Information technology —

Specification method for cultural conventions

Technologies de l'information —

Méthode de modélisation des conventions culturelles

1

Contents Page

28   **Foreword**
29
30   ISO (the International Organization for Standardization) and IEC (the International
31   Electrotechnical Commission) form the specialized system for worldwide standardization.
32   National bodies that are members of ISO or IEC participate in the development of
33   International Standards through technical committees established by the respective
34   organization to deal with particular fields of technical activity. ISO and IEC technical
35   committees collaborate in fields of mutual interest. Other international organizations,
36   governmental and non-governmental, in liaison with ISO and IEC, also take part in the
37   work. In the field of information technology, ISO and IEC have established a joint
38   technical committee, ISO/IEC JTC 1.
39
40   The main task of a technical committee is to prepare International Standards but in
41   exceptional circumstances, the publication of a Technical Report of one of the following
42   types may be proposed:
43
44   - type 1, when the required support cannot be obtained for the publication of an
45   International Standard, despite repeated efforts;
46
47   - type 2, when the subject is still under technical development or where for any
48   other reason there is the future but not immediate possibility of an agreement on an
49   International Standard;
50
51   - type 3, when a technical committee has collected data of a different kind from
52   that which is normally published as an International Standard ("state of the art", for
53   example).
54
55   Technical Reports are drafted in accordance with the rules given in the ISO/IEC
56   Directives, Part 3.
57
58   Technical Reports of types 1 and 2 are subject to review within three years of publication,
59   to decide whether they can be transformed into International Standards. Technical Report
60   of type 3 do not necessarily have to be reviewed until the date they provide are considered
61   to be no longer valid or useful.
62
63   ISO/IEC TR 14652 is a Technical Report type 1, and it was prepared by Joint Technical
64   Committee ISO/IEC JTC 1, *Information technology*, Subcommittee 22, *Programming*
65   *languages, their environments and system software interfaces*.
66
67   The Annexes A, B, C and D of this Technical Report are for information only.

68 **Introduction**
69
70 This Technical Report defines a general mechanism to specify cultural conventions, and it
71 defines formats for a number of specific cultural conventions in the areas of character
72 classification and conversion, sorting, number formatting, monetary formatting, date
73 formatting, message display, paper formats, addressing of persons, postal address
74 formatting, telephone number handling, and a way to specify how much is covered and the
75 status of it.
76
77 There are a number of benefits coming from this standard:
78
79 Rigid specification   Using this Technical Report, a user can rigidly specify a
80   number of the cultural conventions that apply to the
81   information technology environment of the user.
82
83 Cultural adaptability   If an application has been designed and built in a
84   cultural neutral manner, the application may use the
85   specifications as data to its APIs, and thus the same
86   application may accommodate different users in a
87   culturally acceptable way to each of the users, without
88   change of the binary application.
89
90 Productivity   This standard specifies those cultural conventions and
91   how to specify data for them. With those data an
92   application developer is relieved from getting the
93   different information to support all the cultural
94   environments for the expected customers of the product.
95   The application developer is thus ensured of culturally
96   correct behavior as specified by the customer, and
97   possibly more markets may be reached as customers may
98   have the possibility to provide the data themselves for
99   markets that were not targeted.
100
101 Uniform behaviour   When a number of applications share one cultural
102   specification, which may be supplied from the user or a
103   built-in nature, their behaviour for cultural adaptation
104   become uniform.
105
106 The specification format is independent of platforms and specific encoding, and targeted to
107 be usable from a wide range of programming languages.
108
109 A number of cultural conventions, such as spelling, hyphenation rules and terminology, are
110 not specifiable with this standard, but the standard provides mechanisms to define new
111 categories and also new keywords within existing categories. An internationalized
112 application may take advantage of information provided with the FDCC-set (such as the
113 language) to provide further internationalized services to the user.
114
115 This Technical Report defines a format compatible with the one used in the International
116 String Ordering standard, ISO/IEC 14651. This Technical Report is backwards compatible
117 with the ISO/IEC 9945-2:1993 POSIX shell and utilities standard, particulary its clauses

118   2.4 and 2.5. The major extensions from that text are listed in annex A. This Technical
119   Report has enhanced functionality in a number of areas such as ISO/IEC 10646 support,
120   more classification of characters, transliteration, dual (multi) currency support, enhanced
121   date and time formatting, paper size identification, personal name writing, postal address
122   formatting, telephone number handling, and management of categories. There is enhanced
123   support for character sets including ISO/IEC 2022 handling and an enhanced method to
124   separate the specification of cultural conventions from an actual encoding via a description
125   of the character repertoire employed. A standard set of values for all the categories has
126   been defined covering the repertoire of ISO/IEC 10646-1.
127

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 Standard, 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 programming languages.

2 NORMATIVE REFERENCES

The following normative documents contain provisions which, through reference in this text, constitute provisions of this Technical Report. For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. However, parties to agreements based on this Technical Report are encouraged to investigate the possibility of applying the most recent editions of the normative documents indicated below. For undated references, the latest edition of the normative document referred to applies. Members of ISO and IEC maintain registers of currently valid Technical Reports.

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 9945-2:1993, *Information technology - Portable Operating System Interface (POSIX) - Part 2: Shell and Utilities.*

ISO/IEC 10646-1:1993, *Information technology - Universal Multiple-Octet Coded Character Set (UCS) - Part 1: Architecture and Basic Multilingual Plane (including Cor.1 and AMD 1-9).*

ISO/IEC 14651, *Information technology - International string ordering - Method for comparing character strings and description of a default tailorable ordering.*

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

178
179 ## 3   TERMS, DEFINITIONS AND NOTATIONS
180
181 ### 3.1   Terms and definitions
182
183 For the purposes of this Technical Report, the terms and definitions given in the following
184 apply.
185
186 **3.1.1**
187 **byte**:
188 An individually addressable unit of data storage that is equal to or larger than an octet,
189 used to store a character or a portion of a character.
190
191 A byte is composed of a contiguous sequence of bits, the number of which is
192 implementation defined. The least significant bit is called the low-order bit; the most
193 significant bit is called the high-order bit.
194
195 **3.1.2**
196 **character**:
197 A member of a set of elements used for the organization, control or representation of data.
198
199 **3.1.3**
200 **coded character**:
201 A sequence of one or more bytes representing a single character.
202
203 **3.1.4**
204 **text file**:
205 A file that contains characters organized into one or more lines.
206
207 **3.1.5**
208 **cultural convention**:
209 A data item for information technology that may vary dependent on language, territory, or
210 other cultural habits.
211
212 **3.1.6**
213 **FDCC-set**:
214 A Set of Formal Definitions of Cultural Conventions. The definition of the subset of a
215 user's information technology environment that depends on language and cultural conven-
216 tions. Note: the FDCC-set is a superset of the "locale" term in C and POSIX.
217
218 **3.1.7**
219 **charmap**:
220 A definition of a mapping between symbolic character names and character codes, plus
221 related information"
222
223 **3.1.8**
224 **repertoiremap**:
225 A definition of a mapping between symbolic character names and characters for the
226 repertoire of characters used in a FDCC-set, further described in clause 6.
227

228 **3.1.9**
229 **character class**:
230 A named set of characters sharing an attribute associated with the name of the class.
231
232 **3.1.10**
233 **collation**:
234 The logical ordering of strings according to defined precedence rules.
235
236 **3.1.11**
237 **collating element**:
238 The smallest entity used to determine logical ordering.
239
240 See collating sequence. A collating element shall consist of either a single character, or
241 two or more characters collating as a single entity. The LC_COLLATE category in the
242 associated FDCC-set determines the set of collating elements.
243
244 **3.1.12**
245 **multicharacter collating element**:
246 A sequence of two or more characters that collate as an entity.
247
248 For example, in some languages two characters are sorted as one letter, as in the case for
249 Danish and Norwegian "aa".
250
251 **3.1.13**
252 **collating sequence**:
253 The relative order of collating elements as determined by the setting of the LC_COLLATE
254 category in the applied FDCC-set.
255
256 **3.1.14**
257 **equivalence class**:
258 A set of collating elements with the same primary collation weight.
259
260 Elements in an equivalence class are typically elements that naturally group together, such
261 as all accented letters based on the same letter.
262
263 The collation order of elements within an equivalence class is determined by the weights
264 assigned on any subsequent levels after the primary weight.
265
266 **3.2   Notations**
267
268 The following notations and common conventions for specifications apply to this standard:
269
270 **3.2.1   Notation for defining syntax**
271
272 In this standard, the description of an individual record in a FDCC-set is done using the
273 syntax notation given in the following.
274
275 The syntax notation looks as follows:
276
277   "<format>",[<arg1>,<arg2>,...,<argn>]

278   The <format> is given in a format string enclosed in double quotes, followed by a number
279   of parameters, separated by commas. It is similar to the format specification defined in
280   clause 2.12 in the ISO/IEC 9945-2:1993 standard and the format specification used in C
281   language printf() function. The format of each parameter is given by an escape sequence
282   as follows:
283
284   %s   specifies a string
285   %d   specifies a decimal integer
286   %c   specifies a character
287   %o   specifies an octal integer
288   %x   specifies a hexadecimal integer
289
290   A " " (an empty character position) in the syntax string represent one or more <blank>
291   characters.
292
293   All other characters in the format string except
294
295   %%   specifies a single %
296   \n   specifies an end-of-line
297
298   represent themselves.
299
300   The notation "..." is used to specify that repetition of the previous specification is optional,
301   and this is done in both the format string and in the parameter list.
302
303
304   **3.2.3 Portable character set**
305
306   A set of symbolic names for characters in Table 1, which is called the portable character
307   set, is used in character description text of this specification. The first eight entries in
308   Table 1 are defined in ISO/IEC 6429 and others are defined in ISO/IEC 10646-1.
309
310   **Table 1: Portable character set**
311

```
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
```

4

338	<period>	.	<U002E>	FULL STOP
339	<slash>	/	<U002F>	SOLIDUS
340	<solidus>	/	<U002F>	SOLIDUS
341	<zero>	0	<U0030>	DIGIT ZERO
342	<one>	1	<U0031>	DIGIT ONE
343	<two>	2	<U0032>	DIGIT TWO
344	<three>	3	<U0033>	DIGIT THREE
345	<four>	4	<U0034>	DIGIT FOUR
346	<five>	5	<U0035>	DIGIT FIVE
347	<six>	6	<U0036>	DIGIT SIX
348	<seven>	7	<U0037>	DIGIT SEVEN
349	<eight>	8	<U0038>	DIGIT EIGHT
350	<nine>	9	<U0039>	DIGIT NINE
351	<colon>	:	<U003A>	COLON
352	<semicolon>	;	<U003B>	SEMICOLON
353	<less-than-sign>	<	<U003C>	LESS-THAN SIGN
354	<equals-sign>	=	<U003D>	EQUALS SIGN
355	<greater-than-sign>	>	<U003E>	GREATER-THAN SIGN
356	<question-mark>	?	<U003F>	QUESTION MARK
357	<commercial-at>	@	<U0040>	COMMERCIAL AT
358	<A>	A	<U0041>	LATIN CAPITAL LETTER A
359	<B>	B	<U0042>	LATIN CAPITAL LETTER B
360	<C>	C	<U0043>	LATIN CAPITAL LETTER C
361	<D>	D	<U0044>	LATIN CAPITAL LETTER D
362	<E>	E	<U0045>	LATIN CAPITAL LETTER E
363	<F>	F	<U0046>	LATIN CAPITAL LETTER F
364	<G>	G	<U0047>	LATIN CAPITAL LETTER G
365	<H>	H	<U0048>	LATIN CAPITAL LETTER H
366	<I>	I	<U0049>	LATIN CAPITAL LETTER I
367	<J>	J	<U004A>	LATIN CAPITAL LETTER J
368	<K>	K	<U004B>	LATIN CAPITAL LETTER K
369	<L>	L	<U004C>	LATIN CAPITAL LETTER L
370	<M>	M	<U004D>	LATIN CAPITAL LETTER M
371	<N>	N	<U004E>	LATIN CAPITAL LETTER N
372	<O>	O	<U004F>	LATIN CAPITAL LETTER O
373	<P>	P	<U0050>	LATIN CAPITAL LETTER P
374	<Q>	Q	<U0051>	LATIN CAPITAL LETTER Q
375	<R>	R	<U0052>	LATIN CAPITAL LETTER R
376	<S>	S	<U0053>	LATIN CAPITAL LETTER S
377	<T>	T	<U0054>	LATIN CAPITAL LETTER T
378	<U>	U	<U0055>	LATIN CAPITAL LETTER U
379	<V>	V	<U0056>	LATIN CAPITAL LETTER V
380	<W>	W	<U0057>	LATIN CAPITAL LETTER W
381	<X>	X	<U0058>	LATIN CAPITAL LETTER X
382	<Y>	Y	<U0059>	LATIN CAPITAL LETTER Y
383	<Z>	Z	<U005A>	LATIN CAPITAL LETTER Z
384	<left-square-bracket>	[	<U005B>	LEFT SQUARE BRACKET
385	<backslash>	\	<U005C>	REVERSE SOLIDUS
386	<reverse-solidus>	\	<U005C>	REVERSE SOLIDUS
387	<right-square-bracket>	]	<U005D>	RIGHT SQUARE BRACKET
388	<circumflex-accent>	^	<U005E>	CIRCUMFLEX ACCENT
389	<circumflex>	^	<U005E>	CIRCUMFLEX ACCENT
390	<low-line>	_	<U005F>	LOW LINE
391	<underscore>	_	<U005F>	LOW LINE
392	<grave-accent>	`	<U0060>	GRAVE ACCENT
393	<a>	a	<U0061>	LATIN SMALL LETTER A
394	<b>	b	<U0062>	LATIN SMALL LETTER B
395	<c>	c	<U0063>	LATIN SMALL LETTER C
396	<d>	d	<U0064>	LATIN SMALL LETTER D
397	<e>	e	<U0065>	LATIN SMALL LETTER E
398	<f>	f	<U0066>	LATIN SMALL LETTER F
399	<g>	g	<U0067>	LATIN SMALL LETTER G
400	<h>	h	<U0068>	LATIN SMALL LETTER H
401	<i>	i	<U0069>	LATIN SMALL LETTER I
402	<j>	j	<U006A>	LATIN SMALL LETTER J
403	<k>	k	<U006B>	LATIN SMALL LETTER K
404	<l>	l	<U006C>	LATIN SMALL LETTER L
405	<m>	m	<U006D>	LATIN SMALL LETTER M
406	<n>	n	<U006E>	LATIN SMALL LETTER N
407	<o>	o	<U006F>	LATIN SMALL LETTER O
408	<p>	p	<U0070>	LATIN SMALL LETTER P
409	<q>	q	<U0071>	LATIN SMALL LETTER Q
410	<r>	r	<U0072>	LATIN SMALL LETTER R
411	<s>	s	<U0073>	LATIN SMALL LETTER S
412	<t>	t	<U0074>	LATIN SMALL LETTER T
413	<u>	u	<U0075>	LATIN SMALL LETTER U
414	<v>	v	<U0076>	LATIN SMALL LETTER V
415	<w>	w	<U0077>	LATIN SMALL LETTER W

```
416   <x>                        x              <U0078>    LATIN SMALL LETTER X
417   <y>                        y              <U0079>    LATIN SMALL LETTER Y
418   <z>                        z              <U007A>    LATIN SMALL LETTER Z
419   <left-brace>               {              <U007B>    LEFT CURLY BRACKET
420   <left-curly-bracket>       {              <U007B>    LEFT CURLY BRACKET
421   <vertical-line>            |              <U007C>    VERTICAL LINE
422   <right-brace>              }              <U007D>    RIGHT CURLY BRACKET
423   <right-curly-bracket>      }              <U007D>    RIGHT CURLY BRACKET
424   <tilde>                    ~              <U007E>    TILDE
425
```

426   This Technical Report may use other symbolic character names than the above in
427   examples, to illustrate the use of the range of symbols allowed by the syntax specified in
428   4.1.1.
429

