connection to the font-faulting server
can be through either a local port or a Local Area Transport (LAT) port. If the connection is
through a LAT port, make sure that no other applications or hosts are using that port.

An 8 pin Din to 6 Pos MMJ Adapter is needed to convert the Apple-talk interface on the printer to be the secondary channel that the font-faulting mechanism uses. The baud rate of the secondary interface should match the value of the `$BAUD` variable in the `/sbin/init.d/ffserver` file. By default, this value is 9600 baud.

To configure the DEClaser 1152 to print Chinese files:

1. Add a printer using the `lprsetup` command and select the `dl1152w` printer type. The following prompt is displayed:

   ```
   Do you want to configure this machine as font faulting
   server for the DEClaser 1152 printer? One and only one font
   faulting server should be configured for every DEClaser
   1152 employing font faulting mechanism. [n]
   ```

2. If this machine is the font-faulting server for the printer, answer `y` to the prompt, and then enter the full pathname of the port used to connect the secondary channel for the `ya` capability.

   The following prompt is displayed:

   ```
   Do you want to set up the printer codeset or default font
   for printing non-ASCII text?  If your answer is 'y', please
   consult printcap.4 for the parameters. *** Remember to
   enclose the parameters with a pair of double quotes. [n]
   ```

3. Answer `y` to the prompt, and then define `plocale` and `font` in one of the following ways:

   ```
   "plocale=<Chinese locale> font=<font>"
   ```

   - Define `plocale` as `zh_TW.dechanyu` if you are printing traditional Chinese characters, and define `font` as either `Sung-Light-CNS11643` or `Hei-Light-CNS11643`.

   - Define `plocale` as `zh_CN.dechanzi` if you are printing simplified Chinese characters, and define `font` as either `XiSong-GB2312-80` or `Hei-GB2312-80`.

4. Use the `pfsetup` command to download the fonts. For greatest efficiency, download all the Chinese fonts unless there are some you will never use, plus the most frequently used single-byte fonts. The printer can access the fonts that are not manually downloaded, but there is overhead for dynamically downloading fonts.

5. If you configured this machine as the font-faulting host, enter the following commands to notify the font-faulting daemon about the new printer:

```
% /sbin/init.d/ffserver stop
% /sbin/init.d/ffserver start
```

You need to enter these commands only after you add or modify the printer queue. The next time you reboot the server, the font-faulting daemon will locate the printer.

6. Each time the font-faulting server or the DEClaser 1152 is restarted, you should use the `pfsetup` command to download the fonts again.

For details about the font-faulting daemon and the `pfsetup` utility, see Section 8.4.

To send Chinese text files to print queues connected to a DEClaser 1152, use the `lpr` command. You can either set the LANG environment variable or use the `-A` option to denote the codeset of the text files. For example, the following command prints the file encoded in the Taiwanese EUC codeset:

```
% lpr -A "flocale=zh_TW.eucTW" my.file1
```

You can override the default `plocale` and `font` setting in the `/etc/printcap` file using the `-A` option of the `lpr` command, but be sure the font you use matches the codeset of `plocale`. For example, the following command gets the file encoded in Taiwanese EUC, converts it to DEC Hanzi, and prints it using the XiSong-GB2312-80 font:

```
% lpr -A "flocale=zh_TW.eucTW
plocale=zh_CN.dechanzi font=XiSong-GB2312-80" my.file2
```

8.5.3 DEClaser 5100

The DEClaser 5100 printer can be used to print Chinese characters by using the font-faulting mechanism with built-in hard disk. The LN90X-HD model that supports the font-faulting mechanism also includes the 128 MB hard disk option. The printer must have at least 6 MB of memory.

To configure the DEClaser 5100 to print Chinese files:

1. Add a printer using the `lprsetup` command and select the `dl5100w` printer type.

The following prompt is displayed:

```
Do you want to set up the printer codeset or default font
for printing non-ASCII text?  If your answer is 'y', please
consult printcap.4 for the parameters. *** Remember to
enclose the parameters with a pair of double quotes. [n]
```

2. Answer y to the prompt and define the `plocale` and `font` in one of the following ways:

```
"plocale=<Chinese locale> font=<font>"
```

- Define `plocale` as `zh_TW.dechanyu` if you are printing traditional Chinese characters, and define `font` as either `Sung-Light-CNS11643` or `Hei-Light-CNS11643`.
- Define `plocale` as `zh_CN.dechanzi` if you are printing simplified Chinese characters, and define `font` as either `XiSong-GB2312-80` or `Hei-GB2312-80`.

