

TECHNICAL REPORT

**ISO/IEC
TR
14652**

First edition
2004-09-15

Information technology — Specification method for cultural conventions

*Technologies de l'information — Méthode de modélisation des
conventions culturelles*



Reference number
ISO/IEC TR 14652:2004(E)

AVA Documents
Document Sharing Hub
© ISO/IEC 2004

PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

© ISO/IEC 2004

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents	Page
FOREWORD	iv
INTRODUCTION	v
1 SCOPE	1
2 NORMATIVE REFERENCES	1
3 TERMS, DEFINITIONS AND NOTATIONS	2
4 FDCC-set	7
4.1 FDCC-set description	7
4.2 LC_IDENTIFICATION	12
4.3 LC_CTYPE (controversial)	13
4.4 LC_COLLATE	27
4.5 LC_MONETARY (controversial)	39
4.6 LC_NUMERIC	43
4.7 LC_TIME (controversial)	44
4.8 LC_MESSAGES	51
4.9 LC_XLITERATE (controversial)	52
4.10 LC_NAME	54
4.11 LC_ADDRESS	55
4.12 LC_TELEPHONE	57
5 CHARMAP	58
6 REPERTOIREMAP (controversial)	63
Annex A (informative) DIFFERENCES FROM POSIX (ISO/IEC 9945)	92
Annex B (informative) RATIONALE	94
Annex C (informative) BNF GRAMMAR	107
Annex D (informative) OUTSTANDING ISSUES	112
INDEX	116
BIBLIOGRAPHY	119

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

In exceptional circumstances, the joint technical committee may propose the publication of a Technical Report of one of the following types:

- type 1, when the required support cannot be obtained for the publication of an International Standard, despite repeated efforts;
- type 2, when the subject is still under technical development or where for any other reason there is the future but not immediate possibility of an agreement on an International Standard;
- type 3, when the joint technical committee has collected data of a different kind from that which is normally published as an International Standard (“state of the art”, for example).

Technical Reports of types 1 and 2 are subject to review within three years of publication, to decide whether they can be transformed into International Standards. Technical Reports of type 3 do not necessarily have to be reviewed until the data they provide are considered to be no longer valid or useful.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

ISO/IEC TR 14652, which is a Technical Report of type 1, was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 22, *Programming languages, their environments and system software interfaces*.

This document is published as a type 1 Technical Report on the request of a number of member bodies, whose concerns are recorded in Annex D.

Introduction

This Technical Report defines a general mechanism to specify cultural conventions, and it defines formats for a number of specific cultural conventions in the areas of character classification and conversion, sorting, number formatting, monetary formatting, date formatting, message display, addressing of persons, postal address formatting, and telephone number handling.

There are a number of benefits coming from this Technical Report:

Rigid specification	Using this Technical Report, a user can rigidly specify a number of the cultural conventions that apply to the information technology environment of the user.
Cultural adaptability	If an application has been designed and built in a culturally neutral manner, the application may use the specifications as data to its APIs, and thus the same application may accommodate different users in a culturally acceptable way to each of the users, without change of the binary application.
Productivity	This Technical Report specifies those cultural conventions and how to specify data for them. With that data an application developer is relieved from getting the different information to support all the cultural environments for the expected customers of the product. The application developer is thus ensured of culturally correct behaviour as specified by the customer, and possibly more markets may be reached as customers may have the possibility to provide the data themselves for markets that were not targeted.
Uniform behaviour	When a number of applications share one cultural specification, which may be supplied from the user or provided by the application or operating system, their behaviour for cultural adaptation becomes uniform.

The specification format is independent of platforms and specific encoding, and targeted to be usable from a wide range of programming languages.

A number of cultural conventions, such as spelling, hyphenation rules and terminology, are not specifiable with this Technical Report, but it provides mechanisms to define new categories and also new keywords within existing categories. An internationalized application may take advantage of information provided with the FDCC-set (such as the language) to provide further internationalized services to the user.

This Technical Report defines a format compatible with the one used in the International string ordering standard, ISO/IEC 14651. This Technical Report is upward compatible with the ISO/IEC 9945-1:2003 POSIX, particularly its clauses 6 and 7. The major extensions

from that text are listed in annex A. This Technical Report has enhanced functionality in a number of areas such as ISO/IEC 10646 support, more classification of characters, transliteration, dual (multi) currency support, enhanced date and time formatting, personal name writing, postal address formatting, telephone number handling, and management of categories. There is enhanced support for character sets including ISO/IEC 2022 handling and an enhanced method to separate the specification of cultural conventions from an actual encoding via a description of the character repertoire employed. A standard set of values for all the categories has been defined covering the repertoire of ISO/IEC 10646-1, as referenced in the normative references clause.

The Technical Report was originally scheduled for adoption as an International Standard, but a number of members of ISO and IEC found the specification problematical. It was then decided to convert the specification into a Technical Report type I. Annex D lists a number of issues that some members of ISO and IEC have with the specification.

The following clauses are thus marked as controversial:

4.3 LC_CTYPE

4.5 LC_MONETARY

4.7 LC_TIME

4.9 LC_XLITERATE

6 REPERTOIREMAP

Information technology — Specification method for cultural conventions

1 SCOPE

This Technical Report specifies a description format for the specification of cultural conventions, a description format for character sets, and a description format for binding character names to ISO/IEC 10646, plus a set of default values for some of these items.

The specification is upward compatible with POSIX locale specifications - a locale conformant to POSIX specifications will also be conformant to the specifications in this Technical Report, while the reverse condition will not hold. The descriptions are intended to be coded in text files to be used via Application Programming Interfaces, that are expected to be developed for a number of systems which comply with ISO/IEC 9945. An alignment effort has been undertaken for this specification to be aligned with the revision of the ISO/IEC 9945 standard published in 2003.

2 NORMATIVE REFERENCES

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 639 (all parts), *Codes for the representation of names of languages*.

ISO/IEC 2022, *Information technology - Character code structure and extension techniques*.

ISO 3166 (all parts), *Codes for the representation of names of countries and their subdivisions*.

ISO 4217, *Codes for the representation of currencies and funds*.

ISO 8601, *Data elements and interchange formats - Information interchange - Representation of dates and times*.

ISO/IEC 9945:2003 *Information technology - Portable Operating System Interface (POSIX)*.

ISO/IEC 10646:2003, *Information technology - Universal Multiple-Octet Coded Character Set (UCS)*. Only the fixed collection 301 plus the characters U20AC EURO SIGN and the UFFFC OBJECT REPLACEMENT CHARACTER (corresponding to UNICODE 2.1) and the control characters U0000..U001F are used in this document.

ISO/IEC 14651:2001, *Information technology - International string ordering and comparison - Method for comparing character strings and description of the common template tailorable ordering*.

ISO/IEC 15897:1999, *Information technology - Procedures for registration of cultural elements*.

3 TERMS, DEFINITIONS AND NOTATIONS

3.1 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

3.1.1 Bytes and characters

3.1.1.1

byte:

An individually addressable unit of data storage that is equal to or larger than an octet, used to store a character or a portion of a character.

A byte is composed of a contiguous sequence of bits, the number of which is implementation defined. The least significant bit is called the low-order bit; the most significant bit is called the high-order bit.

3.1.1.2

character:

A member of a set of elements used for the organization, control or representation of data.

3.1.1.3

coded character:

A sequence of one or more bytes representing a single character.

3.1.1.4

text file:

A file that contains characters organized into one or more lines.

3.1.2 cultural and other major concepts

3.1.2.1

cultural convention:

A data item for information technology that may vary dependent on language, territory, or other cultural habits.

3.1.2.2

FDCC:

A Formal Definition of a Cultural Convention, that is a cultural convention put into a formal definition scheme.

3.1.2.3

FDCC-set:

A Set of Formal Definitions of Cultural Conventions (FDCC's). The definition of the subset of a user's information technology environment that depends on language and cultural conventions.

Note: the FDCC-set is a superset of the "locale" term in C and POSIX.

3.1.2.4**charmap:**

A definition of a mapping between symbolic character names and character codes, plus related information.

3.1.2.5**repertoiremap:**

A definition of a mapping between symbolic character names and characters for the repertoire of characters used in a FDCC-set, further described in clause 6.

3.1.3 FDCC categories related**3.1.3.1****character class:**

A named set of characters sharing an attribute associated with the name of the class.

3.1.3.2**collation:**

The logical ordering of strings according to defined precedence rules.

3.1.3.3**collating element:**

The smallest entity used to determine logical ordering.

See collating sequence. A collating element consists of either a single character, or two or more characters collating as a single entity. The LC_COLLATE category in the associated FDCC-set determines the set of collating elements.

3.1.3.4**multicharacter collating element:**

A sequence of two or more characters that collate as an entity.

For example, in some languages two characters are sorted as one letter, as in the case for Danish and Norwegian "aa".

3.1.3.5**collating sequence:**

The relative order of collating elements as determined by the setting of the LC_COLLATE category in the applied FDCC-set.

3.1.3.6**equivalence class:**

A set of collating elements with the same primary collation weight.

Elements in an equivalence class are typically elements that naturally group together, such as all accented letters based on the same letter.

The collation order of elements within an equivalence class is determined by the weights assigned on any subsequent levels after the primary weight.

3.2 Notations

The following notations and common conventions for specifications apply to this Technical Report:

3.2.1 Notation for defining syntax

In this Technical Report, the description of an individual record in a FDCC-set is done using the syntax notation given in the following.

The syntax notation looks as follows:

```
"<format>",<arg1>,<arg2>,...,<argn>]
```

The <format> is given in a format string enclosed in double quotes, followed by a number of parameters, separated by commas. It is similar to the format specification defined in clause 5 in the ISO/IEC 9945-1:2003 standard and the format specification used in C language printf() function. The format of each parameter is given by an escape sequence as follows:

%s	specifies a string
%d	specifies a decimal integer
%c	specifies a character
%o	specifies an octal integer
%x	specifies a hexadecimal integer

A " " (an empty character position) in the syntax string represents one or more <blank> characters.

All other characters in the format string represent themselves, except:

%%	specifies a single %
\n	specifies an end-of-line

The notation "..." is used to specify that repetition of the previous specification is optional, and this is done in both the format string and in the parameter list.

3.2.3 Portable character set

A set of symbolic names for characters in Table 1, which is called the portable character set, is used in character description text of this specification. The first eight entries in Table 1 are defined in ISO/IEC 6429 and the rest is defined in ISO/IEC 9945-1:2003 with some definitions from ISO/IEC 10646-1.

Table 1: Portable character set

Symbolic name	Glyph	UCS	Description
<NUL>		<U0000>	NULL (NUL)
<alert>		<U0007>	BELL (BEL)
<backspace>		<U0008>	BACKSPACE (BS)
<tab>		<U0009>	CHARACTER TABULATION (HT)

<carriage-return>		<U000D>	CARRIAGE RETURN (CR)
<newline>		<U000A>	LINE FEED (LF)
<vertical-tab>		<U000B>	LINE TABULATION (VT)
<form-feed>		<U000C>	FORM FEED (FF)
<space>		<U0020>	SPACE
<exclamation-mark>	!	<U0021>	EXCLAMATION MARK
<quotation-mark>	"	<U0022>	QUOTATION MARK
<number-sign>	#	<U0023>	NUMBER SIGN
<dollar-sign>	\$	<U0024>	DOLLAR SIGN
<percent-sign>	%	<U0025>	PERCENT SIGN
<ampersand>	&	<U0026>	AMPERSAND
<apostrophe>	'	<U0027>	APOSTROPHE
<left-parenthesis>	(	<U0028>	LEFT PARENTHESIS
<right-parenthesis>	)	<U0029>	RIGHT PARENTHESIS
<asterisk>	*	<U002A>	ASTERISK
<plus-sign>	+	<U002B>	PLUS SIGN
<comma>	,	<U002C>	COMMA
<hyphen-minus>	-	<U002D>	HYPHEN-MINUS
<hyphen>	-	<U002D>	HYPHEN-MINUS
<full-stop>	.	<U002E>	FULL STOP
<period>	.	<U002E>	FULL STOP
<slash>	/	<U002F>	SOLIDUS
<solidus>	/	<U002F>	SOLIDUS
<zero>	0	<U0030>	DIGIT ZERO
<one>	1	<U0031>	DIGIT ONE
<two>	2	<U0032>	DIGIT TWO
<three>	3	<U0033>	DIGIT THREE
<four>	4	<U0034>	DIGIT FOUR
<five>	5	<U0035>	DIGIT FIVE
<six>	6	<U0036>	DIGIT SIX
<seven>	7	<U0037>	DIGIT SEVEN
<eight>	8	<U0038>	DIGIT EIGHT
<nine>	9	<U0039>	DIGIT NINE
<colon>	:	<U003A>	COLON
<semicolon>	;	<U003B>	SEMICOLON
<less-than-sign>	<	<U003C>	LESS-THAN SIGN
<equals-sign>	=	<U003D>	EQUALS SIGN
<greater-than-sign>	>	<U003E>	GREATER-THAN SIGN
<question-mark>	?	<U003F>	QUESTION MARK
<commercial-at>	@	<U0040>	COMMERCIAL AT
<A>	A	<U0041>	LATIN CAPITAL LETTER A
	B	<U0042>	LATIN CAPITAL LETTER B
<C>	C	<U0043>	LATIN CAPITAL LETTER C
<D>	D	<U0044>	LATIN CAPITAL LETTER D
<E>	E	<U0045>	LATIN CAPITAL LETTER E
<F>	F	<U0046>	LATIN CAPITAL LETTER F
<G>	G	<U0047>	LATIN CAPITAL LETTER G
<H>	H	<U0048>	LATIN CAPITAL LETTER H
<I>	I	<U0049>	LATIN CAPITAL LETTER I
<J>	J	<U004A>	LATIN CAPITAL LETTER J
<K>	K	<U004B>	LATIN CAPITAL LETTER K
<L>	L	<U004C>	LATIN CAPITAL LETTER L
<M>	M	<U004D>	LATIN CAPITAL LETTER M
<N>	N	<U004E>	LATIN CAPITAL LETTER N
<O>	O	<U004F>	LATIN CAPITAL LETTER O
<P>	P	<U0050>	LATIN CAPITAL LETTER P
<Q>	Q	<U0051>	LATIN CAPITAL LETTER Q
<R>	R	<U0052>	LATIN CAPITAL LETTER R
<S>	S	<U0053>	LATIN CAPITAL LETTER S
<T>	T	<U0054>	LATIN CAPITAL LETTER T
<U>	U	<U0055>	LATIN CAPITAL LETTER U
<V>	V	<U0056>	LATIN CAPITAL LETTER V
<W>	W	<U0057>	LATIN CAPITAL LETTER W
<X>	X	<U0058>	LATIN CAPITAL LETTER X
<Y>	Y	<U0059>	LATIN CAPITAL LETTER Y
<Z>	Z	<U005A>	LATIN CAPITAL LETTER Z
<left-square-bracket>	[	<U005B>	LEFT SQUARE BRACKET
<backslash>	\	<U005C>	REVERSE SOLIDUS
<reverse-solidus>	\	<U005C>	REVERSE SOLIDUS
<right-square-bracket>	]	<U005D>	RIGHT SQUARE BRACKET