430   4   FDCC-set

431

432   A FDCC-set is the definition of the subset of a user's information technology environment
433   that depends on language and cultural conventions. It is made up from one or more
434   categories. Each category is identified by its name and controls specific aspects of the
435   behaviour of components of the system. This Technical Report defines the following
436   categories:
437

438   LC_IDENTIFICATION   Versions and status of categories
439   LC_CTYPE   Character classification, case conversion and code
440   transformation.
441   LC_COLLATE   Collation order.
442   LC_TIME   Date and time formats.
443   LC_NUMERIC   Numeric, non-monetary formatting.
444   LC_MONETARY   Monetary formatting.
445   LC_MESSAGES   Formats of informative and diagnostic messages and
446   interactive responses.
447   LC_PAPER   Paper format
448   LC_NAME   Format of writing personal names
449   LC_ADDRESS   Format of postal addresses
450   LC_TELEPHONE   Format for telephone numbers, and other telephone
451   information
452

453   In future editions of this Technical Report further categories may be added. Other category
454   names beginning with the 3 characters "LC_" are intended for future standardization,
455   except for category names beginning with the five characters "LC_X_" which shall not be
456   used for future addition of categories specified in this Technical Report. An application
457   may thus use category names beginning with the five characters "LC_X_" for application
458   defined categories to avoid clashes with future standardized categories.
459

460   This Technical Report also defines an FDCC-set named "i18n" with values for some of the
461   above categories in order to simplify FDCC-set descriptions for a number of cultures. The
462   contents of "i18n" categories should not necessarily be considered as the most commonly
463   accepted values, while it in many cases could be the recommended values.
464

465   4.1   FDCC-set definition

466

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

470   The **FDCC-set definition text** shall contain one or more FDCC-set category source
471   definitions, and shall not contain more than one definition for the same FDCC-set
472   category. If the text contains source definitions for more than one category, application-
473   defined categories, if present, shall appear after the categories defined by this clause. A
474   category source definition shall contain either the definition of a category or a copy
475   directive. In the event that some of the information for a FDCC-set category, as specified
476   in this Technical Report, is missing from the FDCC-set source definition, the behaviour of
477   that category, if it is referenced, is unspecified. A FDCC-set category is the normal way of
478   specifying a single FDCC.
479
480   There are no **naming conventions** for FDCC-sets specified in this Technical Report, but
481   ISO/IEC 15897:1999 specifies naming rules for POSIX locales, charmaps and
482   repertoiremaps, that may also be applied to FDCC-sets, charmaps and repertoiremaps
483   specified according to this Technical Report.
484
485   A **category source definition** shall consist of a category header, a category body, and a
486   category trailer. A category header shall consist of the character string naming of the
487   category, beginning with the characters "LC_". The category trailer shall consist of the
488   string "END", followed by one or more "blank"s and the string used in the corresponding
489   category header.
490
491   The **category body** shall consist of one or more lines of text. Each line shall be one of the
492   following:
493
494   - a line containing an identifier, optionally followed by one or more operands. Identifiers
495   shall be either keywords, identifying a particular FDCC, or collating elements, or
496   section symbols,
497   - one of transliteration statements defined in 4.3.
498
499   In addition to the keywords defined in this Technical Report, the source can contain
500   application-defined keywords. Each **keyword** within a category shall have a unique name
501   (i.e., two categories can have a commonly-named keyword); no keyword shall start with
502   the characters "LC_". Identifiers shall be separated from the operands by one or more
503   "blank"s.
504
505   **Operands** shall be characters, collating elements, section symbols, or strings of characters.
506   Strings shall be enclosed in double-quotes. Literal double-quotes within strings shall be
507   preceded by the <escape character>, described below. When a keyword is followed by
508   more than one operand, the operands shall be separated by semicolons; "blank"s shall be
509   allowed before and/or after a semicolon.
510
511

512   **4.1.1   Character representation**
513

514   Individual characters, characters in strings, and collating elements shall be represented
515   using symbolic names, UCS notation or characters themselves, or as octal, hexadecimal, or
516   decimal constants as defined below. When constant notation is used, the resultant
517   FDCC-set definitions need not be portable between systems.
518

519   (0)   The left angle bracket (<) is a reserved symbol, denoting the

520 start of a symbolic name; when used to represent itself
521 outside a symbolic name it shall be preceded by the escape
522 character.
523

524 (1) A character can be represented via a **symbolic name**,
525 enclosed within angle brackets (< and >). The symbolic
526 name, including the angle brackets, shall exactly match a
527 symbolic name defined in a charmap or a repertoiremap to
528 be used, and shall be replaced by a character value
529 determined from the value associated with the symbolic
530 name in the charmap or a value associated via a
531 repertoiremap. Repertoiremaps have predefined symbolic
532 names for UCS characters, see clause 6. A FDCC-set may
533 also use the UCS notation of clause 6 to represent characters,
534 without a repertoiremap being defined for the FDCC-set. Use
535 of the escape character or a right angle bracket within a
536 symbolic name shall be invalid unless the character is
537 preceded by the escape character.
538

539 Example: <c>;<c-cedilla> "<M><a><y>"
540

541 The items (2), (3), (4) and (5) are deprecated and are retained for compatibility with the
542 POSIX standard. FDCC-sets should be specified in a coded character set independent way,
543 using symbolic names. To make actual use of the FDCC-set, it shall be used together with
544 charmaps and/or repertoiremaps, so that the symbolic character names can be resolved into
545 the actual character encoding used.
546

547 (2) A character can be represented by the character itself, in
548 which case the value of the character is application-defined.
549 Within a string, the double-quote character, the escape
550 character, and the right angle bracket character shall be
551 escaped (preceded by the escape character) to be interpreted
552 as the character itself. Outside strings, the characters
553

554 , ; < > escape_char
555

556 shall be escaped to be interpreted as the character itself.
557

558 Example: c ä "May"
559

560 (3) A character can be represented as an octal constant. An octal
561 constant shall be specified as the escape character followed
562 by two or more octal digits. Each constant shall represent a
563 byte value.
564

565 Example: \143; \347; "\115"
566

567 (4) A character can be represented as a hexadecimal constant. A
568 hexadecimal constant shall be specified as the escape
569 character followed by an x followed by two or more
570 hexadecimal digits. Each constant shall represent a byte

571   value.

572
573   Example: \x63;\xe7;

574
575   (5)   A character can be represented as a decimal constant. A
576   decimal constant shall be specified as the escape character
577   followed by a d followed by two or more decimal digits.
578   Each constant shall represent a byte value.

579
580   Example: \d99; \d231;

581
582   (6)   Multibyte characters can be represented by concatenated
583   constants specified in byte order with the last constant
584   specifying the least significant byte of the character.
585   Concatenated constants can include a mix of the above
586   character representations.

587
588   Example: \143\xe7; "\115\xe7\d171"

589
590   Only characters existing in the character set for which the FDCC-set definition is created
591   shall be specified, whether using symbolic names, the characters themselves, or octal,
592   decimal, or hexadecimal constants. If a charmap is present, only characters defined in the
593   charmap can be specified using octal, decimal, or hexadecimal constants. Symbolic names
594   not present in the charmap can be specified and shall be ignored, as specified under item
595   (1) above.

596
597   **4.1.2   Continuation of lines**

598
599   A line in a specification can be continued by placing an escape character as the last visible
600   graphic character on the line; this continuation character shall be discarded from the input.
601   The line is continued to the next non-comment line.

602
603   **4.1.3   Names for copy keyword**

604
605   In most of the categories a "copy" keyword is allowed. The name specified wth this copy
606   keyword shall be one of:

607
608   - "i18n" which indicate the "i18n" FDCC-set defined in this specification,
609   - the name of a FDCC-set or POSIX locale registered by the process defined in ISO/IEC
610   15897,
611   - any other name which may be recognized in some local context - not being
612   recommended as an international specification.

613
614   **4.1.4   Pre-category statements**

615
616   In a FDCC-set the following statements can precede category specifications, and they
617   apply to all categories in the specified FDCC-set.

618
619   **4.1.4.1   comment_char**

620
621   The following line in a FDCC-set modifies the comment character. It shall have the

622 following syntax, starting in column 1:

623

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

625

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

631

632 **4.1.4.2   escape_char**

633

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

636

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

638

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

642

643 **4.1.4.3   repertoiremap**

644

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

648

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

650

651 The name shall be one of:
652 - "i18nrep" which indicate the "i18nrep" repertoiremap defined in this specification,
653 - the name of a <repertoiremap> registered by the process defined in ISO/IEC 15897,
654 - any other name which may be recognized in some local context - not being
655   recommended as an international specification.

656

657 **4.1.4.4   charmap**

658

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

661

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

663

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

668

669 The name shall be one of:
670 - the name of a <charmap> registered by the process defined in ISO/IEC 15897,
671 - any other name which may be recognized in some local context - not being

672   recommended as an international specification.
673
674   **4.2   LC_IDENTIFICATION**
675
676   The LC_IDENTIFICATION category defines properties of the FDCC-set, and which
677   specification methods the FDCC-set is conforming to. All keywords are mandatory unless
678   otherwise noted, and the operands are strings. The following keywords shall be defined:
679

680	**title**	Title of the FDCC-set.
681	**source**	Organization name of provider of the source.
682	**address**	Organization postal address.
683	**contact**	Name of contact person. This keyword is optional.
684 685	**email**	Electronic mail address of the organization, or contact person.
686 687	**tel**	Telephone number for the organization, in international format.
688	**fax**	Fax number for the organization, in international format.
689 690	**language**	Natural language to which the FDCC-set applies, as specified in ISO 639.
691 692	**territory**	The geographic extent where the FDCC-set applies (need not be a national extent), as two-letter form of ISO 3166.
693 694	**audience**	If not for general use, an indication of the intended user audience. This keyword is optional.
695 696	**application**	If for use of a special application, a description of the application. This keyword is optional.
697 698	**abbreviation**	Short name for provider of the source. This keyword is optional.
699 700	**revision**	Revision number consisting of digits and zero or more full stops (".").
701 702	**date**	Revision date in the format according to this example: "1995-02-05" meaning the 5th of February, 1995.

703
704   If any of the above information is non-existent, it must be stated in each case; the
705   corresponding string is then the empty string. If required information is not present in ISO
706   639 or ISO 3166, the relevant Maintenance Authority should be approached to get the
707   needed item registered.
708
709   Note: Only one language can be addressed with the concepts of a FDCC-set; to address
710   for example a bilingual culture, one need to have 2 FDCC-sets.
711

712	**category**	Shall be used to define that a category is present and what
713		specification the category is claiming conformance to. The
714		first operand is a string in double-quotes that describes the
715		specification that the category is claiming conformance to,
716		and the following values shall be defined:
717		"i18n:1999"
718		"posix:1993"
719		The second operand is a string with the category name,
720		where the category names of clause 4 shall be defined. More
721		than one "category" keyword may be given, but only one per

722   category name.

723

724 The "i18n" LC_IDENTIFICATION category is:

725

```
726   LC_IDENTIFICATION
727   % This is the ISO/IEC TR 14652 "i18n" definition for
728   % the LC_IDENTIFICATION category.
729   %
730   title                 "ISO/IEC 14652 i18n FDCC-set"
731   source                "ISO/IEC Copyright Office"
732   address               "Case postale 56, CH-1211 Geneve 20, Switzerland"
733   contact               ""
734   email                 ""
735   tel                   ""
736   fax                   ""
737   language              ""
738   territory             "ISO"
739   revision              "1.0"
740   date                  "1999-06-28"
741   %
742   category  "i18n:1999";LC_IDENTIFICATION
743   category  "i18n:1999";LC_CTYPE
744   category  "i18n:1999";LC_COLLATE
745   category  "i18n:1999";LC_TIME
746   category  "i18n:1999";LC_NUMERIC
747   category  "i18n:1999";LC_MONETARY
748   category  "i18n:1999";LC_MESSAGES
749   category  "i18n:1999";LC_PAPER
750   category  "i18n:1999";LC_NAME
751   category  "i18n:1999";LC_ADDRESS
752   category  "i18n:1999";LC_TELEPHONE
753
754   END LC_IDENTIFICATION
```

755

756

4.3   LC_CTYPE

757

758

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

762

763 A series of characters in a specification can be represented by the hexadecimal symbolic
764 ellipsis symbol ".." (two dots), the decimal symbolic ellipses symbols "...." (4 dots), the
765 double increment hexadecimal symbolic ellipses "..(2)..", or the absolute ellipses "..." (3
766 dots).

767

768 The **hexadecimal symbolic ellipsis** ("..") specification is only valid between symbolic
769 character names. The symbolic names shall consist of zero or more nonnumeric characters
770 from the set shown with visible glyphs in Table 1, followed by an integer formed by one
771 or more hexadecimal digits, using uppercase letters only for the range "A" to "F". The
772 characters preceding the hexadecimal integer shall be identical in the two symbolic names,
773 and the integer formed by the hexadecimal digits in the second symbolic name shall be
774 identical to or greater than the integer formed by the hexadecimal digits in the first name.
775 This shall be interpreted as a series of symbolic names formed from the common part and
776 each of the integers in hexadecimal format using uppercase letters only between the first
777 and the second integer, inclusive, and with a length of the symbolic names generated that
778 is equal to the length of the first (and also the second) symbolic name. As an example,
779 <U010E>..<U0111> is interpreted as the symbolic names <U010E>, <U010F>, <U0110>,
780 and <U0111>, in that order.

781

782 The **decimal symbolic ellipsis** ("....") specification is only valid between symbolic

783   character names. The symbolic names shall consist of zero or more nonnumeric characters
784   from the set shown with visible glyphs in Table 1, followed by an integer formed by one
785   or more decimal digits. The characters preceding the decimal integer shall be identical in
786   the two symbolic names, and the integer formed by the decimal digits in the second
787   symbolic name shall be identical to or greater than the integer formed by the decimal
788   digits in the first name. This shall be interpreted as a series of symbolic names formed
789   from the common part and each of the integers in decimal format between the first and the
790   second integer, inclusive, and with a length of the symbolic names generated that is equal
791   to the length of the first (and also the second) symbolic name. As an example,
792   <j0101>....<j0104> is interpreted as the symbolic names <j0101>, <j0102>, <j0103>, and
793   <j0104>, in that order.

794

795   The **double increment hexadecimal symbolic ellipses** ("..(2)..") works like the
796   hexadecimal symbolic ellipses, but generates only every other of the symbolic character
797   names. As an example. <U01AC>..(2)..<U01B2> is interpreted as the symbolic character
798   names <U01AC>, <U01AE>, <U01B0>, and <U01B2>, in that order.

799

800   The **absolute ellipsis** specification is only valid within a single encoded character set. An
801   ellipsis shall be interpreted as including in the list all characters with an encoded value
802   higher than the encoded value of the character preceding the ellipsis and lower than the
803   encoded value of the character following the ellipsis. The absolute ellipsis specification is
804   deprecated, as this is only relevant to FDCC-sets not using symbolic characters.
805   As an example, \x30;...;\x39 includes in the character class all characters with encoded
806   values between the endpoints.

807

808   ### 4.3.1   Basic keywords

809

810   The following keywords shall be recognized. In the descriptions, the term "automatically
811   included" means that it shall not be an error to either include the referenced characters or
812   to omit them; the interpreting system shall provide them if missing and accept them
813   silently if present.

814

815	**copy**	Specify the name of an existing FDCC-set to be used as the source for the
816		definition of this category. If this keyword is specified, no other keyword
817		shall be specified.
818	**upper**	Define characters to be classified as uppercase letters. No character
819		specified for the keywords "cntrl", "digit", "punct", or "space" shall be
820		specified. The uppercase letters A through Z of the portable character set,
821		shall automatically belong to this class, with application-defined character
822		values. The keyword may be omitted.
823	**lower**	Define characters to be classified as lowercase letters. No character
824		specified for the keywords "cntrl", "digit", "punct", or "space" shall be
825		specified. The lowercase letters a through z of the portable character set,
826		shall automatically belong to this class, with application-defined character
827		values. The keyword may be omitted.
828	**alpha**	Define characters to be classified as used to spell out the words for natural
829		languages; such as letters, syllabic or ideographic characters. No character
830		specified for the keywords "cntrl", "digit", "punct", or "space" shall be
831		specified. In addition, characters classified as either "upper" or "lower" shall
832		automatically belong to this class. The keyword may be omitted.