3. Use the `pfsetup` command to download the fonts. For greatest efficiency, download all fonts that you expect the printer to need; any fonts not manually downloaded are not accessible to the printer.

You need to download fonts only once. The fonts remain on the printer until they are manually removed or the hard disk is reformatted.

For details about the `pfsetup` utility, see Section 8.4.

To send Chinese text files to print queues connected to a DEClaser 5100, use the `lpr` command. You can either set the LANG environment variable or use the `-A` option to denote the codeset of the text files. For example, the following command prints the file encoded in the Taiwanese EUC codeset:

```
% lpr -A "flocale=zh_TW.eucTW" my.file1
```

You can override the default `plocale` and `font` setting in the `/etc/printcap` file by using the `-A` option to the `lpr` command, but be sure the font you use matches the codeset of the `plocale`. For example, the following command gets the file encoded in Taiwanese EUC, converts it to DEC Hanzi, and prints it using the XiSong-GB2312-80 font:

```
% lpr -A "flocale=zh_TW.eucTW
plocale=zh_CN.dechanzi font=XiSong-GB2312-80"
my.file2
```

8.5.4 PrintServer 17

The PrintServer 17 printer can be used to print Chinese characters by using the font-faulting mechanism through the network. There are no special hardware requirements.

To configure the PrintServer 17 to print Chinese files:

1. Install the PrintServer Software Version 5.0 or later for Tru64 UNIX. This is a layered product, purchased separately. See the installation guide for installing and configuring the PrintServer software.

2. If you want to define default the printer codeset or default font to print Chinese text, use the `lprsetup` command or manually update the `/etc/printcap` file to include the `ya` option. Add the following string to the file:

    ```
    :ya="plocale=<Chinese locale> font=<font>":\
    ```

    - Define `plocale` as `zh_TW.dechanyu` if you are printing traditional Chinese characters, and define `font` as either `Sung-Light-CNS11643` or `Hei-Light-CNS11643`.

    - Define `plocale` as `zh_CN.dechanzi` if you are printing simplified Chinese characters, and define `font` as either `XiSong-GB2312-80` or `Hei-GB2312-80`.

3. Use the `pfsetup` command to define the fonts to be downloaded. For greatest efficiency, download all Chinese fonts unless there are some you will never use, plus the most frequently used single-byte fonts. The printer can access the fonts that are not manually downloaded, but there is overhead for dynamically downloading fonts.

    Note that you should execute the `pfsetup` command in the Configuration Host of the PrintServer Printer. After executing `pfsetup`, restart the Management Client to initialize the new configuration. See the PrintServer Software documentation for further details.

4. Turn off the PrintServer 17 and then turn it on. The fonts defined in step 3 will be downloaded automatically.

    Hereafter, each time you restart the PrintServer 17, the fonts download automatically. Thus, the `pfsetup` command is executed only once as long as you do not modify the configuration.

For details about the `pfsetup` utility, see Section 8.4.

To send Chinese text files to print queues connected to a PrintServer 17, use the `wwlpspr` command. You can either set the LANG environment variable or use the `-A` option to denote the codeset of the text files. For example, the following command prints the file encoded in the Taiwanese EUC codeset:

```
% wwlpspr -A "flocale=zh_TW.eucTW" my.file1
```

You can override the default `plocale` and `font` setting in the `/etc/printcap` file by using the `-A` option of the `wwlpspr` command, but make sure the font you use matches the codeset of the `plocale`. For example, the following command gets the file encoded in Taiwanese EUC, converts it to DEC Hanzi, and prints it using the XiSong-GB2312-80 font:

```
% wwlpspr -A "flocale=zh_TW.eucTW
plocale=zh_CN.dechanzi font=XiSong-GB2312-80" my.file2
```

8.5.5 Generic PostScript Printers

Any PostScript printer that supports PostScript level 2 or level 1 with multibyte font extension and sufficient memory can be used to print Chinese using the font-embedding mechanism.

To configure a printer to use font embedding, use the `lprsetup` or `printconfig` utility to set up a print queue using the `wwpsof` filter. You do not need to specify the printer locale as this mechanism can print any language as long as the fonts are available in the system. See `wwpsof(8)` for information on how to use the features available for this print filter.