<circumflex-accent>	^	<U005E>	CIRCUMFLEX ACCENT
<circumflex>	^	<U005E>	CIRCUMFLEX ACCENT
<low-line>	—	<U005F>	LOW LINE
<underscore>	—	<U005F>	LOW LINE
<grave-accent>	`	<U0060>	GRAVE ACCENT
<a>	a	<U0061>	LATIN SMALL LETTER A
	b	<U0062>	LATIN SMALL LETTER B
<c>	c	<U0063>	LATIN SMALL LETTER C
<d>	d	<U0064>	LATIN SMALL LETTER D
<e>	e	<U0065>	LATIN SMALL LETTER E
<f>	f	<U0066>	LATIN SMALL LETTER F
<g>	g	<U0067>	LATIN SMALL LETTER G
<h>	h	<U0068>	LATIN SMALL LETTER H
<i>	i	<U0069>	LATIN SMALL LETTER I
<j>	j	<U006A>	LATIN SMALL LETTER J
<k>	k	<U006B>	LATIN SMALL LETTER K
<l>	l	<U006C>	LATIN SMALL LETTER L
<m>	m	<U006D>	LATIN SMALL LETTER M
<n>	n	<U006E>	LATIN SMALL LETTER N
<o>	o	<U006F>	LATIN SMALL LETTER O
<p>	p	<U0070>	LATIN SMALL LETTER P
<q>	q	<U0071>	LATIN SMALL LETTER Q
<r>	r	<U0072>	LATIN SMALL LETTER R
<s>	s	<U0073>	LATIN SMALL LETTER S
<t>	t	<U0074>	LATIN SMALL LETTER T
<u>	u	<U0075>	LATIN SMALL LETTER U
<v>	v	<U0076>	LATIN SMALL LETTER V
<w>	w	<U0077>	LATIN SMALL LETTER W
<x>	x	<U0078>	LATIN SMALL LETTER X
<y>	y	<U0079>	LATIN SMALL LETTER Y
<z>	z	<U007A>	LATIN SMALL LETTER Z
<left-brace>	{	<U007B>	LEFT CURLY BRACKET
<left-curly-bracket>	{	<U007B>	LEFT CURLY BRACKET
<vertical-line>		<U007C>	VERTICAL LINE
<right-brace>	}	<U007D>	RIGHT CURLY BRACKET
<right-curly-bracket>	}	<U007D>	RIGHT CURLY BRACKET
<tilde>	~	<U007E>	TILDE

This Technical Report may use other symbolic character names than the above in examples, to illustrate the use of the range of symbols allowed by the syntax specified in 4.1.1.

4 FDCC-set

A FDCC-set is the definition of the subset of a user's information technology environment that depends on language and cultural conventions. It is made up from one or more categories. Each category is identified by its name and controls specific aspects of the behaviour of components of the system. This Technical Report defines the following categories:

LC_IDENTIFICATION	Versions and status of categories
LC_CTYPE	Character classification, case conversion and code transformation.
LC_COLLATE	Collation order.
LC_TIME	Date and time formats.
LC_NUMERIC	Numeric, non-monetary formatting.
LC_MONETARY	Monetary formatting.
LC_MESSAGES	Formats of informative and diagnostic messages and interactive responses.
LC_XLITERATE	Character transliteration.
LC_NAME	Format of writing personal names.
LC_ADDRESS	Format of postal addresses.
LC_TELEPHONE	Format for telephone numbers, and other telephone information.

Note: In future editions of this Technical Report further categories may be added.

Other category names beginning with the 3 characters "LC_" are reserved for future standardization, except for category names beginning with the five characters "LC_X_" which is not used for future addition of categories specified in this Technical Report. An application may thus use category names beginning with the five characters "LC_X_" for application defined categories to avoid clashes with future standardized categories.

This Technical Report also defines an FDCC-set named "i18n" with values for some of the above categories in order to simplify FDCC-set descriptions for a number of cultures. The contents of "i18n" categories should not necessarily be considered as the most commonly accepted values, while in many cases it could be the recommended values. The complete "i18n" FDCC-set is defined as the sum of the "i18n" categories specified in the clauses below.

4.1 FDCC-set description

FDCC-sets are described with the syntax presented in this subclause. For the purposes of this Technical Report, the text is referred to as the FDCC-set definition text or FDCC-set source text.

The **FDCC-set definition text** contains one or more FDCC-set category source definitions, and does not contain more than one definition for the same FDCC-set category. If the text contains source definitions for more than one category, application-defined categories, if present, appears after the categories defined by this clause. A category source definition contains either the definition of a category or a copy directive. In the event that some of the information for a FDCC-set category, as specified in this Technical Report, is missing

from the FDCC-set source definition, the behaviour of that category, if it is referenced, is unspecified. A FDCC-set category is the normal way of specifying a single FDCC.

There are no **naming conventions** for FDCC-sets specified in this Technical Report, but clause 6.8 in ISO/IEC 15897:1999 specifies naming rules for POSIX locales, charmaps and repertoire maps, that may also be applied to FDCC-sets, charmaps and repertoire maps specified according to this Technical Report.

A **category source definition** consists of a category header, a category body, and a category trailer. A category header consists of the character string naming of the category, beginning with the characters "LC_". The category trailer consists of the string "END", followed by one or more "blank"s and the string used in the corresponding category header.

The **category body** consists of one or more lines of text. Each line is one of the following:

- a line containing an identifier, optionally followed by one or more operands. Identifiers are either keywords, identifying a particular FDCC, or collating elements, or section symbols,
- one of transliteration statements defined in 4.3.

In addition to the keywords defined in this Technical Report, the source can contain application-defined keywords. Each **keyword** within a category has a unique name (i.e., two categories can have a commonly-named keyword); no keyword starts with the characters "LC_". Identifiers are separated from the operands by one or more "blank"s.

Operands are characters, collating elements, section symbols, or strings of characters. Strings are enclosed in double-quotes. Literal double-quotes within strings are preceded by the <escape character>, described below. When a keyword is followed by more than one operand, the operands are separated by semicolons; "blank"s are allowed before and/or after a semicolon.

4.1.1 Character representation

Individual characters, characters in strings, and collating elements are represented using symbolic names, UCS notation or characters themselves, or as octal, hexadecimal, or decimal constants as defined below. When constant notation is used, the resultant FDCC-set definitions need not be portable between systems.

- (0) The left angle bracket (<) is a reserved symbol, denoting the start of a symbolic name; when used to represent itself outside a symbolic name it is preceded by the escape character.
- (1) A character can be represented via a **symbolic name**, enclosed within angle brackets (< and >). The symbolic name, including the angle brackets, exactly matches a symbolic name defined in a charmap or a repertoire map to be used, and is replaced by a character value determined from the value associated with the symbolic name in the charmap

or a value associated via a repertoiremap. Repertoiremaps have predefined symbolic names for UCS characters, see clause 6. A FDCC-set may also use the UCS notation of clause 6 to represent characters, without a repertoiremap being defined for the FDCC-set. Use of the escape character or a right angle bracket within a symbolic name is invalid unless the character is preceded by the escape character.

Example: <c>;<c-cedilla> "<M><a><y>"

The items (2), (3), (4) and (5) are deprecated and are retained for compatibility with the POSIX standard. FDCC-sets should be specified in a coded character set independent way, using symbolic names. To make actual use of the FDCC-set, it is used together with charmaps and/or repertoiremaps, so that the symbolic character names can be resolved into the actual character encoding used.

- (2) A character can be represented by the character itself, in which case the value of the character is application-defined. Within a string, the double-quote character, the escape character, and the right angle bracket character are escaped (preceded by the escape character) to be interpreted as the character itself. Outside strings, the characters

, ; < > escape_char

are escaped by the escape character to be interpreted as the character itself.

Example: c ä "May"

- (3) A character can be represented as an octal constant. An octal constant is specified as the escape character followed by two or more octal digits. Each constant represents a byte value.

Example: \143; \347; "\115"

- (4) A character can be represented as a hexadecimal constant. A hexadecimal constant is specified as the escape character followed by an x followed by two or more hexadecimal digits. Each constant represents a byte value.

Example: \x63;\xe7;

- (5) A character can be represented as a decimal constant. A decimal constant is specified as the escape character followed by a d followed by two or more decimal digits. Each constant represents a byte value.

Example: \d99; \d231;

- (6) Multibyte characters can be represented by concatenated constants specified in byte order with the last constant specifying the least significant byte of the character. Concatenated constants can include a mix of the above character representations.

Example: `\143\xe7; "\115\xe7\d171"`

Only characters existing in the character set for which the FDCC-set definition is created are specified, whether using symbolic names, the characters themselves, or octal, decimal, or hexadecimal constants. If a charmap is present, only characters defined in the charmap can be specified using octal, decimal, or hexadecimal constants. Symbolic names not present in the charmap can be specified and are ignored, as specified under item (1) above.

Note: The <character> symbolic character notation is recommended for use of specifying all characters in a FDCC-set, to facilitate portability of the FDCC-sets, as the coded character set of the application of the FDCC-set may be different from the coded character set of the FDCC-set source. This is also recommended for format effectors in strings, such as in `LC_DATE` or `LC_ADDRESS`, where the format effectors are allowed to be stored together with the rest of the string, in a binary string with a different encoding from that of the source FDCC-set.

4.1.2 Continuation of lines

A line in a specification can be continued by placing an escape character as the last visible graphic character on the line; this continuation character is discarded from the input. The line is continued to the next non-comment line.

4.1.3 Names for copy keyword

In most of the categories a "copy" keyword is allowed. The name specified with this copy keyword is one of:

- "i18n" which indicate the "i18n" FDCC-set defined in this specification,
- the name of a FDCC-set or POSIX locale registered by the process defined in ISO/IEC 15897,
- any other name which may be recognized in some local context - not being recommended as an international specification.

4.1.4 Pre-category statements

In a FDCC-set the following statements can precede category specifications, and they apply to all categories in the specified FDCC-set.

4.1.4.1 comment_char

The following line in a FDCC-set modifies the comment character. It has the following syntax, starting in column 1:

`"comment_char %c\n", <comment_character>`

The comment character defaults to the number-sign (#). All examples in this Technical Report use "%" as the <comment_character>, except where otherwise noted. Blank lines and lines containing the <comment_character> in the first position are ignored. In collating statements a <comment_character> occurring where the delimiter ";" may occur, terminates the collating statement.

4.1.4.2 escape_char

The following line in a FDCC-set modifies the escape character to be used in the text. It has the following syntax, starting in column 1:

```
"escape_char %c\n", <escape_character>
```

The escape character is used for representing characters in 4.1.1 and for continuing lines. The escape character defaults to backslash "\". All examples in this Technical Report uses "/" as the escape character, except where otherwise noted.

4.1.4.3 repertoiremap

The following line in a FDCC-set specifies the name of a repertoiremap used to define the symbolic character names in the FDCC-set. There may be at most one "repertoiremap" line. It has the following syntax, starting in column 1:

```
"repertoiremap %s\n", <repertoiremap>
```

The name is one of:

- "i18nrep" which indicates the "i18nrep" repertoiremap defined in this specification,
- the name of a <repertoiremap> registered by the process defined in ISO/IEC 15897,
- any other name which may be recognized in some local context - not being recommended as an international specification.

4.1.4.4 charmap

The following line in a FDCC-set specifies the name of a charmap which may be used with the FDCC-set. It has the following syntax, starting in column 1:

```
"charmap %s\n", <charmap>
```

This keyword gives a hint on which charmaps a FDCC-set is meant to be supported by. There may be more than one charmap specification useful with a FDCC-set. It is an application's responsibility to decide what charmap specification is to be used with that application.

The name is one of:

- the name of a <charmap> registered by the process defined in ISO/IEC 15897,
- any other name which may be recognized in some local context - not being recommended as an international specification.

4.2 LC_IDENTIFICATION

The LC_IDENTIFICATION category defines properties of the FDCC-set, and which specification methods the FDCC-set is conforming to. Values must be supplied for all unless otherwise noted, and the operands are strings. The following keywords are defined:

title	Title of the FDCC-set.
source	Organization name of provider of the source.
address	Organization postal address.
contact	Name of contact person. This keyword is optional.
email	Electronic mail address of the organization, or contact person. This keyword is optional.
tel	Telephone number for the organization, in international format. This keyword is optional.
fax	Fax number for the organization, in international format. This keyword is optional.
language	Natural language to which the FDCC-set applies, as specified in ISO 639. If a two-letter code exists for this language, it is used, else the three-letter code is used. This keyword is optional.
territory	The geographic extent where the FDCC-set applies (where applicable), as two-letter form of ISO 3166. This keyword is optional.
audience	If not for general use, an indication of the intended user audience. This keyword is optional.
application	If for use of a special application, a description of the application. This keyword is optional.
abbreviation	Short name for provider of the source. This keyword is optional.
revision	Revision number consisting of digits and zero or more full stops (".").
date	Revision date in the format according to this example: "1995-02-05" meaning the 5th of February, 1995.

If required information is not present in ISO 639 or ISO 3166, the string should be given as empty, and the relevant Maintenance Authority should be approached to get the needed item registered.

Note: Only one language per territory can be addressed with a single FDCC-set; an additional FDCC-set is required for each additional language for that territory.

category	Is used to define that a category is present and what specification the category is claiming conformance to. The first operand is a string in double-quotes that describes the specification that the category is claiming conformance to, and the following values are defined: "i18n:2004" "posix:1993" The second operand is a string with the category name, where
-----------------	---

the category names of clause 4 are defined. More than one "category" keyword may be given, but only one per category name.

The "i18n" LC_IDENTIFICATION category is:

```
LC_IDENTIFICATION
% This is the ISO/IEC TR 14652 "i18n" definition for
% the LC_IDENTIFICATION category.
%
title          "ISO/IEC TR 14652 i18n FDCC-set"
source         "ISO/IEC Copyright Office"
address        "Case postale 56, CH-1211 Geneve 20, Switzerland"
contact       ""
email          ""
tel            ""
fax            ""
language       ""
territory      ""
revision       "1.0"
date           "2001-12-08"
%
category "i18n:2004";LC_IDENTIFICATION
category "i18n:2004";LC_CTYPE
category "i18n:2004";LC_COLLATE
category "i18n:2004";LC_TIME
category "i18n:2004";LC_NUMERIC
category "i18n:2004";LC_MONETARY
category "i18n:2004";LC_MESSAGES
category "i18n:2004";LC_NAME
category "i18n:2004";LC_ADDRESS
category "i18n:2004";LC_TELEPHONE

END LC_IDENTIFICATION
```

4.3 LC_CTYPE (controversial)

The LC_CTYPE category defines character classification, case conversion, character transformation, and other character attribute mappings. Support for the portable character set is required.