833 834 835 836 837 838 839 840 841 842	**digit**	Define the characters to be classified as numeric 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 shall 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.
843 844 845 846 847 848 849 850 851 852	**outdigit**	Define the characters to be classified as numeric digits for output from an application, such as to a printer or a display or a output text file. 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 digits are used for output, including numeric and monetary formatting, and date and time formatting. Only one set of 10 digits may be specified. If this keyword is not specified, the digits 0 through 9 of the portable character set shall automatically belong to this class, with application-defined character values. The keyword may be omitted.
853 854 855	**blank**	Define characters to be classified as "blank" characters. If this keyword is unspecified, the characters <space> and <tab>, with application-defined character values, shall belong to this character class.
856 857 858 859 860 861 862 863 864 865	**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" shall be specified. If this keyword is not specified, the characters <space>, <form-feed>, <newline>, <carriage-return>, <tab>, and <vertical-tab>, shall automatically belong to this class, with application-defined character values. Any characters included in the class "blank" shall be 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.
866 867 868 869	**cntrl**	Define characters to be classified as control characters. No character specified for the keywords "upper", "lower", "alpha", "digit", "punct", "graph", "print", or "xdigit" shall be specified. The keyword shall be specified.
870 871 872 873	**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 shall be specified. The keyword shall be specified.
874 875 876 877 878 879 880 881 882	**xdigit**	Define the characters to be classified as hexadecimal digits. Only the characters defined for the class "digit" shall be specified, in ascending sequence by numerical value, followed by one or more sets of six characters representing the hexadecimal digits 10 through 15, with each set in ascending order (for example <A>, <B>, <C>, <D>, <E>, <F>, <a>, <b>, <c>, <d>, <e>, <f>). If this keyword is not specified, the digits <0> through <9>, the uppercase letters "A" through <F>, and the lowercase letters <a> through <f>, shall automatically belong to this class, with application-defined character values.

14

883 884 885 886 887	**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" shall belong to this character class. No character specified for the keyword "cntrl" shall be specified.
888 889 890 891 892	**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 shall belong to this character class. No character specified for the keyword "cntrl" shall be specified.
893 894 895 896 897 898 899 900 901	**toupper**	Define the mapping of lowercase letters to uppercase letters. The operand shall consist of character pairs, separated by semicolons. The characters in each character pair shall be separated by a comma and the pair enclosed by parentheses. The first character in each pair shall be the lowercase letter, the second the corresponding uppercase letter. Only characters specified for the keywords "lower" and "upper" shall be specified. If this keyword is not specified, the lowercase letters <a> through <z>, and their corresponding uppercase letters <A> through <Z>, shall automatically be included, with application-defined character values.
902 903 904 905 906 907 908 909 910	**tolower**	Define the mapping of uppercase letters to lowercase letters. The operand shall 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 shall be the uppercase letter, the second the corresponding lowercase letter. Only characters specified for the keywords "lower" and "upper" shall be specified. If this keyword is specified, the uppercase letters <A> through <Z>, and their corresponding lowercase letter, shall be specified. If this keyword is not specified, the mapping shall be the reverse mapping of the one specified for toupper.
911 912 913 914 915 916 917	**class**	Define characters to be classified in the class with the name given in the first operand, which is a string. This string shall only contain 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 shall be recognized:

918 919	**combining**	Characters to form composite graphic symbols, such as characters listed in ISO/IEC 10646:1993 annex B.1.
920 921 922	**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.

923 924 925		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.
926 927 928 929 930 931 932	**map**	Define the mapping of characters. The first operand is a string, defining the name of the mapping. The string shall only contain letters, digits and <hyphen-minus> and <low-line> from the portable character set. The following operands shall consist of character pairs, separated by semicolons. The characters in each character pair shall be separated by a comma and the pair enclosed by parentheses. The first character in each pair shall be the character to map from, the second the corresponding character to map to.

933 This keyword is optional. The keyword can only be specified once per
934 named mapping.
935
936 The mapping names "toupper", and "tolower" are taken to mean the
937 mapping defined by the respective keywords.
938
939 Example of use of the "map" keyword:
940
941 map "kana",(<U30AB>,<U304B>);(<U30AC>,<U304C>);(<U30AD>,<U304D>)
942
943 This example introduces a new mapping "kana" that maps three Katakana characters to corresponding Hiragana
944 characters.
945
946 Table 2 shows the allowed character class combinations.
947
948
949 **Table 2: Valid Character Class Combinations**
950
951

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	

964
965 NOTES:
966 Note 1: Explanation of codes:
967 A Automatically included; see text
968 + Permitted
969 x Mutually exclusive
970 * See note 2
971
972 Note 2: The <space> character, which is part of the "space" and "blank" class, cannot
973 belong to "punct" or "graph", but automatically shall belong to the "print" class. Other
974 "space" or "blank" characters can be classified as "punct", "graph", and/or "print".
975
976 **4.3.2 Character string transliteration**
977
978 The following keywords may be used to transliterate strings, by transforming substrings in
979 the source to substrings in the target string. The capabilities are limited to simple
980 transliteration based on substring substitution, while more advanced transliteration
981 schemes, for example based on pattern matching, is either cumbersome to specify, or not
982 addressed. The transliteration may for example be from the Cyrillic script to the Latin
983 script.
984

985 Transliteration is often language dependent, transliterating one specific language to another
986 specific language. For example transliteration from Russian to English, and from Serbian
987 to German would normaly be quite different, although the same repertoire of characters
988 would be transliterated. Even transliteration of two languages using the same script into
989 one language (for example from Russian to Danish and from Serbian to Danish), or
990 transliteration of the same language (for example Russian into English or German) may be
991 different. The language to be transliterated to is identified with the FDCC-set, which may
992 also be used to identify a specific language to be transliterated from. Transliteration may
993 also be to a specific repertoire of characters, determined for example by limitations of
994 displaying equipment, or what the user can intelligibly read. The capabilities here allows
995 for multiple fallback, so that the specification can be valid for all target character
996 repertoires, eliminating the need for specific data for each target repertoire. Transliteration
997 of an incoming character string to a character string in a FDCC-set can be specified with
998 the following keywords and transliteration statements.
999

1000 **translit_start**   The "translit_start" keyword is followed by one or more
1001   transliteration statements assigning character transliteration
1002   values to transliterating elements, and include statements
1003   copying transliteration specifications from other FDCC-sets.
1004 **translit_end**   The end of the transliteration statements.
1005 **include**   The name of the FDCC-set in text form to transliterate from,
1006   and the repertoiremap for the FDCC-set to be used for the
1007   definition of the transliteration statements. Other transliteration
1008   statements may follow to replace specification of the copied
1009   FDCC-set. This keyword is optional.
1010 **default_missing**   defines a string of one or more characters to be used if no
1011   transliteration statement can be applied to a input
1012   <transliteration-source>.
1013 **translit_ignore**   defines a set of characters, separated by semicolons, that are
1014   to be ignored in the incoming character string. The characters
1015   may use the notations defined in 4.3 for lists of characters.
1016 **redefine**   This keyword introduces a list of transliteration statements
1017   where each of the <transliteration_source> strings have been
1018   defined previously in the specification, and the new
1019   transliteration statements then replaces the old transliteration
1020   statements for the <transliteration_source> strings specified.
1021

1022 **4.3.2.1   Transliteration statements**
1023

1024 The "translit_start" keyword may be followed by transliteration statements. The syntax for
1025 a transliteration statement is:
1026

1027   "%s %s;%s;...;%s\n",<transliteration_source>,<transliteration_string>,...
1028

1029 Each <transliteration_source> shall consist of one or more characters (in any of the forms
1030 defined in 4.1.1). The <transliteration_source> that is the longest in terms of number of
1031 characters that match the input string is the one selected for transliteration.
1032

1033 If a transliteration statement contains more than one <transliteration_string>, the order that
1034 each <transliteration_string> occurs in the transliteration statement defines the precedence

1035 order for choosing a particular <transliteration_string> to substitute for the
1036 <transliteration_source>. When a process makes use of a transliteration statement to
1037 transliterate text, and that transliteration statement contains more than one
1038 <transliteration_string>, that process shall choose the first <transliteration_string>, in the
1039 defined precedence order, that satisfies the requirements of the transliteration.
1040
1041 Note: the exact definition of the concept of satisfying the requirements of the
1042 transliteration is outside the context of this Technical Report. If, for example, a
1043 transliteration involves a change in the coded character set of a string, a
1044 <transliteration_string> must be chosen, all of whose elements are members of that
1045 coded character set. In order to determine this, it would be expected that a
1046 repertoire describing which characters are to be present in the resulting transformed
1047 string be available to the transliteration API. Also, a transliteration may involve
1048 requirements such as that string length not change under transliteration. Such
1049 requirements may also affect the choice among alternative <transliteration_string>
1050 values.
1051
1052 If more than one transliteration statement is given for a given <transliteration_source> this
1053 is an error, and duplicate transliteration statements are ignored. Tailoring of transliteration
1054 statements may be done via the "redefine" keyword.
1055
1056 ### 4.3.2.2   "include" keyword
1057
1058 The "include" keyword specifies a set of transliteration statements in text form to be
1059 included in the applied transliteration.
1060
1061 The syntax of the "include" statement is:
1062
1063 "include %s;%s\n", <FDCC-set>, <repertoiremap>
1064
1065 <FDCC-set> is a string identifying the FDCC-set to be included from.
1066
1067 <repertoiremap> is a string identifying the repertoiremap used in the FDCC-set being
1068 included, and is used to map character specifications from the specified FDCC-set into the
1069 current FDCC-set.
1070
1071 #### 4.3.2.3   Example of use of transliteration
1072
1073 ```
1074 translit_start
1075 include "de_DE";"de_repmap"
1076 default_missing <?>
1077 translit_ignore <U3200>..<UFAFF>
1078 <ae>   <a:>;<e*>;"<a><e>";"<e>"
1079 <s>   <s*>;<s=>
1080 "<K><O>"   <KO>
1081 translit_end
1082 ```

1082 The "translit_start" keyword introduces the transliteration section in the LC_CTYPE category.

1084 The "include" keyword specifies that the FDCC-set "de_DE" is copied and that the repertoiremap "de_repmap" is
1085 used to define the symbolic character names in the FDCC-set "de_DE".

1087 The "default_missing" keyword introduces the character sequence "<?>" as the string to transform into for input
1088 characters that cannot be transformed into other strings, because no transliteration statement is applicable to the
1089 character.

1091 The "translit_ignore" keyword specifies that a set of Ideographic characters (the range <U3200>..<UFAFF>) shall

1092   be ignored for the transliteration.

1094   The next 3 lines are transliteration statements.

1096   The first transliteration statement defines a number of transliterations for the LATIN LETTER AE, including into
1097   LATIN LETTER A WITH DIAERESIS, GREEK LETTER EPSILON, the two Latin letters A and E, and finally
1098   the LATIN LETTER E.

1100   The second transliteration statement defines transliteration of the LATIN LETTER S into GREEK LETTER
1101   SIGMA, and CYRILLIC LETTER ES.

1103   The third transliteration statement transliterates the two Latin letters K and O into the Japanese Hiragana character
1104   KO.