9

Other Chinese Features

This chapter describes features specific to the Chinese locale in Tru64 UNIX that are not described elsewhere.

9.1 Phrase Support in the VT382–D

You can download up to 100 phrase definitions into the built-in memory of the VT382-D traditional Chinese terminal. You can create a phrase definition file containing the definitions and then download the file to the terminal through the serial port.

_____ **Note** _____

The information presented in this section is not applicable to the system Phrase Utility or the dxim Phrase input method discussed in Chapter 7.

_____

9.1.1 Creating a Phrase Definition File

Each phrase definition file can contain up to 100 phrase definitions. You can create the file using any editor (such as vi) that allows you to edit Chinese data.

9.1.2 Syntax of Phrase Definitions

Phrase definitions have the following syntax:

```
DCS Pc SP v phrase-code / phrase-data ST
```

Table 9–1 lists the parameters.

Table 9–1: Phrase Definitions

DCS	A phrase identifier defined by Tru64 UNIX. Its hexadecimal code is 90. For the 7-bit environment, you can use ESC P (hexadecimal code 1B 50) instead.
Pc	A parameter that controls whether the old phrase definitions in the memory should be cleared before a new one is downloaded. If Pc is equal to zero or is omitted, the old phrase definitions are kept. They are cleared if Pc equals 1.
SP	A space character.
V	A lowercase v.
Phrase-code	A phrase code is a string of up to 8 alphanumeric characters. Uppercase and lowercase letters are regarded as same characters.
/	A slash character separates a phrase code from its phrase data.
phrase-data	A phrase containing up to 80 characters. Characters can be Chinese characters, English letters, numerals, or printable symbols.
ST	An identifier that signals the end of the DCS statement. Its hexadecimal code is 9C. For the 7-bit environment, you can use ESC \ (hexadecimal code 1B 5C) instead.

The following examples show two phrase definition files for the 8-bit and 7-bit environments respectively.

- For the 8-bit environment

  <DCS>1 vBL/提貨單<ST>
  <DCS>0 vBW/保稅倉庫<ST>
  <DCS>0 vBTT/銀行電匯<ST>
  <DCS>0 vCBC/中央銀行<ST>
  <DCS>0 vCH/票據交換所<ST>
  <DCS>0 vCL/託收<ST>
  <DCS>0 vCM/佣金<ST>
  <DCS>0 vCPD/運費付訖<ST>
  <DCS>0 vCWO/憑票即付<ST>
  <DCS>0 vFAS/船邊交貨<ST>
  <DCS>0 vLC/信用狀<ST>

- For the 7-bit environment:

  In this example, the second slash in each phrase definition is regarded as part of the phrase definition:

  <ESC>P1 vAMBASSAD/國賓大飯店/AMBASSADOR<ESC>\
  <ESC>P0 vASIA/環亞大飯店/ASIA WORLD PLAZA<ESC>\
  <ESC>P0 vBROTHER/兄弟大飯店/BROTHER<ESC>\
  <ESC>P0 vCENT/世紀大飯店/CENTURY PLAZA<ESC>\
  <ESC>P0 vFORTUNA/富都大飯店/FORTUNA<ESC>\

9.1.3 Phrase Downloading

The procedure for transferring phrase definitions from a disk file to the terminal is called downloading. The downloaded phrases are kept in the terminal memory as long as the terminal is powered on. Consequently, a phrase definition file needs to be downloaded only once in a terminal session.

To download a phrase definition file to a terminal, display the file onto a terminal using the Tru64 UNIX cat command. You can also download phrase definitions using the Phrase Utility.

The following situations may occur during downloading:

- If a phrase code is duplicated, the new phrase definition replaces the old one.

- If a phrase code contains more than eight characters, the phrase definition is not accepted.

- If the ST code is typed incorrectly, all the characters from the first slash to the correct ST code (with the exception of DCS) are treated as part of the phrase. For example:

  <DCS>**1** vPHONETIC/注音輸入法AT
  <DCS>**0** vINTERNAL/內碼輸入法<ST>

  The AT at the end of the first line is incorrect. If you enter the phrase code "PHONETIC", the following string is input:

  注音輸入法AT0 vINTERNAL/內碼輸入法

- If there are more than 100 phrase definitions, the phrase definitions beyond the limit are ignored.