A series of characters in a specification can be represented by the hexadecimal symbolic ellipsis symbol ".." (two dots), the decimal symbolic ellipses symbols "...." (4 dots), or the absolute ellipses "... " (3 dots).

The **hexadecimal symbolic ellipsis** ("..") specification is only valid between symbolic character names. The symbolic names consists of zero or more nonnumeric characters from the set shown with visible glyphs in Table 1 of clause 3.2.3, followed by an integer formed by one or more hexadecimal digits, using uppercase letters only for the range "A" to "F". The characters preceding the hexadecimal integer are identical in the two symbolic names, and the integer formed by the hexadecimal digits in the second symbolic name are identical to or greater than the integer formed by the hexadecimal digits in the first name. This is interpreted as a series of symbolic names formed from the common part and each of the integers in hexadecimal format using uppercase letters only between the first and the second integer, inclusive, and with a length of the symbolic names generated that is equal to the length of the first (and also the second) symbolic name. As an example, <U010E>..

The **decimal symbolic ellipsis** ("...") specification is only valid between symbolic character names. The symbolic names consist of zero or more nonnumeric characters from the set shown with visible glyphs in Table 1 of clause 3.2.3, followed by an integer formed by one or more decimal digits. The characters preceding the decimal integer are identical in the two symbolic names, and the integer formed by the decimal digits in the second symbolic name is identical to or greater than the integer formed by the decimal digits in the first name. This is interpreted as a series of symbolic names formed from the common part and each of the integers in decimal format between the first and the second integer, inclusive, and with a length of the symbolic names generated that is equal to the length of the first (and also the second) symbolic name. As an example, <j0101>....<j0104> is interpreted as the symbolic names <j0101>, <j0102>, <j0103>, and <j0104>, in that order.

The **absolute ellipsis** specification is only valid within a single encoded character set. An ellipsis is interpreted as including in the list all characters with an encoded value higher than the encoded value of the character preceding the ellipsis and lower than the encoded value of the character following the ellipsis. The absolute ellipsis specification is deprecated, as this is only relevant to FDCC-sets not using symbolic characters.

As an example, \x30;...;\x39 includes in the character class all characters with encoded values between the endpoints.

4.3.1 Character classification keywords

The following keywords are recognized. In the descriptions, the term "automatically included" means that it is not an error to either include the referenced characters or to omit them; the interpreting system provides them if missing and accept them silently if present.

copy	Specify the name of an existing FDCC-set to be used as the source for the definition of this category. If this keyword is specified, no other keyword is specified.
upper	Define characters to be classified as uppercase letters. No character specified for the keywords "cntrl", "digit", "punct", or "space" is specified. The uppercase letters A through Z of the portable character set, automatically belong to this class, with application-defined character values. The keyword may be omitted.
lower	Define characters to be classified as lowercase letters. No character specified for the keywords "cntrl", "digit", "punct", or "space" is specified. The lowercase letters a through z of the portable character set, automatically belong to this class, with application-defined character values. The keyword may be omitted.
alpha	Define characters to be classified as used to spell out the words for natural languages; such as letters, syllabic or ideographic characters. No character specified for the keywords "cntrl", "digit", "punct", or "space" is specified. In addition, characters classified as either "upper" or "lower" automatically belong to this class. The keyword may be omitted.
digit	Define the characters to be classified as decimal digits. Digits corresponding to the values 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9 can be specified in groups of 10 digits, and in ascending order of the values they represent. The digits of the portable character set are automatically included. If this keyword is not specified, the digits 0 through 9 of the portable character set automatically

	belong to this class, with application-defined character values. The "digit" keyword is used to specify which characters are accepted as digits in input to an application, such as characters typed in or scanned in from an input text file, and should list digits used with all the scripts supported by the FDCC-set. The keyword may be omitted.
alnum	Define the characters to be classified as used to spell out the words for natural languages, and numeric digits. The characters of the "alpha" and "digit" classes are automatically included in this class. The keyword may be omitted.
outdigit	Define the characters to be classified as decimal digits for output from an application, such as to a printer or a display or a output text file. Decimal digits corresponding to the values <0>, <1>, <2>, <3>, <4>, <5>, <6>, <7>, <8>, and <9> can be specified, and in ascending order of the values they represent. The intended use is for all places where decimal digits are used for output, including numeric and monetary formatting, and date and time formatting. Only one set of 10 decimal digits may be specified. If this keyword is not specified, the decimal digits 0 through 9 of the portable character set automatically belong to this class, with application-defined character values. The keyword may be omitted.
blank	Define characters to be classified as "blank" characters. If this keyword is unspecified, the characters <space> and <tab>, with application-defined character values, belong to this character class.
space	Define characters to be classified as white-space characters, to find syntactical boundaries. No character specified for the keywords "upper", "lower", "alpha", "digit", "graph", or "xdigit" is specified. If this keyword is not specified, the characters <space>, <form-feed>, <newline>, <carriage-return>, <tab>, and <vertical-tab>, automatically belong to this class, with application-defined character values. Any characters included in the class "blank" are automatically included. The class should not include the NO-BREAK spaces characters <U00A0>, <U2007>, <UFEFF>, as these characters should not be used for word boundaries. The keyword may be omitted.
cntrl	Define characters to be classified as control characters. No character specified for the keywords "upper", "lower", "alpha", "digit", "punct", "graph", "print", or "xdigit" is specified. The keyword is specified.
punct	Define characters to be classified as punctuation characters. No character specified for the keywords "upper", "lower", "alpha", "digit", "cntrl", "xdigit", or as the <space> character is specified. The keyword is specified.
xdigit	Define the characters to be classified as hexadecimal digits. Only the characters defined for the class "digit" are specified, in ascending sequence by numerical value, followed by sets of six characters representing the hexadecimal digits 10 through 15 in ascending order (for example <A>, , <C>, <D>, <E>, <F>, <a>, , <c>, <d>, <e>, <f>). The digits <0> through <9>, the uppercase letters <A> through <F>, and the lowercase letters <a> through <f>, automatically belong to this class, with application-defined character values.
graph	Define characters to be classified as printable characters, not including the <space> character. If this keyword is not specified, characters specified for the keywords "upper", "lower", "alpha", "digit", "xdigit", and "punct" belong to this character class. No character specified for the keyword "cntrl" is specified.

print	Define characters to be classified as printable characters, including the <space> character. If this keyword is not provided, characters specified for the keywords upper, lower, alpha, digit, xdigit, punct, graph, and the <space> character belong to this character class. No character specified for the keyword "cntrl" is specified.
toupper	Define the mapping of lowercase letters to uppercase letters. The operand consists of character pairs, separated by semicolons. The characters in each character pair are separated by a comma and the pair enclosed by parentheses. The first character in each pair is the lowercase letter, the second the corresponding uppercase letter. Only characters specified for the keywords "lower" and "upper" are specified. If this keyword is not specified, the lowercase letters <a> through <z>, and their corresponding uppercase letters <A> through <Z>, are automatically included, with application-defined character values.
tolower	Define the mapping of uppercase letters to lowercase letters. The operand consists of character pairs, separated by semicolons. The characters in each character pair are separated by a comma and the pair enclosed by parentheses. The first character in each pair is the uppercase letter, the second the corresponding lowercase letter. Only characters specified for the keywords "lower" and "upper" are specified. If this keyword is specified, the uppercase letters <A> through <Z>, and their corresponding lowercase letter, are specified. If this keyword is not specified, the mapping is the reverse mapping of the one specified for toupper.
class	(Controversial) Define characters to be classified in the class with the name given in the first operand, which is a string. This string only contains characters of the portable character set that either has the string "LETTER" in its description, or is a digit or <hyphen-minus> or <low-line>. The following operands are characters. This keyword is optional. The keyword can only be specified once per named class. The following two names are recognized: combining Characters to form composite graphic symbols, such as characters listed in ISO/IEC 10646:1993 annex B.1. combining_level3 Characters to form composite graphic symbols, that may also be represented by other characters, such as characters listed in ISO/IEC 10646-1:1993 annex B.2. The class names "upper", "lower", "alpha", "digit", "space", "cntrl", "punct", "graph", "print", "xdigit", and "blank" are taken to mean the classes defined by the respective keywords.
width	(Controversial) Define the column width of characters, for example for use of the C function wwidth(). The operands are first a list for characters, possibly using various ellipses, and semicolon separated, then a <colon>, and then the width of these characters given as an unsigned positive integer. Such width-lists separated by <semicolon> may be given for the various widths. The default value of width of characters in class "cntrl" and class "combining" is 0, else the default value of width is 1. A width for a character may be overridden by a WIDTH specification in a charmap. This keyword is optional.
map	(Controversial) Define the mapping of characters to other characters. The first operand is a string, defining the name of the mapping. The string only

contains letters, digits and <hyphen-minus> and <low-line> from the portable character set. The following operands consist of character pairs, separated by semicolons. The characters in each character pair are separated by a comma and the pair enclosed by parentheses. The first character in each pair is the character to map from, the second the corresponding character to map to. This keyword is optional. The keyword can only be specified once per named mapping.

The mapping names "toupper", and "tolower" are taken to mean the mapping defined by the respective keywords.

Example of use of the "map" keyword:

```
map "kana",(<U30AB>,<U304B>);(<U30AC>,<U304C>);(<U30AD>,<U304D>)
```

This example introduces a new mapping "kana" that maps three Katakana characters to corresponding Hiragana characters.

Table 2 shows the allowed character class combinations.

Table 2: Valid Character Class Combinations

Class	upper	lower	alpha	digit	space	cntrl	punct	graph	print	xdigit	blank
upper		+	A	x	x	x	x	A	A	+	x
lower	+		A	x	x	x	x	A	A	+	x
alpha	+	+		x	x	x	x	A	A	+	x
digit	x	x	x		x	x	x	A	A	A	x
space	x	x	x	x		+	*	*	*	x	+
cntrl	x	x	x	x	+		x	x	x	x	+
punct	x	x	x	x	+	x		A	A	x	+
graph	+	+	+	+	+	x	+		A	+	+
print	+	+	+	+	+	x	+	+		+	+
xdigit	+	+	+	+	x	x	x	A	A		x
blank	x	x	x	x	A	+	*	*	*	x	

Note 1: Explanation of codes:

A Automatically included; see text

+ Permitted

x Mutually exclusive

* See note 2

Note 2: The <space> character, which is part of the "space" and "blank" class, cannot belong to "punct" or "graph", but automatically belong to the "print" class. Other "space" or "blank" characters can be classified as "punct", "graph", and/or "print".