1106   The transliteration sections is terminated via the "translit_end" keyword in the above example.

4.3.3   "i18n" LC_CTYPE category

1110   The "i18n" FDCC-set for the LC_CTYPE is defined as follows:

```
LC_CTYPE
% The following is the 14652 i18n fdcc-set LC_CTYPE category.
% It covers ISO/IEC 10646-1 including Cor.1 and AMD 1 thru 9
% The "upper" class reflects the uppercase characters of class "alpha"
upper /
% TABLE 1 BASIC LATIN
   <U0041>..<U005A>;/
% TABLE 2 LATIN-1 SUPPLEMENT
   <U00C0>..<U00D6>;<U00D8>..<U00DE>;/
% TABLE 3 LATIN EXTENDED-A
   <U0100>..(2)..<U0136>;/
   <U0139>..(2)..<U0147>;/
   <U014A>..(2)..<U0178>;/
   <U0179>..(2)..<U017D>;/
% TABLE 4 LATIN EXTENDED-B
   <U0181>;<U0182>..(2)..<U0186>;<U0187>;/
   <U0189>..<U018B>;<U018E>..<U0191>;<U0193>;<U0194>;/
   <U0196>..<U0198>;<U019C>;<U019D>;<U019F>;/
   <U01A0>..(2)..<U01A4>;/
   <U01A7>;<U01A9>;<U01AC>;<U01AE>;<U01AF>;<U01B1>..<U01B3>;/
   <U01B5>;<U01B7>;<U01B8>;<U01BC>;<U01C4>;<U01C5>;<U01C7>;<U01C8>;/
   <U01CA>;<U01CB>;/
   <U01CD>..(2)..<U01DB>;/
   <U01DE>..(2)..<U01EE>;/
   <U01F1>;<U01F2>;<U01F4>;<U01FA>..(2)..<U01FE>/
% TABLE 5 LATIN EXTENDED-B
   <U0200>..(2)..<U0216>;/
% TABLE 6 IPA EXTENSIONS
   <U0262>;<U026A>;<U0274>;<U0276>;/
   <U0280>;<U0281>;<U028F>;<U0299>;<U029B>;<U029C>;<U029F>;/
% TABLE 9 BASIC GREEK
   <U0386>;<U0388>..<U038A>;<U038C>;<U038E>;<U038F>;<U0391>..<U03A1>;/
   <U03A3>..<U03AB>;/
% TABLE 10 GREEK SYMBOLS AND COPTIC
   <U03E3>..(2)..<U03EF>;/
% TABLE 11 CYRILLIC
   <U0401>..<U040C>;<U040E>..<U042F>;<U0460>..(2)..<U047E>;/
% TABLE 12 CYRILLIC
   <U0480>;<U0490>..(2)..<U04BE>;<U04C1>;<U04C3>;<U04C7>;<U04CB>;/
   <U04D0>..(2)..<U04EA>;<U04EE>..(2)..<U04F4>;<U04F8>;/
% TABLE 13 ARMENIAN
   <U0531>..<U0556>;/
% TABLE 28 GEORGIAN
   <U10A0>..<U10C5>;/
% TABLE 31 LATIN EXTENDED ADDITIONAL
   <U1E00>..(2)..<U1E7E>;/
% TABLE 32 LATIN EXTENDED ADDITIONAL
   <U1E80>..(2)..<U1E94>;/
   <U1EA0>..(2)..<U1EF8>;/
% TABLE 33 GREEK EXTENDED
   <U1F08>..<U1F0F>;<U1F18>..<U1F1D>;<U1F28>..<U1F2F>;<U1F38>..<U1F3F>;/
   <U1F48>..<U1F4D>;<U1F59>..(2)..<U1F5F>;<U1F68>..<U1F6F>;/
```

```
1164        % TABLE 34 GREEK EXTENDED
1165           <U1F88>..<U1F8F>;<U1F98>..<U1F9F>;<U1FA8>..<U1FAF>;<U1FB8>..<U1FBC>;/
1166           <U1FC8>..<U1FCC>;<U1FD8>..<U1FDB>;<U1FE8>..<U1FEC>;<U1FF8>..<U1FFC>
1167        % TABLE 28 GEORGIAN is not addressed as the letters does not have
1168        %          a uppercase/lowercase relation
1169        %
1170        % The "lower" class reflects the lowercase characters of class "alpha"
1171        lower /
1172        % TABLE 1 BASIC LATIN
1173           <U0061>..<U007A>;/
1174        % TABLE 2 LATIN-1 SUPPLEMENT
1175           <U00DF>..<U00F6>;<U00F8>..<U00FF>;/
1176        % TABLE 3 LATIN EXTENDED-A
1177           <U0101>..(2)..<U0137>;<U0138>..(2)..<U0148>;/
1178          <U0149>..(2)..<U0177>;<U017A>..(2)..<U017E>;<U017F>;/
1179        % TABLE 4 LATIN EXTENDED-B
1180           <U0180>;<U0183>;<U0185>;<U0188>;<U018C>;<U018D>;<U0192>;<U0195>;/
1181           <U0199>..<U019B>;<U019E>;<U01A1>;<U01A3>;<U01A5>;<U01A8>;<U01AB>;<U01AD>;/
1182           <U01B0>;<U01B4>;<U01B6>;<U01B9>;<U01BA>;<U01BD>;<U01C5>;<U01C6>;/
1183           <U01C8>;<U01C9>;<U01CB>;<U01CC>..(2)..<U01DC>;/
1184           <U01DD>;..(2)..<U01F2>;<U01F3>;<U01F5>;<U01FB>;<U01FD>;<U01FF>;/
1185        % TABLE 5 LATIN EXTENDED-B
1186           <U0201>..(2)..<U0217>;/
1187        % TABLE 6 IPA EXTENSIONS
1188           <U0250>..<U0293>;<U0299>..<U02A0>;<U02A3>..<U02A8>;/
1189        % TABLE 9 BASIC GREEK
1190           <U0390>;<U03AC>..<U03CE>;/
1191        % TABLE 10 GREEK SYMBOLS AND COPTIC
1192           <U03E2>..(2)..<U03EE>/
1193        % TABLE 11 CYRILLIC
1194           <U0430>..<U044F>;<U0451>..<U045C>;<U045E>;<U045F>;<U460>..(2)..<U047F>;/
1195        % TABLE 12 CYRILLIC
1196           <U04801>;<U0490>..(2)..<U04BF>;<U04C2>;<U04C4>;<U04C8>;<U04CC>;/
1197           <U04D1>..(2)..<U04EB>;<U04EF>..(2)..<U04F5>;<U04F9>;/
1198        % TABLE 13 ARMENIAN
1199           <U0561>..<U0587>;/
1200        % TABLE 28 GEORGIAN
1201           <U10D0>..<U10F6>;/
1202        % TABLE 31 and 32 LATIN EXTENDED ADDITIONAL
1203           <U1E01>..(2)..<U1E95>;<U1EA1>..(2)..<U1EF9>;/
1204        % TABLE 33 and 34 GREEK EXTENDED
1205           <U1F08>..<U1F0F>;<U1F18>..<U1F1D>;<U1F28>..<U1F2F>;<U1F38>..<U1F3F>;/
1206           <U1F48>..<U1F4D>;<U1F59>..(2)..<U1F5F>;<U1F68>..<U1F6F>;/
1207        % TABLE 34 GREEK EXTENDED
1208           <U1F00>..<U1F07>;<U1F10>..<U1F15>;<U1F20>..<U1F27>;<U1F30>..<U1F37>;/
1209           <U1F40>..<U1F45>;<U1F50>..<U1F57>;<U1F60>..<U1F67>;<U1F70>..<U1F7D>;/
1210           <U1F80>..<U1F87>;<U1F90>..<U1F97>;<U1FA0>..<U1FA7>;<U1FB0>..<U1FB4>;/
1211           <U1FB6>;<U1FB7>;<U1FC2>..<U1FC4>;<U1FC6>;<U1FC7>;<U1FD0>..<U1FD3>;/
1212           <U1FD6>;<U1FD7>;<U1FE0>..<U1FE7>;<U1FF2>..<U1FF4>;<U1FF6>;<U1FF7>;
1213        % TABLE 35 SUPERSCRIPTS AND SUBSCRIPTS, CURRENCY SYMBOLS
1214           <U207F>
1215        %
1216        % The "alpha" class of the "i18n" FDCC-set is reflecting
1217        % the recommendations in TR 10176 annex A
1218        alpha /
1219        % TABLE 1 BASIC LATIN
1220           <U0041>..<U005A>;<U0061>..<U007A>;/
1221        % TABLE 2 LATIN-1 SUPPLEMENT
1222           <U00AA>;<U00BA>;<U00C0>..<U00D6>;<U00D8>..<U00F6>;<U00F8>..<U00FF>;/
1223        % TABLE 3 LATIN EXTENDED-A
1224           <U0100>..<U017F>;/
1225        % TABLE 4 and 5 LATIN EXTENDED-B
1226           <U0180>..<U01F5>;<U01FA>..<U0217>;/
1227        % TABLE 6 IPA EXTENSIONS
1228           <U0250>..<U02A8>;/
1229        % TABLE 31 and 32 LATIN EXTENDED ADDITIONAL
1230           <U1E00>..<U1E9B>;<U1EA0>..<U1EF9>;/
1231        % TABLE 35 SUPERSCRIPTS AND SUBSCRIPTS, CURRENCY SYMBOLS
1232           <U207F>;/
1233        % TABLE 9 BASIC GREEK
1234           <U0386>;<U0388>..<U038A>;<U038C>;<U038E>..<U03A1>;<U03A3>..<U03CE>;/
1235        % TABLE 10 GREEK SYMBOLS AND COPTIC
1236           <U03D0>..<U03D6>;<U03DA>;<U03DC>;<U03DE>;<U03E0>;<U03E2>..<U03F3>;/
1237        % TABLE 33 and 34 GREEK EXTENDED
1238           <U1F00>..<U1F15>;<U1F18>..<U1F1D>;<U1F20>..<U1F45>;<U1F48>..<U1F4D>;/
1239           <U1F50>..<U1F57>;<U1F59>;<U1F5B>;<U1F5D>;<U1F5F>..<U1F7D>;/
1240           <U1F80>..<U1FB4>;<U1FB6>..<U1FBC>;<U1FC2>..<U1FC4>;<U1FC6>..<U1FCC>;/
1241           <U1FD0>..<U1FD3>;<U1FD6>..<U1FDB>;<U1FE0>..<U1FEC>;<U1FF2>..<U1FF4>;/
1242           <U1FF6>..<U1FFC>;/
```

```
1243        % TABLE 11 and 12 CYRILLIC
1244          <U0401>..<U040C>;<U040E>..<U044F>;<U0451>..<U045C>;<U045E>..<U0481>;/
1245          <U0490>..<U04C4>;<U04C7>..<U04C8>;<U04CB>..<U04CC>;<U04D0>..<U04EB>;/
1246          <U04EE>..<U04F5>;<U04F8>..<U04F9>;/
1247        % TABLE 13 ARMENIAN
1248          <U0531>..<U0556>;<U0561>..<U0587>;/
1249        % TABLE 14 HEBREW
1250          <U05B0>..<U05B9>;<U05BB>..<U05BD>;<U05BF>;<U05C1>..<U05C2>;/
1251          <U05D0>..<U05EA>;<U05F0>..<U05F2>;/
1252        % TABLE 15 and 16 ARABIC
1253          <U0621>..<U063A>;<U0640>..<U0652>;<U0670>..<U06B7>;<U06BA>..<U06BE>;/
1254          <U06C0>..<U06CE>;<U06D0>..<U06D3>;<U06D5>..<U06DC>;<U06E5>..<U06E8>;/
1255          <U06EA>..<U06ED>;/
1256        % TABLE 17 DEVANAGARI
1257          <U0901>..<U0903>;<U0905>..<U0939>;<U093E>..<U094D>;<U0950>..<U0952>;/
1258          <U0958>..<U0963>;/
1259        % TABLE 18 BENGALI
1260          <U0981>..<U0983>;<U0985>..<U098C>;<U098F>..<U0990>;/
1261          <U0993>..<U09A8>;<U09AA>..<U09B0>;<U09B2>;<U09B6>..<U09B9>;/
1262          <U09BE>..<U09C4>;<U09C7>..<U09C8>;<U09CB>..<U09CD>;<U09DC>..<U09DD>;/
1263          <U09DF>..<U09E3>;<U09F0>..<U09F1>;/
1264        % TABLE 19 GURMUKHI
1265          <U0A02>;<U0A05>..<U0A0A>;<U0A0F>..<U0A10>;<U0A13>..<U0A28>;/
1266          <U0A2A>..<U0A30>;<U0A32>..<U0A33>;<U0A35>..<U0A36>;<U0A38>..<U0A39>;/
1267          <U0A3E>..<U0A42>;<U0A47>..<U0A48>;<U0A4B>..<U0A4D>;<U0A59>..<U0A5C>;/
1268          <U0A5E>;<U0A74>;/
1269        % TABLE 20 GUJARATI
1270          <U0A81>..<U0A83>;<U0A85>..<U0A8B>;<U0A8D>;<U0A8F>..<U0A91>;/
1271          <U0A93>..<U0AA8>;<U0AAA>..<U0AB0>;<U0AB2>..<U0AB3>;<U0AB5>..<U0AB9>;/
1272          <U0ABD>..<U0AC5>;<U0AC7>..<U0AC9>;<U0ACB>..<U0ACD>;<U0AD0>;<U0AE0>;/
1273        % TABLE 21 ORIYA
1274          <U0B01>..<U0B03>;<U0B05>..<U0B0C>;<U0B0F>..<U0B10>;<U0B13>..<U0B28>;/
1275          <U0B2A>..<U0B30>;<U0B32>..<U0B33>;<U0B36>..<U0B39>;<U0B3E>..<U0B43>;/
1276          <U0B47>..<U0B48>;<U0B4B>..<U0B4D>;<U0B5C>..<U0B5D>;<U0B5F>..<U0B61>;/
1277        % TABLE 22 TAMIL
1278          <U0B82>..<U0B83>;<U0B85>..<U0B8A>;<U0B8E>..<U0B90>;<U0B92>..<U0B95>;/
1279          <U0B99>..<U0B9A>;<U0B9C>;<U0B9E>..<U0B9F>;<U0BA3>..<U0BA4>;/
1280          <U0BA8>..<U0BAA>;<U0BAE>..<U0BB5>;<U0BB7>..<U0BB9>;<U0BBE>..<U0BC2>;/
1281          <U0BC6>..<U0BC8>;<U0BCA>..<U0BCD>;/
1282        % TABLE 23 TELUGU
1283          <U0C01>..<U0C03>;<U0C05>..<U0C0C>;<U0C0E>..<U0C10>;<U0C12>..<U0C28>;/
1284          <U0C2A>..<U0C33>;<U0C35>..<U0C39>;<U0C3E>..<U0C44>;<U0C46>..<U0C48>;/
1285          <U0C4A>..<U0C4D>;<U0C60>..<U0C61>;/
1286        % TABLE 24 KANNADA
1287          <U0C82>..<U0C83>;<U0C85>..<U0C8C>;<U0C8E>..<U0C90>;<U0C92>..<U0CA8>;/
1288          <U0CAA>..<U0CB3>;<U0CB5>..<U0CB9>;<U0CBE>..<U0CC4>;<U0CC6>..<U0CC8>;/
1289          <U0CCA>..<U0CCD>;<U0CDE>;<U0CE0>..<U0CE1>;/
1290        % TABLE 25 MALAYALAM
1291          <U0D02>..<U0D03>;<U0D05>..<U0D0C>;<U0D0E>..<U0D10>;<U0D12>..<U0D28>;/
1292          <U0D2A>..<U0D39>;<U0D3E>..<U0D43>;<U0D46>..<U0D48>;<U0D4A>..<U0D4D>;/
1293          <U0D60>..<U0D61>;/
1294        % TABLE 26 THAI
1295          <U0E01>..<U0E3A>;<U0E40>..<U0E4E>;<U0E50>..<U0E59>;/
1296        % TABLE 27 LAO
1297          <U0E81>..<U0E82>;<U0E84>;<U0E87>..<U0E88>;<U0E8A>;<U0E8D>;/
1298          <U0E94>..<U0E97>;<U0E99>..<U0E9F>;<U0EA1>..<U0EA3>;<U0EA5>;<U0EA7>;/
1299          <U0EAA>..<U0EAB>;<U0EAD>..<U0EAE>;<U0EB0>..<U0EB9>;<U0EBB>..<U0EBD>;/
1300          <U0EC0>..<U0EC4>;<U0EC6>;<U0EC8>..<U0ECD>;<U0EDC>..<U0EDD>;/
1301        % TIBETAN Amendment 6
1302          <U0F00>;<U0F18>..<U0F19>;<U0F35>;<U0F37>;<U0F39>;<U0F40>..<U0F47>;/
1303          <U0F49>..<U0F69>;/
1304          <U0F71>..<U0F84>;<U0F86>..<U0F8B>;<U0F90>..<U0F95>;<U0F97>;/
1305          <U0F99>..<U0FAD>;<U0FB1>..<U0FB7>;<U0FB9>;/
1306        % TABLE 28 GEORGIAN
1307          <U10A0>..<U10C5>;<U10D0>..<U10F6>;/
1308        % TABLE 50 HIRAGANA
1309          <U3041>..<U3093>;<U309B>..<U309C>;/
1310        % TABLE 51 KATAKANA
1311          <U30A1>..<U30F6>;<U30FB>..<U30FC>;/
1312        % TABLE 52 BOPOMOFO
1313          <U3105>..<U312C>;/
1314        % CJK unified ideographs
1315          <U4E01>..<U9FA5>;/
1316        % HANGUL amendment 5
1317          <UAC00>..<UD7A3>;/
1318        % Miscellaneous
1319          <U00B5>;<U00B7>;<U02B0>..<U02B8>;<U02BB>;<U02BD>..<U02C1>;/
1320          <U02D0>..<U02D1>;<U02E0>..<U02E4>;<U037A>;<U0559>;<U093D>;<U0B3D>;/
```

```
1321        <U1FBE>;<U203F>..<U2040>;<U2102>;<U2107>;<U210A>..<U2113>;<U2115>;/
1322        <U2118>..<U211D>;<U2124>;<U2126>;<U2128>;<U212A>..<U2131>;/
1323        <U2133>..<U2138>;<U2160>..<U2182>;<U3005>..<U3006>;<U3021>..<U3029>
1324     %
1325     % The "digit" class of the "i18n" FDCC-set is reflecting
1326     % the recommendations in TR 10176 annex A
1327     digit /
1328     % TABLE 1 BASIC LATIN
1329        <U0030>..<U0039>;/
1330     % TABLE 15 and 16 ARABIC
1331        <U0660>..<U0669>;<U06F0>..<U06F9>;/
1332     % TABLE 17 DEVANAGARI
1333        <U0966>..<U096F>;/
1334     % TABLE 18 BENGALI
1335        <U09E6>..<U09EF>;/
1336     % TABLE 19 GURMUKHI
1337        <U0A66>..<U0A6F>;/
1338     % TABLE 20 GUJARATI
1339        <U0AE6>..<U0AEF>;/
1340     % TABLE 21 ORIYA
1341        <U0B66>..<U0B6F>;/
1342     % TABLE 22 TAMIL
1343        <0>;<U0BE7>..<U0BEF>;/
1344     % TABLE 23 TELUGU
1345        <U0C66>..<U0C6F>;/
1346     % TABLE 24 KANNADA
1347        <U0CE6>..<U0CEF>;/
1348     % TABLE 25 MALAYALAM
1349        <U0D66>..<U0D6F>;/
1350     % TABLE 26 THAI
1351        <U0E50>..<U0E59>;/
1352     % TABLE 27 LAO
1353        <U0ED0>..<U0ED9>;/
1354     % TIBETAN Amendment 6
1355        <U0F20>..<U0F29>
1356     %
1357     outdigit <U0030>..<U0039>
1358     %
1359     space /
1360     % ISO/IEC 6429
1361        <U0008>;<U000A>..<U000D>;
1362     % TABLE 1 BASIC LATIN
1363        <U0020>;/
1364     % TABLE 35 GENERAL PUNCTUATION
1365        <U2000>..<U2006>;<U2008>..<U200B>;/
1366     % TABLE 50 CJK SYMBOLS AND PUNCTUATION, HIRAGANA
1367        <U3000>
1368     %
1369     cntrl   <U0000>..<U001F>;<U0077>..<U009F>
1370     %
1371     punct /
1372        <U0021>..<U002F>;<U003A>..<U0040>;<U005B>..<U0060>;<U007B>..<U007E>;/
1373        <U00A0>..<U00A9>;<U00AB>..<U00B9>;<U00BB>..<U00BF>;<U00D7>;<U00F7>;/
1374        <U037E>;<U0482>;<U055A>..<U055F>;<U0589>;<U05BE>;<U05C0>;<U05C3>;/
1375        <U05F3>;<U05F4>;<U060C>;<U061B>;<U061F>;<U0640>;<U064B>..<U0652>;/
1376        <U066A>..<U066D>;<U06D4>;<U06DD>..<U06E1>;<U06E9>..<U06EC>;<U10FB>;/
1377        <U2010>..<U2029>;<U2030>..<U2046>;<U20A0>..<U20AA>;<U2100>..<U210B>;/
1378        <U210D>..<U2110>;<U2112>..<U211B>;<U211D>..<U2127>;<U212A>..<U212C>;/
1379        <U212E>..<U2138>;<U2200>..<U22F1>;<U2300>;<U2302>..<U237A>;<U2400>..<U2424>;/
1380        <U2440>..<U244A>;<U2580>..<U2595>;<U25A0>..<U25EF>;<U2600>..<U2613>;/
1381        <U261A>..<U266F>;<U2701>..<U2704>;<U2706>..<U2709>;<U270C>..<U2727>;/
1382        <U2729>..<U274B>;<U274D>;<U274F>..<U2752>;<U2756>;<U2758>..<U275E>;/
1383        <U2761>..<U2767>;<U3000>..<U3020>;<U3030>;<U3036>;<U3037>;<U303F>;<U3164>;/
1384        <U3190>..<U319F>;<U3200>..<U321C>;<U3220>..<U3243>;<U3260>..<U327B>;/
1385        <U327F>..<U32B0>;<U32C0>..<U32CB>;<U32D0>..<U32FE>;<U3300>..<U3376>;/
1386        <U337B>..<U33DD>;<U33E0>..<U33FE>;<UFD3E>;<UFD3F>;<UFE49>..<UFE52>;/
1387        <UFE54>..<UFE66>;<UFE68>..<UFE6B>;<UFEFF>;<UFF01>..<UFF0F>;<UFF1A>..<UFF20>;/
1388        <UFF3B>..<UFF40>;<UFF5B>..<UFF5E>;<UFF61>..<UFF65>;<UFF70>;<UFF9E>..<UFFA0>;/
1389        <UFFE0>..<UFFE6>;<UFFE8>..<UFFEE>;<UFFFD>
1390     %
1391     graph /
1392        <U0021>..<U007E>;<U00A0>..<U01F5>;<U01FA>..<U0217>;/
1393        <U0250>..<U02A8>;<U02B0>..<U02DE>;<U02E0>..<U02E9>;<U0300>..<U0345>;/
1394        <U0360>;<U0361>;<U0374>;<U0375>;<U037A>;<U037E>;<U0384>..<U038A>;<U038C>;/
1395        <U038E>..<U03A1>;<U03A3>..<U03CE>;<U03D0>..<U03D6>;<U03DA>;<U03DC>;<U03DE>;/
1396        <U03E0>;<U03E2>..<U03F3>;<U0401>..<U040C>;<U040E>..<U044F>;/
1397        <U0451>..<U045C>;<U045E>..<U0486>;<U0490>..<U04C4>;<U04C7>;<U04C8>;/
1398        <U04CB>;<U04CC>;<U04D0>..<U04EB>;<U04EE>..<U04F5>;<U04F8>;<U04F9>;/
1399        <U0531>..<U0556>;<U0559>..<U055F>;<U0561>..<U0587>;<U0589>;/
```