9.2 Sorting Utility

The methods for sorting Chinese characters are shown in Table 9–2 and Table 9–3:

Table 9–2: Traditional Chinese Sorting Methods

Sorting Method	Full Option Name	Short Option Name
Internal code	Code	C
Phonetic	Phonetic	P
Radical	Radical	R
Stroke	Stroke	S

Table 9–3: Simplified Chinese Sorting Methods

Sorting Method	Full Option Name	Short Option Name
Qu-Wei	Quwei	Q
Pinyin	Pinyin	P
Radical	Radical	R
Stroke	Stroke	S

You can sort Chinese data using the internationalized sort utility. This utility allows you to use one sorting method by selecting the respective locale as described in Chapter 3. In some cases, you may find that using one sorting method is insufficient to meet your needs. You might need to sort your data with multiple collating sequences. For instance, many characters can have the same number of strokes and you might want to sort these characters further according to their radicals.

To sort characters according to their radicals, Tru64 UNIX provides an extended sort utility, called `asort`, that you can use to sort or merge files containing Chinese characters according to specified collating sequences. The `asort` utility has the same syntax as that of the `sort` utility, but it provides two additional options:

-C"collate_sequence"	Defines the collating sequences where collate_sequence is a list of identifiers or abbreviations of the collating sequences for sorting or merging a file.
-v	Sorts Chinese data in breadth-first comparison, just like the behavior of the VMS/Hanyu or VMS/Hanzi sorting mechanism. By default, depth-first comparison is used.

9.2.1 Multiple Collating Sequences

The `asort` utility allows you to specify multiple collating sequences. By default, Chinese data is sorted by internal code. You can specify collating sequences with the -C option so that Chinese data can be sorted using other collation methods. For example, the following command sorts DEC Hanyu data files in the order of stroke, radical, and then phonetic:

```
% setenv LANG zh_TW.dechanyu

% asort -C"Stroke Radical Phonetic" input.dat> output.dat
```

Alternatively you can enter:

```
% setenv LANG zh_TW.dechanyu
% asort -C"srp" input.dat > output.dat
```

These commands first sort the input data file according to the number of strokes. If multiple characters have the same number of strokes, they are then sorted by radical. If multiple characters within this group start with the same radical, they will then be sorted by phonetic order.

_____ **Note** _____

The `asort` utility is locale sensitive. You should first set the LANG environment variable to the required Chinese locales before using the `asort` utility.

_____

9.2.2 Comparing Depth–First and Breadth–First Sorting

By default, the `asort` utility compares Chinese data according to the specified collating sequences using depth-first comparison. That is, each character in a sort field is compared using all the specified collating sequences until the collating order is resolved. When two characters have the same collating order, the next pair of characters is compared.

OpenVMS/Hanyu or OpenVMS/Hanzi use a slightly different sorting mechanism. The HSORT utility provided with OpenVMS/Hanyu sorts characters in the whole sort field using the first collating method. The second collating method applies only if the collating order of the two sorting fields is identical. This is called breadth-first comparison. If you want your sorting results to be compatible with that generated by OpenVMS/Hanyu or OpenVMS/Hanzi, you can specify the -v option:

```
% asort -C"srp" -v input.dat > output.dat
```

9.2.3 User–Defined Characters

The asort utility supports the sorting of user-defined characters with the collating values defined through the cedit utility. If required, the asort utility looks up the collating values from the User-Defined Character (UDC) database and sorts the data accordingly. The mechanism for sorting UDCs is totally transparent to you.

9.3 Hanyu and Hanzi DECterm

Hanyu DECterm is a VT382-D terminal emulator; Hanzi DECterm is a VT382-C terminal emulator. This section describes the Chinese features that are specific to the Hanyu and Hanzi DECterm. For details on the common internationalization features provided by DECterm, see *Using International Software*.

9.3.1 Creating a Hanyu or Hanzi DECterm

The terminal type that DECterm emulates is sensitive to the session language.

To create a Hanyu DECterm through the Session Manager, set the session language to one of the traditional Chinese locales, for example, Chinese Taiwan, and then select DECterm from the Applications menu of Session Manager. To create a Hanzi DECterm, select one of the simplified Chinese locales, for example, Chinese China.