4.3.2 "i18n" LC_CTYPE category

The "i18n" FDCC-set for the LC_CTYPE is defined as follows:

```
LC_CTYPE
```

```
% The following is the ISO/IEC TR 14652 i18n fdcc-set LC_CTYPE category.
```

```
% It covers ISO/IEC 10646-1 including Cor.1 and AMD 1 thru 9
```

```
% COLLECTION numbers and names are from ISO/IEC 10646-1 Annex A
```

```
%
```

```
% The "upper" class reflects the uppercase characters of class "alpha"
```

```

upper /
% COLLECTION 1 BASIC LATIN/
<U0041>..<>U005A>;/
% COLLECTION 2 LATIN-1 SUPPLEMENT/
<U00C0>..<>U00D6>;<U00D8>..<>U00DE>;/
% COLLECTION 3 LATIN EXTENDED-A/
<U0100>;<U0102>;<U0104>;<U0106>;<U0108>;<U010A>;<U010C>;<U010E>;/
<U0110>;<U0112>;<U0114>;<U0116>;<U0118>;<U011A>;<U011C>;<U011E>;/
<U0120>;<U0122>;<U0124>;<U0126>;<U0128>;<U012A>;<U012C>;<U012E>;/
<U0130>;<U0132>;<U0134>;<U0136>;/
<U0138>;<U013A>;<U013C>;<U013E>;/
<U0140>;<U0142>;<U0144>;<U0146>;/
<U0148>;<U014A>;<U014C>;<U014E>;/
<U0150>;<U0152>;<U0154>;<U0156>;<U0158>;<U015A>;<U015C>;<U015E>;/
<U0160>;<U0162>;<U0164>;<U0166>;<U0168>;<U016A>;<U016C>;<U016E>;/
<U0170>;<U0172>;<U0174>;<U0176>;/
<U0178>;<U017A>;<U017C>;<U017E>;/
% COLLECTION 4 LATIN EXTENDED-B/
<U0181>;<U0183>;<U0185>;<U0187>;/
<U0189>..<>U018B>;<U018D>..<>U018F>;<U0191>;<U0193>;<U0195>;/
<U0197>..<>U0199>;<U019B>;<U019D>;<U019F>;/
<U01A0>;<U01A2>;<U01A4>;<U01A6>;/
<U01A8>;<U01AA>;<U01AC>;<U01AE>;<U01B0>;<U01B2>..<>U01B4>;/
<U01B6>;<U01B8>;<U01BA>;<U01BC>;<U01BE>;<U01C0>;<U01C2>;<U01C4>;<U01C6>;<U01C8>;/
<U01CA>;<U01CB>;/
<U01CD>;<U01CE>;<U01D0>;<U01D2>;<U01D4>;<U01D6>;<U01D8>;<U01DA>;<U01DC>;<U01DE>;/
<U01E0>;<U01E2>;<U01E4>;<U01E6>;<U01E8>;<U01EA>;<U01EC>;<U01EE>;/
<U01F0>;<U01F2>;<U01F4>;<U01F6>;<U01F8>;<U01FA>;<U01FC>;<U01FE>;/
<U0200>;<U0202>;<U0204>;<U0206>;<U0208>;<U020A>;<U020C>;<U020E>;/
<U0210>;<U0212>;<U0214>;<U0216>;/
% COLLECTION 8 BASIC GREEK/
<U0386>;<U0388>..<>U038A>;<U038C>;<U038E>;<U038F>;<U0391>..<>U0393>;/
<U0395>..<>U0397>;<U0399>;<U039B>;<U039D>;/
% COLLECTION 9 GREEK SYMBOLS AND COPTIC/
<U03E2>;<U03E4>;<U03E6>;<U03E8>;<U03EA>;<U03EC>;<U03EE>;/
% COLLECTION 10 CYRILLIC/
<U0401>..<>U0403>;<U0405>..<>U0407>;/
<U0409>;<U040B>;<U040D>;<U040F>;<U0411>;<U0413>;<U0415>;<U0417>;/
<U0419>;<U041B>;<U041D>;<U041F>;<U0421>;<U0423>;<U0425>;<U0427>;/
<U0429>;<U042B>;<U042D>;<U042F>;<U0431>;<U0433>;<U0435>;<U0437>;/
<U0439>;<U043B>;<U043D>;<U043F>;<U0441>;<U0443>;<U0445>;<U0447>;/
<U0449>;<U044B>;<U044D>;<U044F>;<U0451>;<U0453>;<U0455>;<U0457>;/
<U0459>;<U045B>;<U045D>;<U045F>;<U0461>;<U0463>;<U0465>;<U0467>;/
<U0469>;<U046B>;<U046D>;<U046F>;<U0471>;<U0473>;<U0475>;<U0477>;/
<U0479>;<U047B>;<U047D>;<U047F>;<U0481>;<U0483>;<U0485>;<U0487>;/
<U0489>;<U048B>;<U048D>;<U048F>;<U0491>;<U0493>;<U0495>;<U0497>;/
<U0499>;<U049B>;<U049D>;<U049F>;<U04A1>;<U04A3>;<U04A5>;<U04A7>;/
<U04A9>;<U04AB>;<U04AD>;<U04AF>;<U04B1>;<U04B3>;<U04B5>;<U04B7>;/
<U04B9>;<U04BB>;<U04BD>;<U04BF>;<U04C1>;<U04C3>;<U04C5>;<U04C7>;/
<U04C9>;<U04CB>;<U04CD>;<U04CF>;<U04D1>;<U04D3>;<U04D5>;<U04D7>;/
<U04D9>;<U04DB>;<U04DD>;<U04DF>;<U04E1>;<U04E3>;<U04E5>;<U04E7>;/
<U04E9>;<U04EB>;<U04ED>;<U04EF>;<U04F1>;<U04F3>;<U04F5>;<U04F7>;/
<U04F9>;<U04FB>;<U04FD>;<U04FF>;/
% COLLECTION 11 ARMENIAN/
<U0531>..<>U0533>;/
% COLLECTION 28 GEORGIAN EXTENDED/
<U10A0>..<>U10C5>;/
% COLLECTION 30 LATIN EXTENDED ADDITIONAL/
<U1E00>;<U1E02>;<U1E04>;<U1E06>;<U1E08>;<U1E0A>;<U1E0C>;<U1E0E>;/
<U1E10>;<U1E12>;<U1E14>;<U1E16>;<U1E18>;<U1E1A>;<U1E1C>;<U1E1E>;/
<U1E20>;<U1E22>;<U1E24>;<U1E26>;<U1E28>;<U1E2A>;<U1E2C>;<U1E2E>;/
<U1E30>;<U1E32>;<U1E34>;<U1E36>;<U1E38>;<U1E3A>;<U1E3C>;<U1E3E>;/
<U1E40>;<U1E42>;<U1E44>;<U1E46>;<U1E48>;<U1E4A>;<U1E4C>;<U1E4E>;/
<U1E50>;<U1E52>;<U1E54>;<U1E56>;<U1E58>;<U1E5A>;<U1E5C>;<U1E5E>;/
<U1E60>;<U1E62>;<U1E64>;<U1E66>;<U1E68>;<U1E6A>;<U1E6C>;<U1E6E>;/
<U1E70>;<U1E72>;<U1E74>;<U1E76>;<U1E78>;<U1E7A>;<U1E7C>;<U1E7E>;/
<U1E80>;<U1E82>;<U1E84>;<U1E86>;<U1E88>;<U1E8A>;<U1E8C>;<U1E8E>;/
<U1E90>;<U1E92>;<U1E94>;/
<U1EA0>;<U1EA2>;<U1EA4>;<U1EA6>;<U1EA8>;<U1EAA>;<U1EAC>;<U1EAE>;/
<U1EB0>;<U1EB2>;<U1EB4>;<U1EB6>;<U1EB8>;<U1EBA>;<U1EBC>;<U1EBE>;/
<U1EC0>;<U1EC2>;<U1EC4>;<U1EC6>;<U1EC8>;<U1ECA>;<U1ECC>;<U1ECE>;/
<U1ED0>;<U1ED2>;<U1ED4>;<U1ED6>;<U1ED8>;<U1EDA>;<U1EDC>;<U1EDE>;/
<U1EE0>;<U1EE2>;<U1EE4>;<U1EE6>;<U1EE8>;<U1EEA>;<U1EEC>;<U1EEE>;/
<U1EF0>;<U1EF2>;<U1EF4>;<U1EF6>;<U1EF8>;/
% COLLECTION 31 GREEK EXTENDED/
<U1F00>..<>U1F0F>;<U1F10>..<>U1F1F>;<U1F20>..<>U1F2F>;<U1F30>..<>U1F3F>;/
<U1F40>..<>U1F4F>;<U1F50>..<>U1F5F>;<U1F60>..<>U1F6F>;/
<U1F70>..<>U1F7F>;<U1F80>..<>U1F8F>;<U1F90>..<>U1F9F>;<U1FA0>..<>U1FAF>;<U1FB0>..<>U1FBC>;/
<U1FC0>..<>U1FCC>;<U1FD0>..<>U1FDB>;<U1FE0>..<>U1FEC>;<U1FF0>..<>U1FFC>;/
%
% The "lower" class reflects the lowercase characters of class "alpha"
lower /
% COLLECTION 1 BASIC LATIN/
<U0061>..<>U007A>;/

```

```

% COLLECTION 2 LATIN-1 SUPPLEMENT/
<U00DF>..

```

```

alpha /
% COLLECTION 1 BASIC LATIN/
  <U0041>..<<U005A>;<U0061>..<<U007A>;/
% COLLECTION 2 LATIN-1 SUPPLEMENT/
  <U00AA>;<U00BA>;<U00C0>..<<U00D6>;<U00D8>..<<U00F6>;<U00F8>..<<U00FF>;/
% COLLECTION 3 LATIN EXTENDED-A/
  <U0100>..<<U017F>;/
% COLLECTION 4 LATIN EXTENDED-B/
  <U0180>..<<U01F5>;<U01FA>..<<U0217>;/
% COLLECTION 5 IPA EXTENSIONS/
  <U0250>..<<U02A8>;/
% COLLECTION 30 LATIN EXTENDED ADDITIONAL/
  <U1E00>..<<U1E9B>;<U1EA0>..<<U1EF9>;/
% COLLECTION 33 SUPERSCRIPTS AND SUBSCRIPTS/
  <U207F>;/
% COLLECTION 8 BASIC GREEK/
  <U0386>;<U0388>..<<U038A>;<U038C>;<U038E>..<<U03A1>;<U03A3>..<<U03CE>;/
% COLLECTION 9 GREEK SYMBOLS AND COPTIC/
  <U03D0>..<<U03D6>;<U03DA>;<U03DC>;<U03DE>;<U03E0>;<U03E2>..<<U03F3>;/
% COLLECTION 31 GREEK EXTENDED/
  <U1F00>..<<U1F15>;<U1F18>..<<U1F1D>;<U1F20>..<<U1F45>;<U1F48>..<<U1F4D>;/
  <U1F50>..<<U1F57>;<U1F59>;<U1F5B>;<U1F5D>;<U1F5F>..<<U1F7D>;/
  <U1F80>..<<U1FB4>;<U1FB6>..<<U1FBC>;<U1FC2>..<<U1FC4>;<U1FC6>..<<U1FCC>;/
  <U1FD0>..<<U1FD3>;<U1FD6>..<<U1FDB>;<U1FE0>..<<U1FEC>;<U1FF2>..<<U1FF4>;/
  <U1FF6>..<<U1FFC>;/
% COLLECTION 10 CYRILLIC/
  <U0401>..<<U040C>;<U040E>..<<U044F>;<U0451>..<<U045C>;<U045E>..<<U0481>;/
  <U0490>..<<U04C4>;<U04C7>..<<U04C8>;<U04CB>..<<U04CC>;<U04D0>..<<U04EB>;/
  <U04EE>..<<U04F5>;<U04F8>..<<U04F9>;/
% COLLECTION 11 ARMENIAN/
  <U0531>..<<U0556>;<U0561>..<<U0587>;/
% COLLECTION 13 HEBREW EXTENDED/
  <U05B0>..<<U05B9>;<U05BB>..<<U05BD>;<U05BF>;<U05C1>..<<U05C2>;/
  <U05D0>..<<U05EA>;<U05F0>..<<U05F2>;/
% COLLECTION 15 ARABIC EXTENDED/
  <U0621>..<<U063A>;<U0641>..<<U064A>;<U0670>..<<U06B7>;<U06BA>..<<U06BE>;/
  <U06C0>..<<U06CE>;<U06D0>..<<U06D3>;<U06D5>..<<U06DC>;<U06E5>..<<U06E8>;/
% COLLECTION 16 DEVANAGARI/
  <U0901>..<<U0903>;<U0905>..<<U0939>;<U093E>..<<U094D>;<U0950>..<<U0952>;/
  <U0958>..<<U0963>;/
% COLLECTION 17 BENGALI/
  <U0981>..<<U0983>;<U0985>..<<U098C>;<U098F>..<<U0990>;/
  <U0993>..<<U09A8>;<U09AA>..<<U09B0>;<U09B2>;<U09B6>..<<U09B9>;/
  <U09BE>..<<U09C4>;<U09C7>..<<U09C8>;<U09CB>..<<U09CD>;<U09DC>..<<U09DD>;/
  <U09DF>..<<U09E3>;<U09F0>..<<U09F1>;/
% COLLECTION 18 GURMUKHI/
  <U0A02>;<U0A05>..<<U0A0A>;<U0A0F>..<<U0A10>;<U0A13>..<<U0A28>;/
  <U0A2A>..<<U0A30>;<U0A32>..<<U0A33>;<U0A35>..<<U0A36>;<U0A38>..<<U0A39>;/
  <U0A3E>..<<U0A42>;<U0A47>..<<U0A48>;<U0A4B>..<<U0A4D>;<U0A59>..<<U0A5C>;/
  <U0A5E>;<U0A74>;/
% COLLECTION 19 GUJARATI/
  <U0A81>..<<U0A83>;<U0A85>..<<U0A8B>;<U0A8D>;<U0A8F>..<<U0A91>;/
  <U0A93>..<<U0AA8>;<U0AAA>..<<U0AB0>;<U0AB2>..<<U0AB3>;<U0AB5>..<<U0AB9>;/
  <U0ABD>..<<U0AC5>;<U0AC7>..<<U0AC9>;<U0ACB>..<<U0ACD>;<U0AD0>;<U0AE0>;/
% COLLECTION 20 ORIYA/
  <U0B01>..<<U0B03>;<U0B05>..<<U0B0C>;<U0B0F>..<<U0B10>;<U0B13>..<<U0B28>;/
  <U0B2A>..<<U0B30>;<U0B32>..<<U0B33>;<U0B36>..<<U0B39>;<U0B3E>..<<U0B43>;/
  <U0B47>..<<U0B48>;<U0B4B>..<<U0B4D>;<U0B5C>..<<U0B5D>;<U0B5F>..<<U0B61>;/
% COLLECTION 21 TAMIL/
  <U0B82>..<<U0B83>;<U0B85>..<<U0B8A>;<U0B8E>..<<U0B90>;<U0B92>..<<U0B95>;/
  <U0B99>..<<U0B9A>;<U0B9C>;<U0B9E>..<<U0B9F>;<U0BA3>..<<U0BA4>;/
  <U0BA8>..<<U0BAA>;<U0BAE>..<<U0BB5>;<U0BB7>..<<U0BB9>;<U0BBE>..<<U0BC2>;/
  <U0BC6>..<<U0BC8>;<U0BCA>..<<U0BCD>;/
% COLLECTION 22 TELUGU/
  <U0C01>..<<U0C03>;<U0C05>..<<U0C0C>;<U0C0E>..<<U0C10>;<U0C12>..<<U0C28>;/
  <U0C2A>..<<U0C33>;<U0C35>..<<U0C39>;<U0C3E>..<<U0C44>;<U0C46>..<<U0C48>;/
  <U0C4A>..<<U0C4D>;<U0C60>..<<U0C61>;/
% COLLECTION 23 KANNADA/
  <U0C82>..<<U0C83>;<U0C85>..<<U0C8C>;<U0C8E>..<<U0C90>;<U0C92>..<<U0CA8>;/
  <U0CAA>..<<U0CB3>;<U0CB5>..<<U0CB9>;<U0CBE>..<<U0CC4>;<U0CC6>..<<U0CC8>;/
  <U0CCA>..<<U0CCD>;<U0CDE>;<U0CE0>..<<U0CE1>;/
% COLLECTION 24 MALAYALAM/
  <U0D02>..<<U0D03>;<U0D05>..<<U0D0C>;<U0D0E>..<<U0D10>;<U0D12>..<<U0D28>;/
  <U0D2A>..<<U0D39>;<U0D3E>..<<U0D43>;<U0D46>..<<U0D48>;<U0D4A>..<<U0D4D>;/
  <U0D60>..<<U0D61>;/
% COLLECTION 25 THAI/
  <U0E01>..<<U0E3A>;<U0E40>..<<U0E4E>;/
% COLLECTION 26 LAO/