22

```
1400        <U0591>..<U05A1>;<U05A3>..<U05AF>;<U05B0>..<U05B9>;/
1401        <U05BB>..<U05C4>;<U05D0>..<U05EA>;<U05F0>..<U05F4>;<U060C>;<U061B>;<U061F>;/
1402        <U0621>..<U063A>;<U0640>..<U0652>;<U0660>..<U066D>;<U0670>..<U06B7>;/
1403        <U06BA>..<U06BE>;<U06C0>..<U06CE>;<U06D0>..<U06ED>;<U06F0>..<U06F9>;/
1404        <U0901>..<U0903>;<U0905>..<U0939>;<U093C>..<U094D>;<U0950>..<U0954>;/
1405        <U0958>..<U0970>;<U0981>..<U0983>;<U0985>..<U098C>;<U098F>;<U0990>;/
1406        <U0993>..<U09A8>;<U09AA>..<U09B0>;<U09B2>;<U09B6>..<U09B9>;<U09BC>;/
1407        <U09BE>..<U09C4>;<U09C7>;<U09C8>;<U09CB>..<U09CD>;<U09D7>;<U09DC>;<U09DD>;/
1408        <U09DF>..<U09E3>;<U09E6>..<U09FA>;<U0A02>;<U0A05>..<U0A0A>;<U0A0F>;<U0A10>;/
1409        <U0A13>..<U0A28>;<U0A2A>..<U0A30>;<U0A32>;<U0A33>;<U0A35>;<U0A36>;/
1410        <U0A38>;<U0A39>;<U0A3C>;<U0A3E>..<U0A42>;<U0A47>;<U0A48>;<U0A4B>..<U0A4D>;/
1411        <U0A59>..<U0A5C>;<U0A5E>;<U0A66>..<U0A74>;<U0A81>..<U0A83>;<U0A85>..<U0A8B>;/
1412        <U0A8D>;<U0A8F>..<U0A91>;<U0A93>..<U0AA8>;<U0AAA>..<U0AB0>;/
1413        <U0AB2>;<U0AB3>;<U0AB5>..<U0AB9>;<U0ABC>..<U0AC5>;<U0AC7>..<U0AC9>;/
1414        <U0ACB>..<U0ACD>;<U0AD0>;<U0AE0>;<U0AE6>..<U0AEF>;<U0B01>..<U0B03>;/
1415        <U0B05>..<U0B0C>;<U0B0F>;<U0B10>;<U0B13>..<U0B28>;<U0B2A>..<U0B30>;/
1416        <U0B32>;<U0B33>;<U0B36>..<U0B39>;<U0B3C>..<U0B43>;<U0B47>;<U0B48>;/
1417        <U0B4B>..<U0B4D>;<U0B56>;<U0B57>;<U0B5C>;<U0B5D>;<U0B5F>..<U0B61>;/
1418        <U0B66>..<U0B70>;<U0B82>;<U0B83>;<U0B85>..<U0B8A>;<U0B8E>..<U0B90>;/
1419        <U0B92>..<U0B95>;<U0B99>;<U0B9A>;<U0B9C>;<U0B9E>;<U0B9F>;<U0BA3>;<U0BA4>;/
1420        <U0BA8>..<U0BAA>;<U0BAE>..<U0BB5>;<U0BB7>..<U0BB9>;<U0BBE>..<U0BC2>;/
1421        <U0BC6>..<U0BC8>;<U0BCA>..<U0BCD>;<U0BD7>;<U0BE7>..<U0BF2>;<U0C01>..<U0C03>;/
1422        <U0C05>..<U0C0C>;<U0C0E>..<U0C10>;<U0C12>..<U0C28>;<U0C2A>..<U0C33>;/
1423        <U0C35>..<U0C39>;<U0C3E>..<U0C44>;<U0C46>..<U0C48>;<U0C4A>..<U0C4D>;/
1424        <U0C55>;<U0C56>;<U0C60>;<U0C61>;<U0C66>..<U0C6F>;<U0C82>;<U0C83>;/
1425        <U0C85>..<U0C8C>;<U0C8E>..<U0C90>;<U0C92>..<U0CA8>;<U0CAA>..<U0CB3>;/
1426        <U0CB5>..<U0CB9>;<U0CBE>..<U0CC4>;<U0CC6>..<U0CC8>;<U0CCA>..<U0CCD>;/
1427        <U0CD5>;<U0CD6>;<U0CDE>;<U0CE0>;<U0CE1>;<U0CE6>..<U0CEF>;<U0D02>;<U0D03>;/
1428        <U0D05>..<U0D0C>;<U0D0E>..<U0D10>;<U0D12>..<U0D28>;<U0D2A>..<U0D39>;/
1429        <U0D3E>..<U0D43>;<U0D46>..<U0D48>;<U0D4A>..<U0D4D>;<U0D57>;<U0D60>;<U0D61>;/
1430        <U0D66>..<U0D6F>;<U0E01>..<U0E3A>;<U0E3F>..<U0E5B>;<U0E81>;<U0E82>;<U0E84>;/
1431        <U0E87>;<U0E88>;<U0E8A>;<U0E8D>;<U0E94>..<U0E97>;<U0E99>..<U0E9F>;/
1432        <U0EA1>..<U0EA3>;<U0EA5>;<U0EA7>;<U0EAA>;<U0EAB>;<U0EAD>..<U0EB9>;/
1433        <U0EBB>..<U0EBD>;<U0EC0>..<U0EC4>;<U0EC6>;<U0EC8>..<U0ECD>;<U0ED0>..<U0ED9>;/
1434        <U0EDC>;<U0EDD>;/
1435        <U0F00>..<U0F47>;<U0F49>..<U0F69>;<U0F71>..<U0F7F>;/
1436        <U10A0>..<U10C5>;<U10D0>..<U10F6>;<U10FB>;<U1100>..<U1159>;/
1437        <U115F>..<U11A2>;<U11A8>..<U11F9>;<U1E00>..<U1E9B>;<U1EA0>..<U1EF9>;/
1438        <U1F00>..<U1F15>;<U1F18>..<U1F1D>;<U1F20>..<U1F45>;<U1F48>..<U1F4D>;/
1439        <U1F50>..<U1F57>;<U1F59>;<U1F5B>;<U1F5D>;<U1F5F>..<U1F7D>;<U1F80>..<U1FB4>;/
1440        <U1FB6>..<U1FC4>;<U1FC6>..<U1FD3>;<U1FD6>..<U1FDB>;<U1FDD>..<U1FEF>;/
1441        <U1FF2>..<U1FF4>;<U1FF6>..<U1FFE>;<U2000>..<U202E>;<U2030>..<U2046>;/
1442        <U206A>..<U2070>;<U2074>..<U208E>;<U20A0>..<U20AB>;<U20D0>..<U20E1>;/
1443        <U2100>..<U2138>;<U2153>..<U2182>;<U2190>..<U21EA>;<U2200>..<U22F1>;<U2300>;/
1444        <U2302>..<U237A>;<U2400>..<U2424>;<U2440>..<U244A>;<U2460>..<U24EA>;/
1445        <U2500>..<U2595>;<U25A0>..<U25EF>;<U2600>..<U2613>;<U261A>..<U266F>;/
1446        <U2701>..<U2704>;<U2706>..<U2709>;<U270C>..<U2727>;<U2729>..<U274B>;<U274D>;/
1447        <U274F>..<U2752>;<U2756>;<U2758>..<U275E>;<U2761>..<U2767>;<U2776>..<U2794>;/
1448        <U2798>..<U27AF>;<U27B1>..<U27BE>;<U3000>..<U3037>;<U303F>;<U3041>..<U3094>;/
1449        <U3099>..<U309E>;<U30A1>..<U30FE>;<U3105>..<U312C>;<U3131>..<U318E>;/
1450        <U3190>..<U319F>;<U3200>..<U321C>;<U3220>..<U3243>;<U3260>..<U327B>;/
1451        <U327F>..<U32B0>;<U32C0>..<U32CB>;<U32D0>..<U32FE>;<U3300>..<U3376>;/
1452        <U337B>..<U33DD>;<U33E0>..<U33FE>;<UFB00>..<UFB06>;<UFB13>..<UFB17>;/
1453        <UFB1E>..<UFB36>;<UFB38>..<UFB3C>;<UFB3E>;<UFB40>;<UFB41>;<UFB43>;<UFB44>;/
1454        <UFB46>..<UFBB1>;<UFBD3>..<UFD3F>;<UFD50>..<UFD8F>;<UFD92>..<UFDC7>;/
1455        <UFDF0>..<UFDFB>;<UFE20>..<UFE23>;<UFE30>..<UFE44>;<UFE49>..<UFE52>;/
1456        <UFE54>..<UFE66>;<UFE68>..<UFE6B>;<UFE70>..<UFE72>;<UFE74>;<UFE76>..<UFEFC>;/
1457        <UFEFF>;<UFF01>..<UFF5E>;<UFF61>..<UFFBE>;<UFFC2>..<UFFC7>;/
1458        <UFFCA>..<UFFCF>;<UFFD2>..<UFFD7>;<UFFDA>..<UFFDC>;<UFFE0>..<UFFE6>;/
1459        <UFFE8>..<UFFEE>;<UFFFD>
1460    %
1461    % "print" is by default "graph", and the <space> character
1462    %
1463    xdigit  <U0030>..<U0039>;<U0041>..<U0046>;<U0061>..<U0066>
1464    %
1465    blank   <U0008>;<U0020>;<U2000>..<U2006>;<U2008>..<U200B>;<U3000>
1466    %
1467    toupper /
1468        (<U0061>,<U0041>);(<U0062>,<U0042>);(<U0063>,<U0043>);(<U0064>,<U0044>);/
1469        (<U0065>,<U0045>);(<U0066>,<U0046>);(<U0067>,<U0047>);(<U0068>,<U0048>);/
1470        (<U0069>,<U0049>);(<U006A>,<U004A>);(<U006B>,<U004B>);(<U006C>,<U004C>);/
1471        (<U006D>,<U004D>);(<U006E>,<U004E>);(<U006F>,<U004F>);(<U0070>,<U0050>);/
1472        (<U0071>,<U0051>);(<U0072>,<U0052>);(<U0073>,<U0053>);(<U0074>,<U0054>);/
1473        (<U0075>,<U0055>);(<U0076>,<U0056>);(<U0077>,<U0057>);(<U0078>,<U0058>);/
1474        (<U0079>,<U0059>);(<U007A>,<U005A>);(<U00E0>,<U00C0>);(<U00E1>,<U00C1>);/
1475        (<U00E2>,<U00C2>);(<U00E3>,<U00C3>);(<U00E4>,<U00C4>);(<U00E5>,<U00C5>);/
1476        (<U00E6>,<U00C6>);(<U00E7>,<U00C7>);(<U00E8>,<U00C8>);(<U00E9>,<U00C9>);/
1477        (<U00EA>,<U00CA>);(<U00EB>,<U00CB>);(<U00EC>,<U00CC>);(<U00ED>,<U00CD>);/
```

```
1478        (<U00EE>,<U00CE>);(<U00EF>,<U00CF>);(<U00F0>,<U00D0>);(<U00F1>,<U00D1>);/
1479        (<U00F2>,<U00D2>);(<U00F3>,<U00D3>);(<U00F4>,<U00D4>);(<U00F5>,<U00D5>);/
1480        (<U00F6>,<U00D6>);(<U00F8>,<U00D8>);(<U00F9>,<U00D9>);(<U00FA>,<U00DA>);/
1481        (<U00FB>,<U00DB>);(<U00FC>,<U00DC>);(<U00FD>,<U00DD>);(<U00FE>,<U00DE>);/
1482        (<U00FF>,<U0178>);(<U0101>,<U0100>);(<U0103>,<U0102>);(<U0105>,<U0104>);/
1483        (<U0107>,<U0106>);(<U0109>,<U0108>);(<U010B>,<U010A>);(<U010D>,<U010C>);/
1484        (<U010F>,<U010E>);(<U0111>,<U0110>);(<U0113>,<U0112>);(<U0115>,<U0114>);/
1485        (<U0117>,<U0116>);(<U0119>,<U0118>);(<U011B>,<U011A>);(<U011D>,<U011C>);/
1486        (<U011F>,<U011E>);(<U0121>,<U0120>);(<U0123>,<U0122>);(<U0125>,<U0124>);/
1487        (<U0127>,<U0126>);(<U0129>,<U0128>);(<U012B>,<U012A>);(<U012D>,<U012C>);/
1488        (<U012F>,<U012E>);(<U0133>,<U0132>);(<U0135>,<U0134>);(<U0137>,<U0136>);/
1489        (<U013A>,<U0139>);(<U013C>,<U013B>);(<U013E>,<U013D>);(<U0140>,<U013F>);/
1490        (<U0142>,<U0141>);(<U0144>,<U0143>);(<U0146>,<U0145>);(<U0148>,<U0147>);/
1491        (<U014B>,<U014A>);(<U014D>,<U014C>);(<U014F>,<U014E>);(<U0151>,<U0150>);/
1492        (<U0153>,<U0152>);(<U0155>,<U0154>);(<U0157>,<U0156>);(<U0159>,<U0158>);/
1493        (<U015B>,<U015A>);(<U015D>,<U015C>);(<U015F>,<U015E>);(<U0161>,<U0160>);/
1494        (<U0163>,<U0162>);(<U0165>,<U0164>);(<U0167>,<U0166>);(<U0169>,<U0168>);/
1495        (<U016B>,<U016A>);(<U016D>,<U016C>);(<U016F>,<U016E>);(<U0171>,<U0170>);/
1496        (<U0173>,<U0172>);(<U0175>,<U0174>);(<U0177>,<U0176>);(<U017A>,<U0179>);/
1497        (<U017C>,<U017B>);(<U017E>,<U017D>);(<U017F>,<U0053>);(<U0183>,<U0182>);/
1498        (<U0185>,<U0184>);(<U0188>,<U0187>);(<U018C>,<U018B>);(<U0192>,<U0191>);/
1499        (<U0199>,<U0198>);(<U01A1>,<U01A0>);(<U01A3>,<U01A2>);(<U01A5>,<U01A4>);/
1500        (<U01A8>,<U01A7>);(<U01AD>,<U01AC>);(<U01B0>,<U01AF>);(<U01B4>,<U01B3>);/
1501        (<U01B6>,<U01B5>);(<U01B9>,<U01B8>);(<U01BD>,<U01BC>);(<U01C5>,<U01C4>);/
1502        (<U01C6>,<U01C4>);(<U01C8>,<U01C7>);/
1503        (<U01C9>,<U01C7>);(<U01CB>,<U01CA>);(<U01CC>,<U01CA>);/
1504        (<U01CE>,<U01CD>);(<U01D0>,<U01CF>);(<U01D2>,<U01D1>);(<U01D4>,<U01D3>);/
1505        (<U01D6>,<U01D5>);(<U01D8>,<U01D7>);(<U01DA>,<U01D9>);(<U01DC>,<U01DB>);/
1506        (<U01DD>,<U018E>);(<U01DF>,<U01DE>);(<U01E1>,<U01E0>);(<U01E3>,<U01E2>);/
1507        (<U01E5>,<U01E4>);(<U01E7>,<U01E6>);(<U01E9>,<U01E8>);(<U01EB>,<U01EA>);/