Alternatively, you can use the -xnllanguage qualifier to specify the terminal type of the DECterm to create. For example, you can use zh_CN.dechanzi as the value for -xnllanguage to create a Hanzi DECterm:

```
% /usr/bin/X11/dxterm -xnllanguage \ zh_CN.dechanzi
```

If you specify an unknown value for -xnllanguage, then ISO Latin-1 DECterm is assumed. If no Chinese font exists, it defaults to ISO Latin-1 DECterm.

The user interface language of Hanyu and Hanzi DECterm always follows the terminal type. The language is independent of the language selection.

9.3.2 Customizing DECterm

Except for customization of NRCS character sets, all customization features applicable to the ISO Latin-1 DECterm window can also be applied to any Hanyu or Hanzi DECterm window.

Customized features can be saved and restored in the same way as in ISO Latin-1 DECterm.

9.3.3 Font Sizes

By choosing the Big Font or Little Font option from the Window... item of the Options menu you can choose either the 24 x 24 or 16 x 18 Chinese fonts.

9.3.4 Terminal ID

By choosing the General... item from the Options menu you can change the general features, such as the terminal type, for the Hanyu and Hanzi DECterm from a dialog box. You can also choose the VT382 ID from the dialog box.

9.3.5 Interaction Style

By choosing the Input Method... item from the Options menu you can select the interaction style for Hanyu and Hanzi DECterm. For example, if you want to select the Root window interaction style, you can click on the Root window button and then apply the change. If you click on the ISO Latin 1 Input button, Hanyu and Hanzi DECterm disable the input of Chinese data until another style is selected.

9.3.6 Input Server

By choosing the Input Method... item from the Options menu you can switch to use another input server for Hanyu and Hanzi DECterm. By default, the traditional Chinese input server is used for Hanyu DECterm, and simplified Chinese input server is used for Hanzi DECterm. To select another input server, you can click on the Other button and then enter the input server name on the input field.

For Hanyu DECterm, you can enter DECCN to switch to the simplified Chinese input server. For Hanzi DECterm, you can enter DECTW to switch to the traditional Chinese input server. For details about these input servers, see Chapter 0.

9.3.7 Copying Information

You can use the Edit menu to copy information within or between DECterm windows. The Cut-and-Paste operation is enhanced to handle mixed ASCII and Chinese characters. Beyond this, conversion between traditional and simplified Chinese data is performed when data is copied between Hanyu DECterm and simplified Chinese applications, and between Hanzi DECterm and traditional Chinese applications, through the cut-and-paste or quick copying operation.

9.3.8 Default Chinese Character Set

Hanyu DECterm supports CNS 11643 (first and second planes), DTSCS and all character sets supported by the ISO Latin-1 DECterm. Hanzi DECterm supports GB2312 and all character sets supported by the ISO Latin-1 DECterm.

ISO Latin-1 DECterm uses the ISO 8859-1 (Latin-1) as the default character set. You can override this setting by choosing another option from the General... item on the Options menu. For Hanyu DECterm, the default character set for 8-bit data is the Hanyu character set (CNS 11643 and DTSCS). For Hanzi DECterm, the default character set for 8-bit data is the Hanzi character set GB2312.

In general, Hanyu and Hanzi DECterm cannot display mixed accented Latin-1 characters and Chinese characters. If you want to achieve this, you can output the data together with the designated character set escape sequences.

9.3.9 Chinese Character Input and Output

You can enter Chinese characters in Hanyu and Hanzi DECterm by invoking any of the Chinese input modes as described in Table 7–4.

Mixed ASCII and Chinese characters can be displayed properly in Hanyu and Hanzi DECterm without any special settings.

9.3.10 Reconnecting the Input Server

The Chinese input server provides you the ability to input Chinese characters. If this process does not exist or terminates for some reason, one of the following messages is displayed:

- `Hanyu input method does not exist`

- `Hanzi input method does not exist`

You can restart the input server and then use the Reset Terminal option from the Commands menu to reconnect the Hanyu and Hanzi DECterm to the input server.