```

```

<U0E81>..<<U0E82>;<U0E84>;<U0E87>..<<U0E88>;<U0E8A>;<U0E8D>;/
<U0E94>..<<U0E97>;<U0E99>..<<U0E9F>;<U0EA1>..<<U0EA3>;<U0EA5>;<U0EA7>;/
<U0EAA>..<<U0EAB>;<U0EAD>..<<U0EAE>;<U0EB0>..<<U0EB9>;<U0EBB>..<<U0EBD>;/
<U0EC0>..<<U0EC4>;<U0EC6>;<U0EC8>..<<U0ECD>;<U0EDC>..<<U0EDD>;/
% COLLECTION 72 BASIC TIBETAN/
<U0F00>;<U0F18>..<<U0F19>;<U0F35>;<U0F37>;<U0F39>;<U0F40>..<<U0F47>;/
<U0F49>..<<U0F69>;/
<U0F71>..<<U0F84>;<U0F86>..<<U0F8B>;<U0F90>..<<U0F95>;<U0F97>;/
<U0F99>..<<U0FAD>;<U0FB1>..<<U0FB7>;<U0FB9>;/
% COLLECTION 28 GEORGIAN EXTENDED/
<U10A0>..<<U10C5>;/
% COLLECTION 27 BASIC GEORGIAN/
<U10D0>..<<U10F6>;/
% COLLECTION 50 HIRAGANA/
<U3041>..<<U3093>;<U309B>..<<U309C>;/
% COLLECTION 51 KATAKANA/
<U30A1>..<<U30F6>;<U30FB>..<<U30FC>;/
% COLLECTION 52 BOPOMOF0/
<U3105>..<<U312C>;/
% CJK unified ideographs/
<U4E00>..<<U9FA5>;/
% COLLECTION 71 HANGUL SYLLABLES/
<UAC00>..<<UD7A3>;/
% Miscellaneous/
<U00B5>;<U02B0>..<<U02B8>;<U02BB>;<U02BD>..<<U02C1>;/
<U02D0>..<<U02D1>;<U02E0>..<<U02E4>;<U037A>;<U0559>;<U093D>;<U0B3D>;/
<U1FBE>;<U2160>..<<U2182>;<U3021>..<<U3029>
%
% The "digit" class of the "i18n" FDCC-set is reflecting
% the recommendations in TR 10176 annex A
digit /
% COLLECTION 1 BASIC LATIN/
<U0030>..<<U0039>;/
% COLLECTION 15 ARABIC EXTENDED/
<U0660>..<<U0669>;<U06F0>..<<U06F9>;/
% COLLECTION 16 DEVANAGARI/
<U0966>..<<U096F>;/
% COLLECTION 18 BENGALI/
<U09E6>..<<U09EF>;/
% COLLECTION 18 GURMUKHI/
<U0A66>..<<U0A6F>;/
% COLLECTION 19 GUJARATI/
<U0AE6>..<<U0AEF>;/
% COLLECTION 20 ORIYA/
<U0B66>..<<U0B6F>;/
% COLLECTION 21 TAMIL/
<0>;<U0BE7>..<<U0BEF>;/
% COLLECTION 22 TELUGU/
<U0C66>..<<U0C6F>;/
% COLLECTION 23 KANNADA/
<U0CE6>..<<U0CEF>;/
% COLLECTION 24 MALAYALAM/
<U0D66>..<<U0D6F>;/
% COLLECTION 25 THAI/
<U0E50>..<<U0E59>;/
% COLLECTION 26 LAO/
<U0ED0>..<<U0ED9>;/
% COLLECTION 72 BASIC TIBETAN/
<U0F20>..<<U0F29>;/
% COLLECTION 68 HALFWIDTH AND FULLWIDTH FORMS/
<UFF10>..<<UFF19>
%
outdigit <U0030>..<<U0039>
%
space /
% ISO/IEC 6429/
<U0008>;<U000A>..<<U000D>;/
% COLLECTION 1 BASIC LATIN/
<U0020>;/
% COLLECTION 35 GENERAL PUNCTUATION/
<U2000>..<<U2006>;<U2008>..<<U200B>;/
% COLLECTION 50 CJK SYMBOLS AND PUNCTUATION, HIRAGANA/
<U3000>
%
cntrl <U0000>..<<U001F>;<U007F>..<<U009F>
%
punct /
<U0021>..<<U002F>;<U003A>..<<U0040>;<U005B>..<<U0060>;<U007B>..<<U007E>;/

```

```

<U00A0>..<U00A9>;<U00AB>..<U00B4>;<U00B6>..<U00B9>;<U00BB>..<U00BF>;/
<U00D7>;<U00F7>;/
<U037E>;<U0482>;<U055A>..<U055F>;<U0589>;<U05BE>;<U05C0>;<U05C3>;/
<U05F3>;<U05F4>;<U060C>;<U061B>;<U061F>;<U0640>;<U064B>..<U0652>;/
<U066A>..<U066D>;<U06D4>;<U06DD>..<U06E1>;<U06E9>..<U06EC>;<U10FB>;/
<U2010>..<U2029>;<U2030>..<U2046>;<U20A0>..<U20AC>;<U2100>..<U210B>;/
<U210D>..<U2110>;<U2112>..<U211B>;<U211D>..<U2127>;<U212A>..<U212C>;/
<U212E>..<U2138>;<U2200>..<U22F1>;<U2300>;<U2302>..<U237A>;<U2400>..<U2424>;/
<U2440>..<U244A>;<U2580>..<U2595>;<U25A0>..<U25EF>;<U2600>..<U2613>;/
<U261A>..<U266F>;<U2701>..<U2704>;<U2706>..<U2709>;<U270C>..<U2727>;/
<U2729>..<U274B>;<U274D>;<U274F>..<U2752>;<U2756>;<U2758>..<U275E>;/
<U2761>..<U2767>;<U3000>..<U3020>;<U3030>;<U3036>;<U3037>;<U303F>;<U3164>;/
<U3190>..<U319F>;<U3200>..<U321C>;<U3220>..<U3243>;<U3260>..<U327B>;/
<U327F>..<U32B0>;<U32C0>..<U32CB>;<U32D0>..<U32FE>;<U3300>..<U3376>;/
<U337B>..<U33DD>;<U33E0>..<U33FE>;<UFD3E>;<UFD3F>;<UFE49>..<UFE52>;/
<UFE54>..<UFE66>;<UFE68>..<UFE6B>;<UFEFF>;<UFF01>..<UFF0F>;<UFF1A>..<UFF20>;/
<UFF3B>..<UFF40>;<UFF5B>..<UFF5E>;<UFF61>..<UFF65>;<UFF70>;<UFF9E>..<UFFA0>;/
<UFFE0>..<UFFE6>;<UFFE8>..<UFFEE>;<UFFFC>;<UFFFD>

%
graph /
<U0021>..<U007E>;<U00A0>..<U01F5>;<U01FA>..<U0217>;/
<U0250>..<U02A8>;<U02B0>..<U02DE>;<U02E0>..<U02E9>;<U0300>..<U0345>;/
<U0360>;<U0361>;<U0374>;<U0375>;<U037A>;<U037E>;<U0384>..<U038A>;<U038C>;/
<U038E>..<U03A1>;<U03A3>..<U03CE>;<U03D0>..<U03D6>;<U03DA>;<U03DC>;<U03DE>;/
<U03E0>;<U03E2>..<U03F3>;<U0401>..<U040C>;<U040E>..<U044F>;/
<U0451>..<U045C>;<U045E>..<U0486>;<U0490>..<U04C4>;<U04C7>;<U04C8>;/
<U04CB>;<U04CC>;<U04D0>..<U04EB>;<U04EE>..<U04F5>;<U04F8>;<U04F9>;/
<U0531>..<U0556>;<U0559>..<U055F>;<U0561>..<U0587>;<U0589>;/
<U0591>..<U05A1>;<U05A3>..<U05AF>;<U05B0>..<U05B9>;/
<U05BB>..<U05C4>;<U05D0>..<U05EA>;<U05F0>..<U05F4>;<U060C>;<U061B>;<U061F>;/
<U0621>..<U063A>;<U0640>..<U0652>;<U0660>..<U066D>;<U0670>..<U06B7>;/
<U06BA>..<U06BE>;<U06C0>..<U06CE>;<U06D0>..<U06ED>;<U06F0>..<U06F9>;/
<U0901>..<U0903>;<U0905>..<U0939>;<U093C>..<U094D>;<U0950>..<U0954>;/
<U0958>..<U0970>;<U0981>..<U0983>;<U0985>..<U098C>;<U098F>;<U0990>;/
<U0993>..<U09A8>;<U09AA>..<U09BC>;<U09B2>;<U09B6>..<U09B9>;<U09BC>;/
<U09BE>..<U09C4>;<U09C7>;<U09C8>;<U09CB>..<U09CD>;<U09D7>;<U09DC>;<U09DD>;/
<U09DF>..<U09E3>;<U09E6>..<U09FA>;<U0A02>;<U0A05>..<U0A0A>;<U0A0F>;<U0A10>;/
<U0A13>..<U0A28>;<U0A2A>..<U0A30>;<U0A32>;<U0A33>;<U0A35>;<U0A36>;/
<U0A38>;<U0A39>;<U0A3C>;<U0A3E>..<U0A42>;<U0A47>;<U0A48>;<U0A4B>..<U0A4D>;/
<U0A59>..<U0A5C>;<U0A5E>;<U0A66>..<U0A74>;<U0A81>..<U0A83>;<U0A85>..<U0A8B>;/
<U0A8D>;<U0A8F>..<U0A91>;<U0A93>..<U0AA8>;<U0AAA>..<U0AB0>;/
<U0AB2>;<U0AB3>;<U0AB5>..<U0AB9>;<U0ABC>..<U0AC5>;<U0AC7>..<U0AC9>;/
<U0ACB>..<U0ACD>;<U0AD0>;<U0AE0>;<U0AE6>..<U0AEF>;<U0B01>..<U0B03>;/
<U0B05>..<U0B0C>;<U0B0F>;<U0B10>;<U0B13>..<U0B28>;<U0B2A>..<U0B30>;/
<U0B32>;<U0B33>;<U0B36>..<U0B39>;<U0B3C>..<U0B43>;<U0B47>;<U0B48>;/
<U0B4B>..<U0B4D>;<U0B56>;<U0B57>;<U0B5C>;<U0B5D>;<U0B5F>..<U0B61>;/
<U0B66>..<U0B70>;<U0B82>;<U0B83>;<U0B85>..<U0B8A>;<U0B8E>..<U0B90>;/
<U0B92>..<U0B95>;<U0B99>;<U0B9A>;<U0B9C>;<U0B9E>;<U0B9F>;<U0BA3>;<U0BA4>;/
<U0BA8>..<U0BAA>;<U0BAE>..<U0BB5>;<U0BB7>..<U0BB9>;<U0BBE>..<U0BC2>;/
<U0BC6>..<U0BC8>;<U0BCA>..<U0BCD>;<U0BD7>;<U0BE7>..<U0BF2>;<U0C01>..<U0C03>;/
<U0C05>..<U0C0C>;<U0C0E>;<U0C10>;<U0C12>..<U0C28>;<U0C2A>..<U0C33>;/
<U0C35>..<U0C39>;<U0C3E>..<U0C44>;<U0C46>..<U0C48>;<U0C4A>..<U0C4D>;/
<U0C55>;<U0C56>;<U0C60>;<U0C61>;<U0C66>..<U0C6F>;<U0C82>;<U0C83>;/
<U0C85>..<U0C8C>;<U0C8E>..<U0C90>;<U0C92>..<U0CA8>;<U0CAA>..<U0CB3>;/
<U0CB5>..<U0CB9>;<U0CBE>..<U0CC4>;<U0CC6>..<U0CC8>;<U0CCA>..<U0CCD>;/
<U0CD5>;<U0CD6>;<U0CDE>;<U0CE0>;<U0CE1>;<U0CE6>..<U0CEF>;<U0D02>;<U0D03>;/
<U0D05>..<U0D0C>;<U0D0E>..<U0D10>;<U0D12>..<U0D28>;<U0D2A>..<U0D39>;/
<U0D3E>..<U0D43>;<U0D46>..<U0D48>;<U0D4A>..<U0D4B>;<U0D57>;<U0D60>;<U0D61>;/
<U0D66>..<U0D6F>;<U0E01>..<U0E3A>;<U0E3F>..<U0E5B>;<U0E81>;<U0E82>;<U0E84>;/
<U0E87>;<U0E88>;<U0E8A>;<U0E8D>;<U0E94>..<U0E97>;<U0E99>..<U0E9F>;/
<U0EA1>..<U0EA3>;<U0EA5>;<U0EA7>;<U0EAA>;<U0EAB>;<U0EAD>..<U0EB9>;/
<U0EBB>..<U0EBD>;<U0EC0>..<U0EC4>;<U0EC6>;<U0EC8>..<U0ECD>;<U0ED0>..<U0ED9>;/
<U0EDC>;<U0EDD>;/
<U0F00>..<U0F47>;<U0F49>..<U0F69>;<U0F71>..<U0F7F>;/
<U10A0>..<U10C5>;<U10D0>..<U10F6>;<U10FB>;<U1100>..<U1159>;/
<U115F>..<U11A2>;<U11A8>..<U11F9>;<U1E00>..<U1E9B>;<U1EA0>..<U1EF9>;/
<U1F00>..<U1F15>;<U1F18>..<U1F1D>;<U1F20>..<U1F45>;<U1F48>..<U1F4D>;/
<U1F50>..<U1F57>;<U1F59>;<U1F5B>;<U1F5D>;<U1F5F>..<U1F7D>;<U1F80>..<U1FB4>;/
<U1FB6>..<U1FC4>;<U1FC6>..<U1FD3>;<U1FD6>..<U1FDB>;<U1FDD>..<U1FEF>;/
<U1FF2>..<U1FF4>;<U1FF6>..<U1FFE>;<U2000>..<U202E>;<U2030>..<U2046>;/
<U206A>..<U2070>;<U2074>..<U208E>;<U20A0>..<U20AC>;<U20D0>..<U20E1>;/
<U2100>..<U2138>;<U2153>..<U2182>;<U2190>..<U21EA>;<U2200>..<U22F1>;<U2300>;/
<U2302>..<U237A>;<U2400>..<U2424>;<U2440>..<U244A>;<U2460>..<U24EA>;/
<U2500>..<U2595>;<U25A0>..<U25EF>;<U2600>..<U2613>;<U261A>..<U266F>;/
<U2701>..<U2704>;<U2706>..<U2709>;<U270C>..<U2727>;<U2729>..<U274B>;<U274D>;/
<U274F>..<U2752>;<U2756>;<U2758>..<U275E>;<U2761>..<U2767>;<U2776>..<U2794>;/
<U2798>..<U27AF>;<U27B1>..<U27BE>;<U3000>..<U3037>;<U303F>;<U3041>..<U3094>;/
<U3099>..<U309E>;<U30A1>..<U30FE>;<U3105>..<U312C>;<U3131>..<U318E>;/