1508        (<U01ED>,<U01EC>);(<U01EF>,<U01EE>);(<U01F2>,<U01F1>);/
1509        (<U01F3>,<U01F1>);(<U01F5>,<U01F4>);(<U01FB>,<U01FA>);(<U01FD>,<U01FC>);/
1510        (<U01FF>,<U01FE>);(<U0201>,<U0200>);(<U0203>,<U0202>);(<U0205>,<U0204>);/
1511        (<U0207>,<U0206>);(<U0209>,<U0208>);(<U020B>,<U020A>);(<U020D>,<U020C>);/
1512        (<U020F>,<U020E>);(<U0211>,<U0210>);(<U0213>,<U0212>);(<U0215>,<U0214>);/
1513        (<U0217>,<U0216>);(<U0253>,<U0181>);(<U0254>,<U0186>);(<U0256>,<U0189>);/
1514        (<U0257>,<U018A>);(<U0259>,<U018F>);(<U025B>,<U0190>);(<U0260>,<U0193>);/
1515        (<U0263>,<U0194>);(<U0268>,<U0197>);(<U0269>,<U0196>);(<U026F>,<U019C>);/
1516        (<U0272>,<U019D>);(<U0275>,<U019F>);(<U0283>,<U01A9>);(<U0288>,<U01AE>);/
1517        (<U028A>,<U01B1>);(<U028B>,<U01B2>);(<U0292>,<U01B7>);(<U03AC>,<U0386>);/
1518        (<U03AD>,<U0388>);(<U03AE>,<U0389>);(<U03AF>,<U038A>);(<U03B1>,<U0391>);/
1519        (<U03B2>,<U0392>);(<U03B3>,<U0393>);(<U03B4>,<U0394>);(<U03B5>,<U0395>);/
1520        (<U03B6>,<U0396>);(<U03B7>,<U0397>);(<U03B8>,<U0398>);(<U03B9>,<U0399>);/
1521        (<U03BA>,<U039A>);(<U03BB>,<U039B>);(<U03BC>,<U039C>);(<U03BD>,<U039D>);/
1522        (<U03BE>,<U039E>);(<U03BF>,<U039F>);(<U03C0>,<U03A0>);(<U03C1>,<U03A1>);/
1523        (<U03C2>,<U03A3>);(<U03C3>,<U03A3>);(<U03C4>,<U03A4>);(<U03C5>,<U03A5>);/
1524        (<U03C6>,<U03A6>);(<U03C7>,<U03A7>);(<U03C8>,<U03A8>);(<U03C9>,<U03A9>);/
1525        (<U03CA>,<U03AA>);(<U03CB>,<U03AB>);(<U03CC>,<U038C>);(<U03CD>,<U038E>);/
1526        (<U03CE>,<U038F>);/
1527        (<U03E3>,<U03E2>);(<U03E5>,<U03E4>);(<U03E7>,<U03E6>);(<U03E9>,<U03E8>);/
1528        (<U03EB>,<U03EA>);(<U03ED>,<U03EC>);(<U03EF>,<U03EE>);/
1529        (<U0430>,<U0410>);(<U0431>,<U0411>);(<U0432>,<U0412>);/
1530        (<U0433>,<U0413>);(<U0434>,<U0414>);(<U0435>,<U0415>);(<U0436>,<U0416>);/
1531        (<U0437>,<U0417>);(<U0438>,<U0418>);(<U0439>,<U0419>);(<U043A>,<U041A>);/
1532        (<U043B>,<U041B>);(<U043C>,<U041C>);(<U043D>,<U041D>);(<U043E>,<U041E>);/
1533        (<U043F>,<U041F>);(<U0440>,<U0420>);(<U0441>,<U0421>);(<U0442>,<U0422>);/
1534        (<U0443>,<U0423>);(<U0444>,<U0424>);(<U0445>,<U0425>);(<U0446>,<U0426>);/
1535        (<U0447>,<U0427>);(<U0448>,<U0428>);(<U0449>,<U0429>);(<U044A>,<U042A>);/
1536        (<U044B>,<U042B>);(<U044C>,<U042C>);(<U044D>,<U042D>);(<U044E>,<U042E>);/
1537        (<U044F>,<U042F>);(<U0451>,<U0401>);(<U0452>,<U0402>);(<U0453>,<U0403>);/
1538        (<U0454>,<U0404>);(<U0455>,<U0405>);(<U0456>,<U0406>);(<U0457>,<U0407>);/
1539        (<U0458>,<U0408>);(<U0459>,<U0409>);(<U045A>,<U040A>);(<U045B>,<U040B>);/
1540        (<U045C>,<U040C>);(<U045E>,<U040E>);(<U045F>,<U040F>);(<U0461>,<U0460>);/
1541        (<U0463>,<U0462>);(<U0465>,<U0464>);(<U0467>,<U0466>);(<U0469>,<U0468>);/
1542        (<U046B>,<U046A>);(<U046D>,<U046C>);(<U046F>,<U046E>);(<U0471>,<U0470>);/
1543        (<U0473>,<U0472>);(<U0475>,<U0474>);(<U0477>,<U0476>);(<U0479>,<U0478>);/
1544        (<U047B>,<U047A>);(<U047D>,<U047C>);(<U047F>,<U047E>);(<U0481>,<U0480>);/
1545        (<U0491>,<U0490>);(<U0493>,<U0492>);(<U0495>,<U0494>);(<U0497>,<U0496>);/
1546        (<U0499>,<U0498>);(<U049B>,<U049A>);(<U049D>,<U049C>);(<U049F>,<U049E>);/
1547        (<U04A1>,<U04A0>);(<U04A3>,<U04A2>);(<U04A5>,<U04A4>);(<U04A7>,<U04A6>);/
1548        (<U04A9>,<U04A8>);(<U04AB>,<U04AA>);(<U04AD>,<U04AC>);(<U04AF>,<U04AE>);/
1549        (<U04B1>,<U04B0>);(<U04B3>,<U04B2>);(<U04B5>,<U04B4>);(<U04B7>,<U04B6>);/
1550        (<U04B9>,<U04B8>);(<U04BB>,<U04BA>);(<U04BD>,<U04BC>);(<U04BF>,<U04BE>);/
1551        (<U04C2>,<U04C1>);(<U04C4>,<U04C3>);(<U04C8>,<U04C7>);(<U04CC>,<U04CB>);/
1552        (<U04D1>,<U04D0>);(<U04D3>,<U04D2>);(<U04D5>,<U04D4>);(<U04D7>,<U04D6>);/
1553        (<U04D9>,<U04D8>);(<U04DB>,<U04DA>);(<U04DD>,<U04DC>);(<U04DF>,<U04DE>);/
1554        (<U04E1>,<U04E0>);(<U04E3>,<U04E2>);(<U04E5>,<U04E4>);(<U04E7>,<U04E6>);/
1555        (<U04E9>,<U04E8>);(<U04EB>,<U04EA>);(<U04EF>,<U04EE>);(<U04F1>,<U04F0>);/
1556        (<U04F3>,<U04F2>);(<U04F5>,<U04F4>);(<U04F9>,<U04F8>);(<U0561>,<U0531>);/
```

```
1557        (<U0562>,<U0532>);(<U0563>,<U0533>);(<U0564>,<U0534>);(<U0565>,<U0535>);/
1558        (<U0566>,<U0536>);(<U0567>,<U0537>);(<U0568>,<U0538>);(<U0569>,<U0539>);/
1559        (<U056A>,<U053A>);(<U056B>,<U053B>);(<U056C>,<U053C>);(<U056D>,<U053D>);/
1560        (<U056E>,<U053E>);(<U056F>,<U053F>);(<U0570>,<U0540>);(<U0571>,<U0541>);/
1561        (<U0572>,<U0542>);(<U0573>,<U0543>);(<U0574>,<U0544>);(<U0575>,<U0545>);/
1562        (<U0576>,<U0546>);(<U0577>,<U0547>);(<U0578>,<U0548>);(<U0579>,<U0549>);/
1563        (<U057A>,<U054A>);(<U057B>,<U054B>);(<U057C>,<U054C>);(<U057D>,<U054D>);/
1564        (<U057E>,<U054E>);(<U057F>,<U054F>);(<U0580>,<U0550>);(<U0581>,<U0551>);/
1565        (<U0582>,<U0552>);(<U0583>,<U0553>);(<U0584>,<U0554>);(<U0585>,<U0555>);/
1566        (<U0586>,<U0556>);/
1567        (<U10D0>,<U10A0>);(<U10D1>,<U10A1>);(<U10D2>,<U10A2>);(<U10D3>,<U10A3>);/
1568        (<U10D4>,<U10A4>);(<U10D5>,<U10A5>);(<U10D6>,<U10A6>);(<U10D7>,<U10A7>);/
1569        (<U10D8>,<U10A8>);(<U10D9>,<U10A9>);(<U10DA>,<U10AA>);(<U10DB>,<U10AB>);/
1570        (<U10DC>,<U10AC>);(<U10DD>,<U10AD>);(<U10DE>,<U10AE>);(<U10DF>,<U10AF>);/
1571        (<U10E0>,<U10B0>);(<U10E1>,<U10B1>);(<U10E2>,<U10B2>);(<U10E3>,<U10B3>);/
1572        (<U10E4>,<U10B4>);(<U10E5>,<U10B5>);(<U10E6>,<U10B6>);(<U10E7>,<U10B7>);/
1573        (<U10E8>,<U10B8>);(<U10E9>,<U10B9>);(<U10EA>,<U10BA>);(<U10EB>,<U10BB>);/
1574        (<U10EC>,<U10BC>);(<U10ED>,<U10BD>);(<U10EE>,<U10BE>);(<U10EF>,<U10BF>);/
1575        (<U10F0>,<U10C0>);(<U10F1>,<U10C1>);(<U10F2>,<U10C2>);(<U10F3>,<U10C3>);/
1576        (<U10F4>,<U10C4>);(<U10F5>,<U10C5>);/
1577        (<U1E01>,<U1E00>);(<U1E03>,<U1E02>);(<U1E05>,<U1E04>);/
1578        (<U1E07>,<U1E06>);(<U1E09>,<U1E08>);(<U1E0B>,<U1E0A>);(<U1E0D>,<U1E0C>);/
1579        (<U1E0F>,<U1E0E>);(<U1E11>,<U1E10>);(<U1E13>,<U1E12>);(<U1E15>,<U1E14>);/
1580        (<U1E17>,<U1E16>);(<U1E19>,<U1E18>);(<U1E1B>,<U1E1A>);(<U1E1D>,<U1E1C>);/
1581        (<U1E1F>,<U1E1E>);(<U1E21>,<U1E20>);(<U1E23>,<U1E22>);(<U1E25>,<U1E24>);/
1582        (<U1E27>,<U1E26>);(<U1E29>,<U1E28>);(<U1E2B>,<U1E2A>);(<U1E2D>,<U1E2C>);/
1583        (<U1E2F>,<U1E2E>);(<U1E31>,<U1E30>);(<U1E33>,<U1E32>);(<U1E35>,<U1E34>);/
1584        (<U1E37>,<U1E36>);(<U1E39>,<U1E38>);(<U1E3B>,<U1E3A>);(<U1E3D>,<U1E3C>);/
1585        (<U1E3F>,<U1E3E>);(<U1E41>,<U1E40>);(<U1E43>,<U1E42>);(<U1E45>,<U1E44>);/
1586        (<U1E47>,<U1E46>);(<U1E49>,<U1E48>);(<U1E4B>,<U1E4A>);(<U1E4D>,<U1E4C>);/
1587        (<U1E4F>,<U1E4E>);(<U1E51>,<U1E50>);(<U1E53>,<U1E52>);(<U1E55>,<U1E54>);/
1588        (<U1E57>,<U1E56>);(<U1E59>,<U1E58>);(<U1E5B>,<U1E5A>);(<U1E5D>,<U1E5C>);/
1589        (<U1E5F>,<U1E5E>);(<U1E61>,<U1E60>);(<U1E63>,<U1E62>);(<U1E65>,<U1E64>);/
1590        (<U1E67>,<U1E66>);(<U1E69>,<U1E68>);(<U1E6B>,<U1E6A>);(<U1E6D>,<U1E6C>);/
1591        (<U1E6F>,<U1E6E>);(<U1E71>,<U1E70>);(<U1E73>,<U1E72>);(<U1E75>,<U1E74>);/
1592        (<U1E77>,<U1E76>);(<U1E79>,<U1E78>);(<U1E7B>,<U1E7A>);(<U1E7D>,<U1E7C>);/
1593        (<U1E7F>,<U1E7E>);(<U1E81>,<U1E80>);(<U1E83>,<U1E82>);(<U1E85>,<U1E84>);/
1594        (<U1E87>,<U1E86>);(<U1E89>,<U1E88>);(<U1E8B>,<U1E8A>);(<U1E8D>,<U1E8C>);/
1595        (<U1E8F>,<U1E8E>);(<U1E91>,<U1E90>);(<U1E93>,<U1E92>);(<U1E95>,<U1E94>);/
1596        (<U1E9B>,<U1E60>);/
1597        (<U1EA1>,<U1EA0>);(<U1EA3>,<U1EA2>);(<U1EA5>,<U1EA4>);(<U1EA7>,<U1EA6>);/
1598        (<U1EA9>,<U1EA8>);(<U1EAB>,<U1EAA>);(<U1EAD>,<U1EAC>);(<U1EAF>,<U1EAE>);/
1599        (<U1EB1>,<U1EB0>);(<U1EB3>,<U1EB2>);(<U1EB5>,<U1EB4>);(<U1EB7>,<U1EB6>);/
1600        (<U1EB9>,<U1EB8>);(<U1EBB>,<U1EBA>);(<U1EBD>,<U1EBC>);(<U1EBF>,<U1EBE>);/
1601        (<U1EC1>,<U1EC0>);(<U1EC3>,<U1EC2>);(<U1EC5>,<U1EC4>);(<U1EC7>,<U1EC6>);/
1602        (<U1EC9>,<U1EC8>);(<U1ECB>,<U1ECA>);(<U1ECD>,<U1ECC>);(<U1ECF>,<U1ECE>);/
1603        (<U1ED1>,<U1ED0>);(<U1ED3>,<U1ED2>);(<U1ED5>,<U1ED4>);(<U1ED7>,<U1ED6>);/
1604        (<U1ED9>,<U1ED8>);(<U1EDB>,<U1EDA>);(<U1EDD>,<U1EDC>);(<U1EDF>,<U1EDE>);/
1605        (<U1EE1>,<U1EE0>);(<U1EE3>,<U1EE2>);(<U1EE5>,<U1EE4>);(<U1EE7>,<U1EE6>);/
1606        (<U1EE9>,<U1EE8>);(<U1EEB>,<U1EEA>);(<U1EED>,<U1EEC>);(<U1EEF>,<U1EEE>);/
1607        (<U1EF1>,<U1EF0>);(<U1EF3>,<U1EF2>);(<U1EF5>,<U1EF4>);(<U1EF7>,<U1EF6>);/
1608        (<U1EF9>,<U1EF8>);(<U1F00>,<U1F08>);(<U1F01>,<U1F09>);(<U1F02>,<U1F0A>);/
1609        (<U1F03>,<U1F0B>);(<U1F04>,<U1F0C>);(<U1F05>,<U1F0D>);(<U1F06>,<U1F0E>);/