9.3.11 VT382–D and VT382–C Terminal Functions

The following functions of the VT382-D and VT382-C terminal are implemented in the Hanyu and Hanzi DECterm terminal emulator respectively:

- Display characteristics and capabilities

- Text capabilities

    — Level 3 terminal compatibility

        * VT300 mode

        * VT100 mode

        * VT52 mode

    — ANSI compatible control functions

- Support for Terminal State Interrogation (TSI)

- Support for all of the Chinese input methods

- Support for the following character sets:

  — DEC Special Graphics Character Set (line drawing)

  — DEC Supplemental Character Set

  — DEC Technical Character Set

  — ISO Latin-1 Character Set

  — CNS11643-1986 and DTSCS-1990 Character Sets for VT382-D, and GB2312-80 Characters Sets for VT382-C

- Control Representation mode

- Support for sixel graphics

- Support for UDK editing function

- Support for Chinese character display attributes: reverse, underline, bold, blink, double height/width

The following functions of the VT382-D terminal are implemented in the Hanyu DECterm terminal emulator:

- Display/Suppress leading code

  A selection button is added in the Display... item under the Options menu for users to enable or disable the display of a symbol for the leading code in a 4-byte EDPC character.

- The escape sequence DECLCSM, that is, Leading Code Suppression Mode, is also supported in Hanyu DECterm.

For details about the VT382-D terminal functions, see the *VT382-D Programming Reference Manual* and *VT382-D User's Manual*. For the details of the VT382-C terminal functions, see the *VT382-C Programming Reference Manual* and *VT382-C User's Manual*.

9.4 Phrase Conversion

Chapter 2 describes how Tru64 UNIX supports conversion between different codesets using the iconv utility. This utility can also be used for phrase conversion. When phrase conversion is activated, a phrase in traditional Chinese can be converted to a phrase in simplified Chinese, or the reverse. Phrase conversion does not apply to traditional Chinese encoded in Telecode, but this is the only exception.

To activate the phrase conversion option, you can define the ICONV_PHRCONV environment variable. If this environment variable is set to mark, the converted phrases are enclosed in brackets ([]) to highlight the conversion result for visual checking.

The phrase conversion databases in the /usr/share/phrdb directory are normal text files with the same file names as those of the algorithmic converters in /usr/lib/nls/loc/iconv/*. These phrase conversion databases contain entries for phrase conversion pairs.

9.5 Special Characters in nroff

The nroff utility has been internationalized to format text of various languages. When a Chinese document is formatted using nroff, its contents are handled according to Chinese formatting rules:

- Text line wrapping - Chinese text, unlike English text, does not use space characters as separators between words or as break points for text line wrapping. Instead, a Chinese text line can be broken between any two consecutive Chinese characters, with the following exceptions:

  - Some Chinese characters cannot be placed at the beginning of a text line. They are called no-first characters. For traditional Chinese, the no-first characters include the following:

  ，、。．・；：？！‥……｜、、
  ・；：？！）‾｝‾〕‾〗━》∨
  〉∨」═』═）｝〉'〞〝、＞＝

  For simplified Chinese, the no-first characters include the following:

  、。‾˘¨〃－～‖…'"〕〉》」』
  〗】：！），．：；＞？］｝

  - Some Chinese characters cannot be placed at the end of a text line. They are called no-last characters. For traditional Chinese, the no-last characters include the following:

  （‾｛‾〔‾〘━《∧〈∧「‾『
  ═（｛（'"〝〉＜＜

  For simplified Chinese, the no-last characters include the following:

  ～'"〔〈《「『〖〗（［｛

  - Some English characters are handled similarly according to this rule. No-first English characters include the following:

  !),.:;>?])

No-last English characters include the following:

(< [(

- Right justification - To justify the right margin of a paragraph, `nroff` inserts space characters at proper break points.

  For English, space characters are usually inserted at word breaks.

  For Chinese, space characters are inserted only at the following places:

  - Before a can-space-before character if it is not placed at the beginning of a text line.

    For traditional Chinese, the can-space-before characters include the following:

    (⌒ { ⌒ 〔 ⌒ 〖 ▬ 《 ≪ 〈 ⌄ 「 ¬ 『
    ¬({ (" ＼' < <

    For simplified Chinese, the can-space-before characters include the following:

    ' " 〔 〈 《 「 『 〖 ([{

  - After a can-space-after character if it is not placed at the end of a text line.

    For traditional Chinese, the can-space-after characters include the following:

    , 、 。 ? ! ? !) ⌣ } ⌣) ⌣ 〗 ▬
    》 ⋙ 〉 ⌣ 」 ╚ 』 ╚) }) " ″ ＼ > =

    For simplified Chinese, the can-space-after characters include the following:

    、 。 ' ") 〉 》 」 』 〗 〗 !) , . > ?] }