```

```

<U3190>..<<U319F>;<U3200>..<<U321C>;<U3220>..<<U3243>;<U3260>..<<U327B>;/
<U327F>..<<U32B0>;<U32C0>..<<U32CB>;<U32D0>..<<U32FE>;<U3300>..<<U3376>;/
<U337B>..<<U33DD>;<U33E0>..<<U33FE>;/
<UAC00>..<<UD7A3>;<U4E00>..<<U9FA5>;/
<UFB00>..<<UFB06>;<UFB13>..<<UFB17>;/
<UFB1E>..<<UFB36>;<UFB38>..<<UFB3C>;<UFB3E>;<UFB40>;<UFB41>;<UFB43>;<UFB44>;/
<UFB46>..<<UFB51>;<UFB53>..<<UFD3F>;<UFD50>..<<UFD8F>;<UFD92>..<<UFD97>;/
<UFD00>..<<UFD0B>;<UFE20>..<<UFE23>;<UFE30>..<<UFE44>;<UFE49>..<<UFE52>;/
<UFE54>..<<UFE66>;<UFE68>..<<UFE6B>;<UFE70>..<<UFE72>;<UFE74>;<UFE76>..<<UFEFC>;/
<UFEFF>;<UFF01>..<<UFF5E>;<UFF61>..<<UFFBE>;<UFFC2>..<<UFFC7>;/
<UFFCA>..<<UFFCF>;<UFFD2>..<<UFFD7>;<UFFDA>..<<UFFDC>;<UFFE0>..<<UFFE6>;/
<UFFE8>..<<UFFEE>;<UFFFC>;<UFFFD>

%
% "print" is by default "graph", and the <space> character
xdigit <U0030>..<<U0039>;<U0041>..<<U0046>;<U0061>..<<U0066>
blank <U0008>;<U0020>;<U2000>..<<U2006>;<U2008>..<<U200B>;<U3000>
%
%
toupper /
(<U0061>,<U0041>);(<U0062>,<U0042>);(<U0063>,<U0043>);(<U0064>,<U0044>);/
(<U0065>,<U0045>);(<U0066>,<U0046>);(<U0067>,<U0047>);(<U0068>,<U0048>);/
(<U0069>,<U0049>);(<U006A>,<U004A>);(<U006B>,<U004B>);(<U006C>,<U004C>);/
(<U006D>,<U004D>);(<U006E>,<U004E>);(<U006F>,<U004F>);(<U0070>,<U0050>);/
(<U0071>,<U0051>);(<U0072>,<U0052>);(<U0073>,<U0053>);(<U0074>,<U0054>);/
(<U0075>,<U0055>);(<U0076>,<U0056>);(<U0077>,<U0057>);(<U0078>,<U0058>);/
(<U0079>,<U0059>);(<U007A>,<U005A>);(<U00E0>,<U00C0>);(<U00E1>,<U00C1>);/
(<U00E2>,<U00C2>);(<U00E3>,<U00C3>);(<U00E4>,<U00C4>);(<U00E5>,<U00C5>);/
(<U00E6>,<U00C6>);(<U00E7>,<U00C7>);(<U00E8>,<U00C8>);(<U00E9>,<U00C9>);/
(<U00EA>,<U00CA>);(<U00EB>,<U00CB>);(<U00EC>,<U00CC>);(<U00ED>,<U00CD>);/
(<U00EE>,<U00CE>);(<U00EF>,<U00CF>);(<U00F0>,<U00D0>);(<U00F1>,<U00D1>);/
(<U00F2>,<U00D2>);(<U00F3>,<U00D3>);(<U00F4>,<U00D4>);(<U00F5>,<U00D5>);/
(<U00F6>,<U00D6>);(<U00F7>,<U00D7>);(<U00F8>,<U00D8>);(<U00F9>,<U00D9>);/
(<U00FA>,<U00DA>);(<U00FB>,<U00DB>);(<U00FC>,<U00DC>);(<U00FD>,<U00DD>);/
(<U00FE>,<U00DE>);(<U00FF>,<U0178>);(<U0101>,<U0100>);(<U0103>,<U0102>);/
(<U0105>,<U0104>);(<U0107>,<U0106>);(<U0109>,<U0108>);(<U010B>,<U010A>);/
(<U010D>,<U010C>);(<U010F>,<U010E>);(<U0111>,<U0110>);(<U0113>,<U0112>);/
(<U0115>,<U0114>);(<U0117>,<U0116>);(<U0119>,<U0118>);(<U011B>,<U011A>);/
(<U011D>,<U011C>);(<U011F>,<U011E>);(<U0121>,<U0120>);(<U0123>,<U0122>);/
(<U0125>,<U0124>);(<U0127>,<U0126>);(<U0129>,<U0128>);(<U012B>,<U012A>);/
(<U012D>,<U012C>);(<U012F>,<U012E>);(<U0131>,<U0049>);/
(<U0133>,<U0132>);(<U0135>,<U0134>);(<U0137>,<U0136>);/
(<U013A>,<U0139>);(<U013C>,<U013B>);(<U013E>,<U013D>);(<U0140>,<U013F>);/
(<U0142>,<U0141>);(<U0144>,<U0143>);(<U0146>,<U0145>);(<U0148>,<U0147>);/
(<U014B>,<U014A>);(<U014D>,<U014C>);(<U014F>,<U014E>);(<U0151>,<U0150>);/
(<U0153>,<U0152>);(<U0155>,<U0154>);(<U0157>,<U0156>);(<U0159>,<U0158>);/
(<U015B>,<U015A>);(<U015D>,<U015C>);(<U015F>,<U015E>);(<U0161>,<U0160>);/
(<U0163>,<U0162>);(<U0165>,<U0164>);(<U0167>,<U0166>);(<U0169>,<U0168>);/
(<U016B>,<U016A>);(<U016D>,<U016C>);(<U016F>,<U016E>);(<U0171>,<U0170>);/
(<U0173>,<U0172>);(<U0175>,<U0174>);(<U0177>,<U0176>);(<U017A>,<U0179>);/
(<U017C>,<U017B>);(<U017E>,<U017D>);(<U017F>,<U0053>);(<U0183>,<U0182>);/
(<U0185>,<U0184>);(<U0188>,<U0187>);(<U018C>,<U018B>);(<U0192>,<U0191>);/
(<U0199>,<U0198>);(<U01A1>,<U01A0>);(<U01A3>,<U01A2>);(<U01A5>,<U01A4>);/
(<U01A8>,<U01A7>);(<U01AD>,<U01AC>);(<U01B0>,<U01AF>);(<U01B4>,<U01B3>);/
(<U01B6>,<U01B5>);(<U01B9>,<U01B8>);(<U01BD>,<U01BC>);(<U01C5>,<U01C4>);/
(<U01C6>,<U01C4>);(<U01C8>,<U01C7>);/
(<U01C9>,<U01C7>);(<U01CB>,<U01CA>);(<U01CC>,<U01CA>);/
(<U01CE>,<U01CD>);(<U01D0>,<U01CF>);(<U01D2>,<U01D1>);(<U01D4>,<U01D3>);/
(<U01D6>,<U01D5>);(<U01D8>,<U01D7>);(<U01DA>,<U01D9>);(<U01DC>,<U01DB>);/
(<U01DD>,<U018E>);(<U01DF>,<U01DE>);(<U01E1>,<U01E0>);(<U01E3>,<U01E2>);/
(<U01E5>,<U01E4>);(<U01E7>,<U01E6>);(<U01E9>,<U01E8>);(<U01EB>,<U01EA>);/
(<U01ED>,<U01EC>);(<U01EF>,<U01EE>);(<U01F2>,<U01F1>);/
(<U01F3>,<U01F1>);(<U01F5>,<U01F4>);(<U01FB>,<U01FA>);(<U01FD>,<U01FC>);/
(<U01FF>,<U01FE>);(<U0201>,<U0200>);(<U0203>,<U0202>);(<U0205>,<U0204>);/
(<U0207>,<U0206>);(<U0209>,<U0208>);(<U020B>,<U020A>);(<U020D>,<U020C>);/
(<U020F>,<U020E>);(<U0211>,<U0210>);(<U0213>,<U0212>);(<U0215>,<U0214>);/
(<U0217>,<U0216>);(<U0253>,<U0181>);(<U0254>,<U0186>);(<U0256>,<U0189>);/
(<U0257>,<U018A>);(<U0259>,<U018F>);(<U025B>,<U0190>);(<U0260>,<U0193>);/
(<U0263>,<U0194>);(<U0268>,<U0197>);(<U0269>,<U0196>);(<U026F>,<U019C>);/
(<U0280>,<U01A6>);/
(<U0272>,<U019D>);(<U0275>,<U019F>);(<U0283>,<U01A9>);(<U0288>,<U01AE>);/
(<U028A>,<U01B1>);(<U028B>,<U01B2>);(<U0292>,<U01B7>);(<U03AC>,<U0386>);/
(<U03AD>,<U0388>);(<U03AE>,<U0389>);(<U03AF>,<U038A>);(<U03B1>,<U0391>);/
(<U03B2>,<U0392>);(<U03B3>,<U0393>);(<U03B4>,<U0394>);(<U03B5>,<U0395>);/
(<U03B6>,<U0396>);(<U03B7>,<U0397>);(<U03B8>,<U0398>);(<U03B9>,<U0399>);/
(<U03BA>,<U039A>);(<U03BB>,<U039B>);(<U03BC>,<U039C>);(<U03BD>,<U039D>);/
(<U03BE>,<U039E>);(<U03BF>,<U039F>);(<U03C0>,<U03A0>);(<U03C1>,<U03A1>);/
(<U03C2>,<U03A3>);(<U03C3>,<U03A3>);(<U03C4>,<U03A4>);(<U03C5>,<U03A5>);/
(<U03C6>,<U03A6>);(<U03C7>,<U03A7>);(<U03C8>,<U03A8>);(<U03C9>,<U03A9>);/
(<U03CA>,<U03AA>);(<U03CB>,<U03AB>);(<U03CC>,<U038C>);(<U03CD>,<U038E>);/

```

```

(<U03CE>, <U038F>); /
(<U03E3>, <U03E2>); (<U03E5>, <U03E4>); (<U03E7>, <U03E6>); (<U03E9>, <U03E8>); /
(<U03EB>, <U03EA>); (<U03ED>, <U03EC>); (<U03EF>, <U03EE>); /
(<U0430>, <U0410>); (<U0431>, <U0411>); (<U0432>, <U0412>); /
(<U0433>, <U0413>); (<U0434>, <U0414>); (<U0435>, <U0415>); (<U0436>, <U0416>); /
(<U0437>, <U0417>); (<U0438>, <U0418>); (<U0439>, <U0419>); (<U043A>, <U041A>); /
(<U043B>, <U041B>); (<U043C>, <U041C>); (<U043D>, <U041D>); (<U043E>, <U041E>); /
(<U043F>, <U041F>); (<U0440>, <U0420>); (<U0441>, <U0421>); (<U0442>, <U0422>); /
(<U0443>, <U0423>); (<U0444>, <U0424>); (<U0445>, <U0425>); (<U0446>, <U0426>); /
(<U0447>, <U0427>); (<U0448>, <U0428>); (<U0449>, <U0429>); (<U044A>, <U042A>); /
(<U044B>, <U042B>); (<U044C>, <U042C>); (<U044D>, <U042D>); (<U044E>, <U042E>); /
(<U044F>, <U042F>); (<U0451>, <U0401>); (<U0452>, <U0402>); (<U0453>, <U0403>); /
(<U0454>, <U0404>); (<U0455>, <U0405>); (<U0456>, <U0406>); (<U0457>, <U0407>); /
(<U0458>, <U0408>); (<U0459>, <U0409>); (<U045A>, <U040A>); (<U045B>, <U040B>); /
(<U045C>, <U040C>); (<U045E>, <U040E>); (<U045F>, <U040F>); (<U0461>, <U0460>); /
(<U0463>, <U0462>); (<U0465>, <U0464>); (<U0467>, <U0466>); (<U0469>, <U0468>); /
(<U046B>, <U046A>); (<U046D>, <U046C>); (<U046F>, <U046E>); (<U0471>, <U0470>); /
(<U0473>, <U0472>); (<U0475>, <U0474>); (<U0477>, <U0476>); (<U0479>, <U0478>); /
(<U047B>, <U047A>); (<U047D>, <U047C>); (<U047F>, <U047E>); (<U0481>, <U0480>); /
(<U0491>, <U0490>); (<U0493>, <U0492>); (<U0495>, <U0494>); (<U0497>, <U0496>); /
(<U0499>, <U0498>); (<U049B>, <U049A>); (<U049D>, <U049C>); (<U049F>, <U049E>); /
(<U04A1>, <U04A0>); (<U04A3>, <U04A2>); (<U04A5>, <U04A4>); (<U04A7>, <U04A6>); /
(<U04A9>, <U04A8>); (<U04AB>, <U04AA>); (<U04AD>, <U04AC>); (<U04AF>, <U04AE>); /
(<U04B1>, <U04B0>); (<U04B3>, <U04B2>); (<U04B5>, <U04B4>); (<U04B7>, <U04B6>); /
(<U04B9>, <U04B8>); (<U04BB>, <U04BA>); (<U04BD>, <U04BC>); (<U04BF>, <U04BE>); /
(<U04C2>, <U04C1>); (<U04C4>, <U04C3>); (<U04C8>, <U04C7>); (<U04CC>, <U04CB>); /
(<U04D1>, <U04D0>); (<U04D3>, <U04D2>); (<U04D5>, <U04D4>); (<U04D7>, <U04D6>); /
(<U04D9>, <U04D8>); (<U04DB>, <U04DA>); (<U04DD>, <U04DC>); (<U04DF>, <U04DE>); /
(<U04E1>, <U04E0>); (<U04E3>, <U04E2>); (<U04E5>, <U04E4>); (<U04E7>, <U04E6>); /
(<U04E9>, <U04E8>); (<U04EB>, <U04EA>); (<U04EF>, <U04EE>); (<U04F1>, <U04F0>); /
(<U04F3>, <U04F2>); (<U04F5>, <U04F4>); (<U04F9>, <U04F8>); (<U0561>, <U0531>); /
(<U0562>, <U0532>); (<U0563>, <U0533>); (<U0564>, <U0534>); (<U0565>, <U0535>); /
(<U0566>, <U0536>); (<U0567>, <U0537>); (<U0568>, <U0538>); (<U0569>, <U0539>); /
(<U056A>, <U053A>); (<U056B>, <U053B>); (<U056C>, <U053C>); (<U056D>, <U053D>); /
(<U056E>, <U053E>); (<U056F>, <U053F>); (<U0570>, <U0540>); (<U0571>, <U0541>); /
(<U0572>, <U0542>); (<U0573>, <U0543>); (<U0574>, <U0544>); (<U0575>, <U0545>); /
(<U0576>, <U0546>); (<U0577>, <U0547>); (<U0578>, <U0548>); (<U0579>, <U0549>); /
(<U057A>, <U054A>); (<U057B>, <U054B>); (<U057C>, <U054C>); (<U057D>, <U054D>); /
(<U057E>, <U054E>); (<U057F>, <U054F>); (<U0580>, <U0550>); (<U0581>, <U0551>); /
(<U0582>, <U0552>); (<U0583>, <U0553>); (<U0584>, <U0554>); (<U0585>, <U0555>); /
(<U0586>, <U0556>); /
(<U1E01>, <U1E00>); (<U1E03>, <U1E02>); (<U1E05>, <U1E04>); /
(<U1E07>, <U1E06>); (<U1E09>, <U1E08>); (<U1E0B>, <U1E0A>); (<U1E0D>, <U1E0C>); /
(<U1E0F>, <U1E0E>); (<U1E11>, <U1E10>); (<U1E13>, <U1E12>); (<U1E15>, <U1E14>); /
(<U1E17>, <U1E16>); (<U1E19>, <U1E18>); (<U1E1B>, <U1E1A>); (<U1E1D>, <U1E1C>); /
(<U1E1F>, <U1E1E>); (<U1E21>, <U1E20>); (<U1E23>, <U1E22>); (<U1E25>, <U1E24>); /
(<U1E27>, <U1E26>); (<U1E29>, <U1E28>); (<U1E2B>, <U1E2A>); (<U1E2D>, <U1E2C>); /
(<U1E2F>, <U1E2E>); (<U1E31>, <U1E30>); (<U1E33>, <U1E32>); (<U1E35>, <U1E34>); /
(<U1E37>, <U1E36>); (<U1E39>, <U1E38>); (<U1E3B>, <U1E3A>); (<U1E3D>, <U1E3C>); /
(<U1E3F>, <U1E3E>); (<U1E41>, <U1E40>); (<U1E43>, <U1E42>); (<U1E45>, <U1E44>); /
(<U1E47>, <U1E46>); (<U1E49>, <U1E48>); (<U1E4B>, <U1E4A>); (<U1E4D>, <U1E4C>); /
(<U1E4F>, <U1E4E>); (<U1E51>, <U1E50>); (<U1E53>, <U1E52>); (<U1E55>, <U1E54>); /
(<U1E57>, <U1E56>); (<U1E59>, <U1E58>); (<U1E5B>, <U1E5A>); (<U1E5D>, <U1E5C>); /
(<U1E5F>, <U1E5E>); (<U1E61>, <U1E60>); (<U1E63>, <U1E62>); (<U1E65>, <U1E64>); /
(<U1E67>, <U1E66>); (<U1E69>, <U1E68>); (<U1E6B>, <U1E6A>); (<U1E6D>, <U1E6C>); /
(<U1E6F>, <U1E6E>); (<U1E71>, <U1E70>); (<U1E73>, <U1E72>); (<U1E75>, <U1E74>); /
(<U1E77>, <U1E76>); (<U1E79>, <U1E78>); (<U1E7B>, <U1E7A>); (<U1E7D>, <U1E7C>); /
(<U1E7F>, <U1E7E>); (<U1E81>, <U1E80>); (<U1E83>, <U1E82>); (<U1E85>, <U1E84>); /
(<U1E87>, <U1E86>); (<U1E89>, <U1E88>); (<U1E8B>, <U1E8A>); (<U1E8D>, <U1E8C>); /
(<U1E8F>, <U1E8E>); (<U1E91>, <U1E90>); (<U1E93>, <U1E92>); (<U1E95>, <U1E94>); /
(<U1E9B>, <U1E9A>); /
(<U1EA1>, <U1EA0>); (<U1EA3>, <U1EA2>); (<U1EA5>, <U1EA4>); (<U1EA7>, <U1EA6>); /
(<U1EA9>, <U1EA8>); (<U1EAB>, <U1EAA>); (<U1EAD>, <U1EAC>); (<U1EAF>, <U1EAE>); /
(<U1EB1>, <U1EB0>); (<U1EB3>, <U1EB2>); (<U1EB5>, <U1EB4>); (<U1EB7>, <U1EB6>); /
(<U1EB9>, <U1EB8>); (<U1EBB>, <U1EBA>); (<U1EBD>, <U1EBC>); (<U1EBF>, <U1EBE>); /
(<U1EC1>, <U1EC0>); (<U1EC3>, <U1EC2>); (<U1EC5>, <U1EC4>); (<U1EC7>, <U1EC6>); /
(<U1EC9>, <U1EC8>); (<U1ECB>, <U1ECA>); (<U1ECD>, <U1ECC>); (<U1ECF>, <U1ECE>); /
(<U1ED1>, <U1ED0>); (<U1ED3>, <U1ED2>); (<U1ED5>, <U1ED4>); (<U1ED7>, <U1ED6>); /
(<U1ED9>, <U1ED8>); (<U1EDB>, <U1EDA>); (<U1EDD>, <U1EDC>); (<U1EDF>, <U1EDE>); /
(<U1EE1>, <U1EE0>); (<U1EE3>, <U1EE2>); (<U1EE5>, <U1EE4>); (<U1EE7>, <U1EE6>); /
(<U1EE9>, <U1EE8>); (<U1EEB>, <U1EEA>); (<U1EED>, <U1EEC>); (<U1EEF>, <U1EEE>); /
(<U1EF1>, <U1EF0>); (<U1EF3>, <U1EF2>); (<U1EF5>, <U1EF4>); (<U1EF7>, <U1EF6>); /
(<U1EF9>, <U1EF8>); (<U1F00>, <U1F08>); (<U1F01>, <U1F09>); (<U1F02>, <U1F0A>); /
(<U1F03>, <U1F0B>); (<U1F04>, <U1F0C>); (<U1F05>, <U1F0D>); (<U1F06>, <U1F0E>); /
(<U1F07>, <U1F0F>); (<U1F10>, <U1F18>); (<U1F11>, <U1F19>); (<U1F12>, <U1F1A>); /
(<U1F13>, <U1F1B>); (<U1F14>, <U1F1C>); (<U1F15>, <U1F1D>); (<U1F20>, <U1F28>); /
(<U1F21>, <U1F29>); (<U1F22>, <U1F2A>); (<U1F23>, <U1F2B>); (<U1F24>, <U1F2C>); /
(<U1F25>, <U1F2D>); (<U1F26>, <U1F2E>); (<U1F27>, <U1F2F>); (<U1F30>, <U1F38>); /