1610        (<U1F07>,<U1F0F>);(<U1F10>,<U1F18>);(<U1F11>,<U1F19>);(<U1F12>,<U1F1A>);/
1611        (<U1F13>,<U1F1B>);(<U1F14>,<U1F1C>);(<U1F15>,<U1F1D>);(<U1F20>,<U1F28>);/
1612        (<U1F21>,<U1F29>);(<U1F22>,<U1F2A>);(<U1F23>,<U1F2B>);(<U1F24>,<U1F2C>);/
1613        (<U1F25>,<U1F2D>);(<U1F26>,<U1F2E>);(<U1F27>,<U1F2F>);(<U1F30>,<U1F38>);/
1614        (<U1F31>,<U1F39>);(<U1F32>,<U1F3A>);(<U1F33>,<U1F3B>);(<U1F34>,<U1F3C>);/
1615        (<U1F35>,<U1F3D>);(<U1F36>,<U1F3E>);(<U1F37>,<U1F3F>);(<U1F40>,<U1F48>);/
1616        (<U1F41>,<U1F49>);(<U1F42>,<U1F4A>);(<U1F43>,<U1F4B>);(<U1F44>,<U1F4C>);/
1617        (<U1F45>,<U1F4D>);(<U1F51>,<U1F59>);(<U1F53>,<U1F5B>);(<U1F55>,<U1F5D>);/
1618        (<U1F57>,<U1F5F>);(<U1F60>,<U1F68>);(<U1F61>,<U1F69>);(<U1F62>,<U1F6A>);/
1619        (<U1F63>,<U1F6B>);(<U1F64>,<U1F6C>);(<U1F65>,<U1F6D>);(<U1F66>,<U1F6E>);/
1620        (<U1F67>,<U1F6F>);(<U1F70>,<U1FBA>);(<U1F71>,<U1FBB>);(<U1F72>,<U1FC8>);/
1621        (<U1F73>,<U1FC9>);(<U1F74>,<U1FCA>);(<U1F75>,<U1FCB>);(<U1F76>,<U1FDA>);/
1622        (<U1F77>,<U1FDB>);(<U1F78>,<U1FF8>);(<U1F79>,<U1FF9>);(<U1F7A>,<U1FEA>);/
1623        (<U1F7B>,<U1FEB>);(<U1F7C>,<U1FFA>);(<U1F7D>,<U1FFB>);(<U1F80>,<U1F88>);/
1624        (<U1F81>,<U1F89>);(<U1F82>,<U1F8A>);(<U1F83>,<U1F8B>);(<U1F84>,<U1F8C>);/
1625        (<U1F85>,<U1F8D>);(<U1F86>,<U1F8E>);(<U1F87>,<U1F8F>);(<U1F90>,<U1F98>);/
1626        (<U1F91>,<U1F99>);(<U1F92>,<U1F9A>);(<U1F93>,<U1F9B>);(<U1F94>,<U1F9C>);/
1627        (<U1F95>,<U1F9D>);(<U1F96>,<U1F9E>);(<U1F97>,<U1F9F>);(<U1FA0>,<U1FA8>);/
1628        (<U1FA1>,<U1FA9>);(<U1FA2>,<U1FAA>);(<U1FA3>,<U1FAB>);(<U1FA4>,<U1FAC>);/
1629        (<U1FA5>,<U1FAD>);(<U1FA6>,<U1FAE>);(<U1FA7>,<U1FAF>);(<U1FB0>,<U1FB8>);/
1630        (<U1FB1>,<U1FB9>);(<U1FB3>,<U1FBC>);(<U1FC3>,<U1FCC>);(<U1FD0>,<U1FD8>);/
1631        (<U1FD1>,<U1FD9>);(<U1FE0>,<U1FE8>);(<U1FE1>,<U1FE9>);(<U1FE5>,<U1FEC>);/
1632        (<U1FF3>,<U1FFC>)
1633    tolower /
1634        (<U0041>,<U0061>);(<U0042>,<U0062>);(<U0043>,<U0063>);(<U0044>,<U0064>);/
```

```
1635          (<U0045>,<U0065>);(<U0046>,<U0066>);(<U0047>,<U0067>);(<U0048>,<U0068>);/
1636          (<U0049>,<U0069>);(<U004A>,<U006A>);(<U004B>,<U006B>);(<U004C>,<U006C>);/
1637          (<U004D>,<U006D>);(<U004E>,<U006E>);(<U004F>,<U006F>);(<U0050>,<U0070>);/
1638          (<U0051>,<U0071>);(<U0052>,<U0072>);(<U0053>,<U0073>);(<U0054>,<U0074>);/
1639          (<U0055>,<U0075>);(<U0056>,<U0076>);(<U0057>,<U0077>);(<U0058>,<U0078>);/
1640          (<U0059>,<U0079>);(<U005A>,<U007A>);(<U00C0>,<U00E0>);(<U00C1>,<U00E1>);/
1641          (<U00C2>,<U00E2>);(<U00C3>,<U00E3>);(<U00C4>,<U00E4>);(<U00C5>,<U00E5>);/
1642          (<U00C6>,<U00E6>);(<U00C7>,<U00E7>);(<U00C8>,<U00E8>);(<U00C9>,<U00E9>);/
1643          (<U00CA>,<U00EA>);(<U00CB>,<U00EB>);(<U00CC>,<U00EC>);(<U00CD>,<U00ED>);/
1644          (<U00CE>,<U00EE>);(<U00CF>,<U00EF>);(<U00D0>,<U00F0>);(<U00D1>,<U00F1>);/
1645          (<U00D2>,<U00F2>);(<U00D3>,<U00F3>);(<U00D4>,<U00F4>);(<U00D5>,<U00F5>);/
1646          (<U00D6>,<U00F6>);(<U00D8>,<U00F8>);(<U00D9>,<U00F9>);(<U00DA>,<U00FA>);/
1647          (<U00DB>,<U00FB>);(<U00DC>,<U00FC>);(<U00DD>,<U00FD>);(<U00DE>,<U00FE>);/
1648          (<U0178>,<U00FF>);(<U0100>,<U0101>);(<U0102>,<U0103>);(<U0104>,<U0105>);/
1649          (<U0106>,<U0107>);(<U0108>,<U0109>);(<U010A>,<U010B>);(<U010C>,<U010D>);/
1650          (<U010E>,<U010F>);(<U0110>,<U0111>);(<U0112>,<U0113>);(<U0114>,<U0115>);/
1651          (<U0116>,<U0117>);(<U0118>,<U0119>);(<U011A>,<U011B>);(<U011C>,<U011D>);/
1652          (<U011E>,<U011F>);(<U0120>,<U0121>);(<U0122>,<U0123>);(<U0124>,<U0125>);/
1653          (<U0126>,<U0127>);(<U0128>,<U0129>);(<U012A>,<U012B>);(<U012C>,<U012D>);/
1654          (<U012E>,<U012F>);(<U0132>,<U0133>);(<U0134>,<U0135>);(<U0136>,<U0137>);/
1655          (<U0139>,<U013A>);(<U013B>,<U013C>);(<U013D>,<U013E>);(<U013F>,<U0140>);/
1656          (<U0141>,<U0142>);(<U0143>,<U0144>);(<U0145>,<U0146>);(<U0147>,<U0148>);/
1657          (<U014A>,<U014B>);(<U014C>,<U014D>);(<U014E>,<U014F>);(<U0150>,<U0151>);/
1658          (<U0152>,<U0153>);(<U0154>,<U0155>);(<U0156>,<U0157>);(<U0158>,<U0159>);/
1659          (<U015A>,<U015B>);(<U015C>,<U015D>);(<U015E>,<U015F>);(<U0160>,<U0161>);/
1660          (<U0162>,<U0163>);(<U0164>,<U0165>);(<U0166>,<U0167>);(<U0168>,<U0169>);/
1661          (<U016A>,<U016B>);(<U016C>,<U016D>);(<U016E>,<U016F>);(<U0170>,<U0171>);/
1662          (<U0172>,<U0173>);(<U0174>,<U0175>);(<U0176>,<U0177>);(<U0179>,<U017A>);/
1663          (<U017B>,<U017C>);(<U017D>,<U017E>);(<U0182>,<U0183>);(<U0184>,<U0185>);/
1664          (<U0187>,<U0188>);(<U0256>,<U0189>);(<U018B>,<U018C>);(<U018E>,<U01DD>);/
1665          (<U0191>,<U0192>);(<U0198>,<U0199>);(<U01A0>,<U01A1>);(<U01A2>,<U01A3>);/
1666          (<U01A4>,<U01A5>);(<U01A7>,<U01A8>);(<U01AC>,<U01AD>);(<U01AF>,<U01B0>);/
1667          (<U01B3>,<U01B4>);(<U01B5>,<U01B6>);(<U01B8>,<U01B9>);(<U01BC>,<U01BD>);/
1668          (<U01C4>,<U01C6>);(<U01C5>,<U01C6>);(<U01C7>,<U01C9>);/
1669          (<U01C8>,<U01C9>);(<U01CA>,<U01CC>);(<U01CB>,<U01CC>);/
1670          (<U01CD>,<U01CE>);(<U01CF>,<U01D0>);(<U01D1>,<U01D2>);/
1671          (<U01D3>,<U01D4>);(<U01D5>,<U01D6>);(<U01D7>,<U01D8>);(<U01D9>,<U01DA>);/
1672          (<U01DB>,<U01DC>);(<U01DE>,<U01DF>);(<U01E0>,<U01E1>);(<U01E2>,<U01E3>);/
1673          (<U01E4>,<U01E5>);(<U01E6>,<U01E7>);(<U01E8>,<U01E9>);(<U01EA>,<U01EB>);/
1674          (<U01EC>,<U01ED>);(<U01EE>,<U01EF>);(<U01F1>,<U01F3>);/
1675          (<U01F2>,<U01F3>);(<U01F4>,<U01F5>);(<U01FA>,<U01FB>);(<U01FC>,<U01FD>);/
1676          (<U01FE>,<U01FF>);(<U0200>,<U0201>);(<U0202>,<U0203>);(<U0204>,<U0205>);/
1677          (<U0206>,<U0207>);(<U0208>,<U0209>);(<U020A>,<U020B>);(<U020C>,<U020D>);/
1678          (<U020E>,<U020F>);(<U0210>,<U0211>);(<U0212>,<U0213>);(<U0214>,<U0215>);/
1679          (<U0216>,<U0217>);(<U0181>,<U0253>);(<U0186>,<U0254>);(<U018A>,<U0257>);/
1680          (<U018E>,<U0258>);(<U018F>,<U0259>);(<U0190>,<U025B>);(<U0193>,<U0260>);/
1681          (<U0194>,<U0263>);(<U0197>,<U0268>);(<U0196>,<U0269>);(<U019C>,<U026F>);/
1682          (<U019D>,<U0272>);(<U01A9>,<U0283>);(<U01AE>,<U0288>);(<U01B1>,<U028A>);/
1683          (<U01B2>,<U028B>);(<U01B7>,<U0292>);(<U0386>,<U03AC>);(<U0388>,<U03AD>);/
1684          (<U0389>,<U03AE>);(<U038A>,<U03AF>);(<U0391>,<U03B1>);(<U0392>,<U03B2>);/
1685          (<U0393>,<U03B3>);(<U0394>,<U03B4>);(<U0395>,<U03B5>);(<U0396>,<U03B6>);/
1686          (<U0397>,<U03B7>);(<U0398>,<U03B8>);(<U0399>,<U03B9>);(<U039A>,<U03BA>);/
1687          (<U039B>,<U03BB>);(<U039C>,<U03BC>);(<U039D>,<U03BD>);(<U039E>,<U03BE>);/
1688          (<U039F>,<U03BF>);(<U03A0>,<U03C0>);(<U03A1>,<U03C1>);(<U03A3>,<U03C3>);/
1689          (<U03A4>,<U03C4>);(<U03A5>,<U03C5>);(<U03A6>,<U03C6>);(<U03A7>,<U03C7>);/
1690          (<U03A8>,<U03C8>);(<U03A9>,<U03C9>);(<U03AA>,<U03CA>);(<U03AB>,<U03CB>);/
1691          (<U038C>,<U03CC>);(<U038E>,<U03CD>);(<U038F>,<U03CE>);(<U03E2>,<U03E3>);/
1692          (<U03E4>,<U03E5>);(<U03E6>,<U03E7>);(<U03E8>,<U03E9>);(<U03EA>,<U03EB>);/
1693          (<U03EC>,<U03ED>);(<U03EE>,<U03EF>);(<U0410>,<U0430>);/
1694          (<U0411>,<U0431>);(<U0412>,<U0432>);(<U0413>,<U0433>);(<U0414>,<U0434>);/
1695          (<U0415>,<U0435>);(<U0416>,<U0436>);(<U0417>,<U0437>);(<U0418>,<U0438>);/
1696          (<U0419>,<U0439>);(<U041A>,<U043A>);(<U041B>,<U043B>);(<U041C>,<U043C>);/
1697          (<U041D>,<U043D>);(<U041E>,<U043E>);(<U041F>,<U043F>);(<U0420>,<U0440>);/
1698          (<U0421>,<U0441>);(<U0422>,<U0442>);(<U0423>,<U0443>);(<U0424>,<U0444>);/
1699          (<U0425>,<U0445>);(<U0426>,<U0446>);(<U0427>,<U0447>);(<U0428>,<U0448>);/
1700          (<U0429>,<U0449>);(<U042A>,<U044A>);(<U042B>,<U044B>);(<U042C>,<U044C>);/
1701          (<U042D>,<U044D>);(<U042E>,<U044E>);(<U042F>,<U044F>);(<U0401>,<U0451>);/
1702          (<U0402>,<U0452>);(<U0403>,<U0453>);(<U0404>,<U0454>);(<U0405>,<U0455>);/
1703          (<U0406>,<U0456>);(<U0407>,<U0457>);(<U0408>,<U0458>);(<U0409>,<U0459>);/
1704          (<U040A>,<U045A>);(<U040B>,<U045B>);(<U040C>,<U045C>);(<U040E>,<U045E>);/
1705          (<U040F>,<U045F>);(<U0460>,<U0461>);(<U0462>,<U0463>);(<U0464>,<U0465>);/
1706          (<U0466>,<U0467>);(<U0468>,<U0469>);(<U046A>,<U046B>);(<U046C>,<U046D>);/
1707          (<U046E>,<U046F>);(<U0470>,<U0471>);(<U0472>,<U0473>);(<U0474>,<U0475>);/
1708          (<U0476>,<U0477>);(<U0478>,<U0479>);(<U047A>,<U047B>);(<U047C>,<U047D>);/
1709          (<U047E>,<U047F>);(<U0480>,<U0481>);(<U0490>,<U0491>);(<U0492>,<U0493>);/