```

```

(<U1F31>, <U1F39>); (<U1F32>, <U1F3A>); (<U1F33>, <U1F3B>); (<U1F34>, <U1F3C>); /
(<U1F35>, <U1F3D>); (<U1F36>, <U1F3E>); (<U1F37>, <U1F3F>); (<U1F40>, <U1F48>); /
(<U1F41>, <U1F49>); (<U1F42>, <U1F4A>); (<U1F43>, <U1F4B>); (<U1F44>, <U1F4C>); /
(<U1F45>, <U1F4D>); (<U1F51>, <U1F59>); (<U1F53>, <U1F5B>); (<U1F55>, <U1F5D>); /
(<U1F57>, <U1F5F>); (<U1F60>, <U1F68>); (<U1F61>, <U1F69>); (<U1F62>, <U1F6A>); /
(<U1F63>, <U1F6B>); (<U1F64>, <U1F6C>); (<U1F65>, <U1F6D>); (<U1F66>, <U1F6E>); /
(<U1F67>, <U1F6F>); (<U1F70>, <U1FBA>); (<U1F71>, <U1FBB>); (<U1F72>, <U1FC8>); /
(<U1F73>, <U1FC9>); (<U1F74>, <U1FCA>); (<U1F75>, <U1FCB>); (<U1F76>, <U1FDA>); /
(<U1F77>, <U1FDB>); (<U1F78>, <U1FF8>); (<U1F79>, <U1FF9>); (<U1F7A>, <U1FEA>); /
(<U1F7B>, <U1FEB>); (<U1F7C>, <U1FFA>); (<U1F7D>, <U1FFB>); (<U1F80>, <U1F88>); /
(<U1F81>, <U1F89>); (<U1F82>, <U1F8A>); (<U1F83>, <U1F8B>); (<U1F84>, <U1F8C>); /
(<U1F85>, <U1F8D>); (<U1F86>, <U1F8E>); (<U1F87>, <U1F8F>); (<U1F90>, <U1F98>); /
(<U1F91>, <U1F99>); (<U1F92>, <U1F9A>); (<U1F93>, <U1F9B>); (<U1F94>, <U1F9C>); /
(<U1F95>, <U1F9D>); (<U1F96>, <U1F9E>); (<U1F97>, <U1F9F>); (<U1FA0>, <U1FA8>); /
(<U1FA1>, <U1FA9>); (<U1FA2>, <U1FAA>); (<U1FA3>, <U1FAB>); (<U1FA4>, <U1FAC>); /
(<U1FA5>, <U1FAD>); (<U1FA6>, <U1FAE>); (<U1FA7>, <U1FAF>); (<U1FB0>, <U1FB8>); /
(<U1FB1>, <U1FB9>); (<U1FB3>, <U1FBC>); (<U1FC3>, <U1FCC>); (<U1FD0>, <U1FD8>); /
(<U1FD1>, <U1FD9>); (<U1FE0>, <U1FE8>); (<U1FE1>, <U1FE9>); (<U1FE5>, <U1FEC>); /
(<U1FF3>, <U1FFC>);
tolower /
(<U0041>, <U0061>); (<U0042>, <U0062>); (<U0043>, <U0063>); (<U0044>, <U0064>); /
(<U0045>, <U0065>); (<U0046>, <U0066>); (<U0047>, <U0067>); (<U0048>, <U0068>); /
(<U0049>, <U0069>); (<U004A>, <U006A>); (<U004B>, <U006B>); (<U004C>, <U006C>); /
(<U004D>, <U006D>); (<U004E>, <U006E>); (<U004F>, <U006F>); (<U0050>, <U0070>); /
(<U0051>, <U0071>); (<U0052>, <U0072>); (<U0053>, <U0073>); (<U0054>, <U0074>); /
(<U0055>, <U0075>); (<U0056>, <U0076>); (<U0057>, <U0077>); (<U0058>, <U0078>); /
(<U0059>, <U0079>); (<U005A>, <U007A>); (<U00C0>, <U00E0>); (<U00C1>, <U00E1>); /
(<U00C2>, <U00E2>); (<U00C3>, <U00E3>); (<U00C4>, <U00E4>); (<U00C5>, <U00E5>); /
(<U00C6>, <U00E6>); (<U00C7>, <U00E7>); (<U00C8>, <U00E8>); (<U00C9>, <U00E9>); /
(<U00CA>, <U00EA>); (<U00CB>, <U00EB>); (<U00CC>, <U00EC>); (<U00CD>, <U00ED>); /
(<U00CE>, <U00EE>); (<U00CF>, <U00EF>); (<U00D0>, <U00F0>); (<U00D1>, <U00F1>); /
(<U00D2>, <U00F2>); (<U00D3>, <U00F3>); (<U00D4>, <U00F4>); (<U00D5>, <U00F5>); /
(<U00D6>, <U00F6>); (<U00D8>, <U00F8>); (<U00D9>, <U00F9>); (<U00DA>, <U00FA>); /
(<U00DB>, <U00FB>); (<U00DC>, <U00FC>); (<U00DD>, <U00FD>); (<U00DE>, <U00FE>); /
(<U0178>, <U00FF>); (<U0100>, <U0101>); (<U0102>, <U0103>); (<U0104>, <U0105>); /
(<U0106>, <U0107>); (<U0108>, <U0109>); (<U010A>, <U010B>); (<U010C>, <U010D>); /
(<U010E>, <U010F>); (<U0110>, <U0111>); (<U0112>, <U0113>); (<U0114>, <U0115>); /
(<U0116>, <U0117>); (<U0118>, <U0119>); (<U011A>, <U011B>); (<U011C>, <U011D>); /
(<U011E>, <U011F>); (<U0120>, <U0121>); (<U0122>, <U0123>); (<U0124>, <U0125>); /
(<U0126>, <U0127>); (<U0128>, <U0129>); (<U012A>, <U012B>); (<U012C>, <U012D>); /
(<U012E>, <U012F>); (<U0130>, <U0069>); /
(<U0132>, <U0133>); (<U0134>, <U0135>); (<U0136>, <U0137>); /
(<U0139>, <U013A>); (<U013B>, <U013C>); (<U013D>, <U013E>); (<U013F>, <U0140>); /
(<U0141>, <U0142>); (<U0143>, <U0144>); (<U0145>, <U0146>); (<U0147>, <U0148>); /
(<U014A>, <U014B>); (<U014C>, <U014D>); (<U014E>, <U014F>); (<U0150>, <U0151>); /
(<U0152>, <U0153>); (<U0154>, <U0155>); (<U0156>, <U0157>); (<U0158>, <U0159>); /
(<U015A>, <U015B>); (<U015C>, <U015D>); (<U015E>, <U015F>); (<U0160>, <U0161>); /
(<U0162>, <U0163>); (<U0164>, <U0165>); (<U0166>, <U0167>); (<U0168>, <U0169>); /
(<U016A>, <U016B>); (<U016C>, <U016D>); (<U016E>, <U016F>); (<U0170>, <U0171>); /
(<U0172>, <U0173>); (<U0174>, <U0175>); (<U0176>, <U0177>); (<U0179>, <U017A>); /
(<U017B>, <U017C>); (<U017D>, <U017E>); (<U0182>, <U0183>); (<U0184>, <U0185>); /
(<U0187>, <U0188>); (<U0256>, <U0189>); (<U018B>, <U018C>); (<U018E>, <U01DD>); /
(<U0191>, <U0192>); (<U0198>, <U0199>); (<U019F>, <U0275>); /
(<U01A0>, <U01A1>); (<U01A2>, <U01A3>); /
(<U01A4>, <U01A5>); (<U01A6>, <U0280>); /
(<U01A7>, <U01A8>); (<U01AC>, <U01AD>); (<U01AF>, <U01B0>); /
(<U01B3>, <U01B4>); (<U01B5>, <U01B6>); (<U01B8>, <U01B9>); (<U01BC>, <U01BD>); /
(<U01C4>, <U01C6>); (<U01C5>, <U01C6>); (<U01C7>, <U01C9>); /
(<U01C8>, <U01C9>); (<U01CA>, <U01CC>); (<U01CB>, <U01CC>); /
(<U01CD>, <U01CE>); (<U01CF>, <U01D0>); (<U01D1>, <U01D2>); /
(<U01D3>, <U01D4>); (<U01D5>, <U01D6>); (<U01D7>, <U01D8>); (<U01D9>, <U01DA>); /
(<U01DB>, <U01DC>); (<U01DE>, <U01DF>); (<U01E0>, <U01E1>); (<U01E2>, <U01E3>); /
(<U01E4>, <U01E5>); (<U01E6>, <U01E7>); (<U01E8>, <U01E9>); (<U01EA>, <U01EB>); /
(<U01EC>, <U01ED>); (<U01EE>, <U01EF>); (<U01F1>, <U01F3>); /
(<U01F2>, <U01F3>); (<U01F4>, <U01F5>); (<U01FA>, <U01FB>); (<U01FC>, <U01FD>); /
(<U01FE>, <U01FF>); (<U0200>, <U0201>); (<U0202>, <U0203>); (<U0204>, <U0205>); /
(<U0206>, <U0207>); (<U0208>, <U0209>); (<U020A>, <U020B>); (<U020C>, <U020D>); /
(<U020E>, <U020F>); (<U0210>, <U0211>); (<U0212>, <U0213>); (<U0214>, <U0215>); /
(<U0216>, <U0217>); (<U0181>, <U0253>); (<U0186>, <U0254>); (<U018A>, <U0257>); /
(<U018E>, <U0258>); (<U018F>, <U0259>); (<U0190>, <U025B>); (<U0193>, <U0260>); /
(<U0194>, <U0263>); (<U0197>, <U0268>); (<U0196>, <U0269>); (<U019C>, <U026F>); /
(<U019D>, <U0272>); (<U01A9>, <U0283>); (<U01AE>, <U0288>); (<U01B1>, <U028A>); /
(<U01B2>, <U028B>); (<U01B7>, <U0292>); (<U0386>, <U03AC>); (<U0388>, <U03AD>); /
(<U0389>, <U03AE>); (<U038A>, <U03AF>); (<U0391>, <U03B1>); (<U0392>, <U03B2>); /
(<U0393>, <U03B3>); (<U0394>, <U03B4>); (<U0395>, <U03B5>); (<U0396>, <U03B6>); /
(<U0397>, <U03B7>); (<U0398>, <U03B8>); (<U0399>, <U03B9>); (<U039A>, <U03BA>); /
(<U039B>, <U03BB>); (<U039C>, <U03BC>); (<U039D>, <U03BD>); (<U039E>, <U03BE>); /
(<U039F>, <U03BF>); (<U03A0>, <U03C0>); (<U03A1>, <U03C1>); (<U03A3>, <U03C3>); /

```

© ISO/IEC 2004. All rights reserved

```

(<U1F3F>,<U1F37>);(<U1F48>,<U1F40>);(<U1F49>,<U1F41>);(<U1F4A>,<U1F42>);/
(<U1F4B>,<U1F43>);(<U1F4C>,<U1F44>);(<U1F4D>,<U1F45>);(<U1F59>,<U1F51>);/
(<U1F5B>,<U1F53>);(<U1F5D>,<U1F55>);(<U1F5F>,<U1F57>);(<U1F68>,<U1F60>);/
(<U1F69>,<U1F61>);(<U1F6A>,<U1F62>);(<U1F6B>,<U1F63>);(<U1F6C>,<U1F64>);/
(<U1F6D>,<U1F65>);(<U1F6E>,<U1F66>);(<U1F6F>,<U1F67>);(<U1FBA>,<U1F70>);/
(<U1FBB>,<U1F71>);(<U1FC8>,<U1F72>);(<U1FC9>,<U1F73>);(<U1FCA>,<U1F74>);/
(<U1FCB>,<U1F75>);(<U1FDA>,<U1F76>);(<U1FDB>,<U1F77>);(<U1FF8>,<U1F78>);/
(<U1FF9>,<U1F79>);(<U1FEA>,<U1F7A>);(<U1FEB>,<U1F7B>);(<U1FFA>,<U1F7C>);/
(<U1FFB>,<U1F7D>);(<U1F88>,<U1F80>);(<U1F89>,<U1F81>);(<U1F8A>,<U1F82>);/
(<U1F8B>,<U1F83>);(<U1F8C>,<U1F84>);(<U1F8D>,<U1F85>);(<U1F8E>,<U1F86>);/
(<U1F8F>,<U1F87>);(<U1F98>,<U1F90>);(<U1F99>,<U1F91>);(<U1F9A>,<U1F92>);/
(<U1F9B>,<U1F93>);(<U1F9C>,<U1F94>);(<U1F9D>,<U1F95>);(<U1F9E>,<U1F96>);/
(<U1F9F>,<U1F97>);(<U1FA8>,<U1FA0>);(<U1FA9>,<U1FA1>);(<U1FAA>,<U1FA2>);/
(<U1FAB>,<U1FA3>);(<U1FAC>,<U1FA4>);(<U1FAD>,<U1FA5>);(<U1FAE>,<U1FA6>);/
(<U1FAF>,<U1FA7>);(<U1FB8>,<U1FB0>);(<U1FB9>,<U1FB1>);(<U1FBC>,<U1FB3>);/
(<U1FCC>,<U1FC3>);(<U1FD8>,<U1FD0>);(<U1FD9>,<U1FD1>);(<U1FE8>,<U1FE0>);/
(<U1FE9>,<U1FE1>);(<U1FEC>,<U1FE5>);(<U1FFC>,<U1FF3>);/

%
% The "combining" class reflects ISO/IEC 10646-1 annex B.1
% That is, all combining characters (level 2+3).
class
    "combining"; /
<U0300>..

```

4.4 LC_COLLATE

A collation sequence definition defines the relative order between collating elements (characters and multicharacter collating elements) in the FDCC-set. This order is expressed in terms of collation values; i.e., by assigning each element one or more collation values (also known as collation weights). This does not imply that applications assign such values, but that ordering of strings using the resultant collation definition in the FDCC-set behaves as if such assignment is done and used in the collation process. The collation sequence definition is used by regular expressions, pattern matching. When no weights are specified the collation sequence definition also is used for sorting, else the weighting defines the sorting. The following capabilities are provided:

- (1) Multicharacter collating elements. Specification of multicharacter collating elements (i.e., sequences of two or more characters to be collated as an entity).
- (2) User-defined ordering of collating elements. Each collating element is assigned a collation value defining its order in the character (or basic) collation sequence. This ordering is used by regular expressions and pattern matching and, unless collation weights are explicitly specified, also as the collation weight to be used in sorting.
- (3) Multiple weights and equivalence classes. Collating elements can be assigned one or more (up to the limit (COLL_WEIGHTS_MAX)) collating weights for use in sorting. The first weight is hereafter referred to as the primary weight.
- (4) One-to Many mapping. A single character is mapped into a string of collating elements.
- (5) Many-to-Many substitution. A string of one or more characters is substituted by another string (or an empty string, i.e., the character or characters are ignored for collation purposes).
- (6) Equivalence class definition. Two or more collating elements have the same collation value (primary weight).
- (7) Ordering by weights. When two strings are compared to determine their relative order, the two strings are first broken up into a series of collating elements, and each successive pair of elements are compared according to the relative primary weights for the elements. If equal, and more than one weight has been assigned, then the pairs of collating elements are recompared according to the relative subsequent weights, until either a pair of collating elements compare unequal or the weights are exhausted.
- (8) Easy reordering of characters. ISO/IEC 14651 has a template for collation specification that with just a few modifications can be culturally correct for a specific culture. Here the "reorder-after" keyword gives a convenient way to modify a FDCC-set template.
- (9) Easy reordering of sections. The template in ISO/IEC 14651 gives an ordering of the sections that may not be culturally acceptable in certain cultures. The keyword "reorder-section-after" gives a convenient way to modify the order of sections in a FDCC-set template.

The following keywords are recognized in a collation sequence definition. Some of them are described in detail in the following subclauses. The keywords are mandatory unless otherwise noted.

copy	Specify the name of an existing FDCC-set to be used as the source for the definition of this category. If this keyword is specified, only the "reorder-after", "reorder-end", "reorder-section-after" and "reorder-section-end" keywords may also be specified. The FDCC-set is copied in source form.
coll_weight_max	Define as a decimal number the number of collation levels that an interpreting system needs to support for this FDCC-set, this value is elsewhere referred to as the COLL_WEIGHT_MAX limit (e.g. in the "order_start" statement). An interpreting system caters for up to 7 collating levels.
section-symbol	Define a section symbol representing a set of collation order statements. The section is defined with

collating-element	the "order_start" keyword until the next "order_start" or "order_end" keyword. This keyword is optional. Define a collating-element symbol representing a multicharacter collating element. This keyword is optional.
collating-symbol	Define one or more collating symbols for use in collation order statements. This keyword is optional.
symbol-equivalence	Define a collating-symbol to be equivalent to another defined collating-symbol.
order_start	Define collation rules. This statement is followed by one or more collation order statements, assigning character collation values and collation weights to collating elements.
order_end	Specify the end of the collation-order statements.
section	Specify a section of collation order statements, and optionally a subrepertoire thereof.
reorder-after	Redefine collating rules. Specify after which collating element the redefinition of collation order takes order. This statement is followed by one or more collation order statements, reassigning character collation values and collation weights to collating elements.
reorder-end	Specify the end of the "reorder-after" collating order statements.
reorder-section-after	Redefine the order of sections. This statement is followed by one or more section symbols, reassigning character collation values and collation weights to collating elements.
reorder-section-end	Specify the end of the "reorder-section" section order statements.

4.4.1 Collation statements

The "order_start", "reorder-after" and "section" keywords are followed by collating statements. The syntax for the collating statements is

```
"%s %s;%s;...;%s\n", <collating-identifier>, <weight>, <weight>, ...
```

Each <collating-identifier> consists of either a character (in any of the forms defined in 4.1.1), a <collating-element>, a <collating-symbol>, an ellipsis, or the special symbol "UNDEFINED". The weights for each of the collation elements determines the character collation sequence - such that each collation statement does not need to be in collation order, and weights could be rearranged via for example the "reorder-after" keyword. No character has any specific predetermined placement in the collation sequence. The order in which collating elements are specified determines the character collation sequence, such that each collating element compares less than the elements following it.

A <collating-element> is used to specify multicharacter collating elements, and indicates that the character sequence specified via the <collating-element> is to be collated as a unit and in the relative order specified by its place in the list of collating statements.

A <collating-symbol> is used to define a position in the relative order for use in weights.

The absolute ellipsis symbol ("...") specifies that a sequence of characters collate according to their encoded character values. It is interpreted as indicating that all characters with a coded character set value higher than the value of the character in the preceding line, and lower than the coded character set value for the character in the following line, in the current coded character set, are placed in the character collation order between the previous and the following character in ascending order according to their coded character set values. An initial ellipsis is interpreted as if the preceding line specified the <NUL> character, and a trailing ellipsis as if the following line specified the highest coded character set value in the current coded character set. An ellipsis is treated as invalid if the preceding or following lines do not specify characters in the current coded character set. The use of the ellipsis symbol ties the definition to a specific coded character set and may preclude the definition from being portable between applications, and is depreciated. Symbolic ellipses may be used as the ellipses symbol, but generating symbolic character names, and thus have a better chance of portability between applications.

The symbolic ellipses (".." or "...") specifies a sequence of collating statements. It is interpreted as indicating that all characters with symbolic names higher than the symbolic name of the character in the preceding line, and lower in the sequence of symbolic names for the character in the following line, is placed in the character collation order between the previous and the following character in ascending order.

The symbol "UNDEFINED" is interpreted as including all coded character set values not specified explicitly or via the ellipsis or one of the symbolic ellipses symbols. Such characters are inserted in the character collation order at the point indicated by the symbol, and in ascending order according to their coded character set values. If no "UNDEFINED" symbol is specified, and the current coded character set contains characters not specified in this clause, the utility issues a warning message and place such characters at the end of the character collation order.

The optional operands for each collation-element are used to define the primary, secondary, or subsequent weights for the collating element. The first operand specifies the relative primary weight, the second the relative secondary weight, and so on. Two or more collation-elements can be assigned the same weight; they belong to the same equivalence class if they have the same primary weight. Collation behaves as if, for each weight level, "IGNORE"d elements are removed. Then each successive pair of elements is compared according to the relative weights for the elements. If the two strings compare equal, the process is repeated for the next weight level, up to the limit "COLL_WEIGHTS_MAX" of the associated FDCC-set.

Weights are expressed as characters (in any of the forms specified here), <collating-symbol>s, <collating-element>s, an ellipsis, or the special symbol "IGNORE". A single character, a <collating-symbol>, or a <collating-element> represent the relative order in the character collating sequence of the character or symbol, rather than the character or characters themselves.

One-to-many mapping is indicated by specifying two or more concatenated characters or symbolic names. Thus, if the character <ss> is given the string <s><s> as a weight,

comparisons are performed as if all occurrences of the character `<ss>` are replaced by `<s><s>`. If it is desirable to define `<ss>` and `<s><s>` as an equivalence class, then a collating-element must be defined for the string "ss", as in the example below.

All characters specified via an ellipsis are by default assigned unique weights, equal to the relative order of characters. Characters specified via an explicit or implicit "UNDEFINED" special symbol are by default assigned the same primary weight (i.e., belong to the same equivalence class). An ellipsis symbol as a weight is interpreted to mean that each character in the sequence has unique weights, equal to the relative order of their character in the character collation sequence. Secondary and subsequent weights have unique values. The use of the ellipsis as a weight is treated as an error if the collating element is neither an ellipsis nor the special symbol "UNDEFINED".

The special keyword "IGNORE" as a weight indicates that when strings are compared using the weights at the level where "IGNORE" is specified, the collating element is ignored; i.e., as if the string did not contain the collating element. In regular expressions and pattern matching, all characters that are "IGNORE"d in their primary weight form an equivalence class.

A `<comment_character>` occurring where the delimiter ";" may occur, terminates the collating statement.

An empty operand is interpreted as the collating-element itself.

For example, the collation statement

```
<a>    <a>;<a>
```

is equal to

```
<a>
```

An ellipsis (absolute or symbolic) can be used as an operand if the collating-element was an ellipsis, and is interpreted as the value of each character defined by the ellipsis.

Example:

```
collating-element <ch> from "<c><h>"
collating-element <Ch> from "<C><h>"
order_start      forward;backward
UNDEFINED        IGNORE;IGNORE
<LOW>
<space>          <LOW>;<space>
...              <LOW>;
<a>              <a>;<a>
<a'>             <a>;<a'>
<A>              <a>;<A>
<A'>             <a>;<A'>
<ch>             <ch>;<ch>
<Ch>             <ch>;<Ch>
<s>              <s>;<s>
<ss>             "<s><s>";"<ss><ss>"
order_end
```

This example is interpreted as follows:

- (1) The UNDEFINED means that all characters not specified in this definition (explicitly or via the ellipsis) is ignored.

- (2) <LOW> defines the first collating weight, and thus the lowest weight in this example.
- (3) All characters between <space> and <a> have the same primary equivalence class <LOW> and individual secondary weights based on their ordinal encoded values. (The use of absolute ellipses is deprecated, but used here to illustrate generic use of ellipses. Symbolic ellipses should be used instead).
- (4) All characters based on the upper or lowercase character "a" belong to the same primary equivalence class.
- (5) The multicharacter collating element <c><h> is represented by the collating symbol <ch> and belongs to the same primary equivalence class as the multicharacter collating element <C><H>.
- (6) The <ss> collating element has two weights on the primary level, and it is in the same primary equivalence class as two consecutive <s>-es; on the secondary level the collating element has two weights of the equivalence class <ss>.

4.4.2 "copy" keyword

This keyword specifies the name of an existing FDCC-set to be used as the source for the definition of this category. The syntax is

```
"copy %s\n", <FDCC-set-name>
```

The <FDCC-set-name> consists of one or more characters (in any of the forms defined in 4.1.1). The FDCC-set is copied in source form.

4.4.3 "coll_weight_max" keyword

This keyword defines as a decimal number the number of collation levels that an interpreting system needs to support. An interpreting system caters for up to 7 collating levels. The syntax is

```
"coll_weight_max %d\n", <value>
```

4.4.4 "section-symbol" keyword

This keyword is used to define symbols for use in section related statements; such as the "order_start", and "reorder-section-after" keywords and section-reordering statements. The syntax is

```
"section-symbol %s\n", <section-symbol>
```

The <section-symbol> is a symbolic name, enclosed between angle brackets (< and >), and does not duplicate any symbolic name in the current charmap (if any), or any other symbolic name defined in this collation definition. A <section-symbol> defined via this keyword is only defined within the LC_COLLATE category.

```
Example:
section-symbol <LATIN>
section-symbol <ARABIC>
```

4.4.5 "collating-element" keyword

In addition to the collating elements in the character set, the collating-element keyword is used to define multicharacter collating elements. The syntax is