1710          (<U0494>,<U0495>);(<U0496>,<U0497>);(<U0498>,<U0499>);(<U049A>,<U049B>);/
1711          (<U049C>,<U049D>);(<U049E>,<U049F>);(<U04A0>,<U04A1>);(<U04A2>,<U04A3>);/
1712          (<U04A4>,<U04A5>);(<U04A6>,<U04A7>);(<U04A8>,<U04A9>);(<U04AA>,<U04AB>);/
1713          (<U04AC>,<U04AD>);(<U04AE>,<U04AF>);(<U04B0>,<U04B1>);(<U04B2>,<U04B3>);/
```

```
1714          (<U04B4>,<U04B5>);(<U04B6>,<U04B7>);(<U04B8>,<U04B9>);(<U04BA>,<U04BB>);/
1715          (<U04BC>,<U04BD>);(<U04BE>,<U04BF>);(<U04C1>,<U04C2>);(<U04C3>,<U04C4>);/
1716          (<U04C7>,<U04C8>);(<U04CB>,<U04CC>);(<U04D0>,<U04D1>);(<U04D2>,<U04D3>);/
1717          (<U04D4>,<U04D5>);(<U04D6>,<U04D7>);(<U04D8>,<U04D9>);(<U04DA>,<U04DB>);/
1718          (<U04DC>,<U04DD>);(<U04DE>,<U04DF>);(<U04E0>,<U04E1>);(<U04E2>,<U04E3>);/
1719          (<U04E4>,<U04E5>);(<U04E6>,<U04E7>);(<U04E8>,<U04E9>);(<U04EA>,<U04EB>);/
1720          (<U04EE>,<U04EF>);(<U04F0>,<U04F1>);(<U04F2>,<U04F3>);(<U04F4>,<U04F5>);/
1721          (<U04F8>,<U04F9>);(<U0531>,<U0561>);(<U0532>,<U0562>);(<U0533>,<U0563>);/
1722          (<U0534>,<U0564>);(<U0535>,<U0565>);(<U0536>,<U0566>);(<U0537>,<U0567>);/
1723          (<U0538>,<U0568>);(<U0539>,<U0569>);(<U053A>,<U056A>);(<U053B>,<U056B>);/
1724          (<U053C>,<U056C>);(<U053D>,<U056D>);(<U053E>,<U056E>);(<U053F>,<U056F>);/
1725          (<U0540>,<U0570>);(<U0541>,<U0571>);(<U0542>,<U0572>);(<U0543>,<U0573>);/
1726          (<U0544>,<U0574>);(<U0545>,<U0575>);(<U0546>,<U0576>);(<U0547>,<U0577>);/
1727          (<U0548>,<U0578>);(<U0549>,<U0579>);(<U054A>,<U057A>);(<U054B>,<U057B>);/
1728          (<U054C>,<U057C>);(<U054D>,<U057D>);(<U054E>,<U057E>);(<U054F>,<U057F>);/
1729          (<U0550>,<U0580>);(<U0551>,<U0581>);(<U0552>,<U0582>);(<U0553>,<U0583>);/
1730          (<U0554>,<U0584>);(<U0555>,<U0585>);(<U0556>,<U0586>);/
1731          (<U10A0>,<U10D0>);(<U10A1>,<U10D1>);(<U10A2>,<U10D2>);(<U10A3>,<U10D3>);/
1732          (<U10A4>,<U10D4>);(<U10A5>,<U10D5>);(<U10A6>,<U10D6>);(<U10A7>,<U10D7>);/
1733          (<U10A8>,<U10D8>);(<U10A9>,<U10D9>);(<U10AA>,<U10DA>);(<U10AB>,<U10DB>);/
1734          (<U10AC>,<U10DC>);(<U10AD>,<U10DD>);(<U10AE>,<U10DE>);(<U10AF>,<U10DF>);/
1735          (<U10B0>,<U10E0>);(<U10B1>,<U10E1>);(<U10B2>,<U10E2>);(<U10B3>,<U10E3>);/
1736          (<U10B4>,<U10E4>);(<U10B5>,<U10E5>);(<U10B6>,<U10E6>);(<U10B7>,<U10E7>);/
1737          (<U10B8>,<U10E8>);(<U10B9>,<U10E9>);(<U10BA>,<U10EA>);(<U10BB>,<U10EB>);/
1738          (<U10BC>,<U10EC>);(<U10BD>,<U10ED>);(<U10BE>,<U10EE>);(<U10BF>,<U10EF>);/
1739          (<U10C0>,<U10F0>);(<U10C1>,<U10F1>);(<U10C2>,<U10F2>);(<U10C3>,<U10F3>);/
1740          (<U10C4>,<U10F4>);(<U10C5>,<U10F5>);/
1741          (<U1E00>,<U1E01>);/
1742          (<U1E02>,<U1E03>);(<U1E04>,<U1E05>);(<U1E06>,<U1E07>);(<U1E08>,<U1E09>);/
1743          (<U1E0A>,<U1E0B>);(<U1E0C>,<U1E0D>);(<U1E0E>,<U1E0F>);(<U1E10>,<U1E11>);/
1744          (<U1E12>,<U1E13>);(<U1E14>,<U1E15>);(<U1E16>,<U1E17>);(<U1E18>,<U1E19>);/
1745          (<U1E1A>,<U1E1B>);(<U1E1C>,<U1E1D>);(<U1E1E>,<U1E1F>);(<U1E20>,<U1E21>);/
1746          (<U1E22>,<U1E23>);(<U1E24>,<U1E25>);(<U1E26>,<U1E27>);(<U1E28>,<U1E29>);/
1747          (<U1E2A>,<U1E2B>);(<U1E2C>,<U1E2D>);(<U1E2E>,<U1E2F>);(<U1E30>,<U1E31>);/
1748          (<U1E32>,<U1E33>);(<U1E34>,<U1E35>);(<U1E36>,<U1E37>);(<U1E38>,<U1E39>);/
1749          (<U1E3A>,<U1E3B>);(<U1E3C>,<U1E3D>);(<U1E3E>,<U1E3F>);(<U1E40>,<U1E41>);/
1750          (<U1E42>,<U1E43>);(<U1E44>,<U1E45>);(<U1E46>,<U1E47>);(<U1E48>,<U1E49>);/
1751          (<U1E4A>,<U1E4B>);(<U1E4C>,<U1E4D>);(<U1E4E>,<U1E4F>);(<U1E50>,<U1E51>);/
1752          (<U1E52>,<U1E53>);(<U1E54>,<U1E55>);(<U1E56>,<U1E57>);(<U1E58>,<U1E59>);/
1753          (<U1E5A>,<U1E5B>);(<U1E5C>,<U1E5D>);(<U1E5E>,<U1E5F>);(<U1E60>,<U1E61>);/
1754          (<U1E62>,<U1E63>);(<U1E64>,<U1E65>);(<U1E66>,<U1E67>);(<U1E68>,<U1E69>);/
1755          (<U1E6A>,<U1E6B>);(<U1E6C>,<U1E6D>);(<U1E6E>,<U1E6F>);(<U1E70>,<U1E71>);/
1756          (<U1E72>,<U1E73>);(<U1E74>,<U1E75>);(<U1E76>,<U1E77>);(<U1E78>,<U1E79>);/
1757          (<U1E7A>,<U1E7B>);(<U1E7C>,<U1E7D>);(<U1E7E>,<U1E7F>);(<U1E80>,<U1E81>);/
1758          (<U1E82>,<U1E83>);(<U1E84>,<U1E85>);(<U1E86>,<U1E87>);(<U1E88>,<U1E89>);/
1759          (<U1E8A>,<U1E8B>);(<U1E8C>,<U1E8D>);(<U1E8E>,<U1E8F>);(<U1E90>,<U1E91>);/
1760          (<U1E92>,<U1E93>);(<U1E94>,<U1E95>);(<U1EA0>,<U1EA1>);(<U1EA2>,<U1EA3>);/
1761          (<U1EA4>,<U1EA5>);(<U1EA6>,<U1EA7>);(<U1EA8>,<U1EA9>);(<U1EAA>,<U1EAB>);/
1762          (<U1EAC>,<U1EAD>);(<U1EAE>,<U1EAF>);(<U1EB0>,<U1EB1>);(<U1EB2>,<U1EB3>);/
1763          (<U1EB4>,<U1EB5>);(<U1EB6>,<U1EB7>);(<U1EB8>,<U1EB9>);(<U1EBA>,<U1EBB>);/
1764          (<U1EBC>,<U1EBD>);(<U1EBE>,<U1EBF>);(<U1EC0>,<U1EC1>);(<U1EC2>,<U1EC3>);/
1765          (<U1EC4>,<U1EC5>);(<U1EC6>,<U1EC7>);(<U1EC8>,<U1EC9>);(<U1ECA>,<U1ECB>);/
1766          (<U1ECC>,<U1ECD>);(<U1ECE>,<U1ECF>);(<U1ED0>,<U1ED1>);(<U1ED2>,<U1ED3>);/
1767          (<U1ED4>,<U1ED5>);(<U1ED6>,<U1ED7>);(<U1ED8>,<U1ED9>);(<U1EDA>,<U1EDB>);/
1768          (<U1EDC>,<U1EDD>);(<U1EDE>,<U1EDF>);(<U1EE0>,<U1EE1>);(<U1EE2>,<U1EE3>);/
1769          (<U1EE4>,<U1EE5>);(<U1EE6>,<U1EE7>);(<U1EE8>,<U1EE9>);(<U1EEA>,<U1EEB>);/
1770          (<U1EEC>,<U1EED>);(<U1EEE>,<U1EEF>);(<U1EF0>,<U1EF1>);(<U1EF2>,<U1EF3>);/
1771          (<U1EF4>,<U1EF5>);(<U1EF6>,<U1EF7>);(<U1EF8>,<U1EF9>);(<U1F08>,<U1F00>);/
1772          (<U1F09>,<U1F01>);(<U1F0A>,<U1F02>);(<U1F0B>,<U1F03>);(<U1F0C>,<U1F04>);/
1773          (<U1F0D>,<U1F05>);(<U1F0E>,<U1F06>);(<U1F0F>,<U1F07>);(<U1F18>,<U1F10>);/
1774          (<U1F19>,<U1F11>);(<U1F1A>,<U1F12>);(<U1F1B>,<U1F13>);(<U1F1C>,<U1F14>);/
1775          (<U1F1D>,<U1F15>);(<U1F28>,<U1F20>);(<U1F29>,<U1F21>);(<U1F2A>,<U1F22>);/
1776          (<U1F2B>,<U1F23>);(<U1F2C>,<U1F24>);(<U1F2D>,<U1F25>);(<U1F2E>,<U1F26>);/
1777          (<U1F2F>,<U1F27>);(<U1F38>,<U1F30>);(<U1F39>,<U1F31>);(<U1F3A>,<U1F32>);/
1778          (<U1F3B>,<U1F33>);(<U1F3C>,<U1F34>);(<U1F3D>,<U1F35>);(<U1F3E>,<U1F36>);/
1779          (<U1F3F>,<U1F37>);(<U1F48>,<U1F40>);(<U1F49>,<U1F41>);(<U1F4A>,<U1F42>);/
1780          (<U1F4B>,<U1F43>);(<U1F4C>,<U1F44>);(<U1F4D>,<U1F45>);(<U1F59>,<U1F51>);/
1781          (<U1F5B>,<U1F53>);(<U1F5D>,<U1F55>);(<U1F5F>,<U1F57>);(<U1F68>,<U1F60>);/
1782          (<U1F69>,<U1F61>);(<U1F6A>,<U1F62>);(<U1F6B>,<U1F63>);(<U1F6C>,<U1F64>);/
1783          (<U1F6D>,<U1F65>);(<U1F6E>,<U1F66>);(<U1F6F>,<U1F67>);(<U1FBA>,<U1F70>);/
1784          (<U1FBB>,<U1F71>);(<U1FC8>,<U1F72>);(<U1FC9>,<U1F73>);(<U1FCA>,<U1F74>);/
1785          (<U1FCB>,<U1F75>);(<U1FDA>,<U1F76>);(<U1FDB>,<U1F77>);(<U1FF8>,<U1F78>);/
1786          (<U1FF9>,<U1F79>);(<U1FEA>,<U1F7A>);(<U1FEB>,<U1F7B>);(<U1FFA>,<U1F7C>);/
1787          (<U1FFB>,<U1F7D>);(<U1F88>,<U1F80>);(<U1F89>,<U1F81>);(<U1F8A>,<U1F82>);/
1788          (<U1F8B>,<U1F83>);(<U1F8C>,<U1F84>);(<U1F8D>,<U1F85>);(<U1F8E>,<U1F86>);/
1789          (<U1F8F>,<U1F87>);(<U1F98>,<U1F90>);(<U1F99>,<U1F91>);(<U1F9A>,<U1F92>);/
1790          (<U1F9B>,<U1F93>);(<U1F9C>,<U1F94>);(<U1F9D>,<U1F95>);(<U1F9E>,<U1F96>);/
1791          (<U1F9F>,<U1F97>);(<U1FA8>,<U1FA0>);(<U1FA9>,<U1FA1>);(<U1FAA>,<U1FA2>);/
```

```
1792        (<U1FAB>,<U1FA3>);(<U1FAC>,<U1FA4>);(<U1FAD>,<U1FA5>);(<U1FAE>,<U1FA6>);/
1793        (<U1FAF>,<U1FA7>);(<U1FB8>,<U1FB0>);(<U1FB9>,<U1FB1>);(<U1FBC>,<U1FB3>);/
1794        (<U1FCC>,<U1FC3>);(<U1FD8>,<U1FD0>);(<U1FD9>,<U1FD1>);(<U1FE8>,<U1FE0>);/
1795        (<U1FE9>,<U1FE1>);(<U1FEC>,<U1FE5>);(<U1FFC>,<U1FF3>)
1796    %
1797    % The "combining" class reflects ISO/IEC 10646-1 annex B.1
1798    % That is, all combining characters (level 2+3).
1799    class        "combining"; /
1800        <U0300>..<U036F>; <U20D0>..<U20FF>; <UFE20>..<UFE2F>;/
1801        <U0483>..<U0486>;<U0591>..<U05A1>;<U05A3>..<U05B9>;/
1802        <U05BB>..<U05BD>;<U05BF>;<U05C1>;<U05C2>;<U05C4>;<U064B>..<U0652>;<U0670>;/
1803        <U06D7>..<U06E4>;<U06E7>;<U06E8>;<U06EA>..<U06ED>;<U0901>..<U0903>;<U093C>;/
1804        <U093E>..<U094D>;<U0951>..<U0954>;<U0962>;<U0963>;<U0981>..<U0983>;<U09BC>;/
1805        <U09BE>..<U09C4>;<U09C7>;<U09C8>;<U09CB>..<U09CD>;<U09D7>;<U09E2>;<U09E3>;/
1806        <U0A02>;<U0A3C>;<U0A3E>..<U0A42>;<U0A47>;<U0A48>;<U0A4B>..<U0A4D>;/
1807        <U0A70>;<U0A71>;<U0A81>..<U0A83>;<U0ABC>;<U0ABE>..<U0AC5>;<U0AC7>..<U0AC9>;/
1808        <U0ACB>..<U0ACD>;<U0B01>..<U0B03>;<U0B3C>;<U0B3E>..<U0B43>;<U0B47>;<U0B48>;/
1809        <U0B4B>..<U0B4D>;<U0B56>;<U0B57>;<U0B82>;<U0B83>;<U0BBE>..<U0BC2>;/
1810        <U0BC6>..<U0BC8>;<U0BCA>..<U0BCD>;<U0BD7>;<U0C01>..<U0C03>;<U0C3E>..<U0C44>;/
1811        <U0C46>..<U0C48>;<U0C4A>..<U0C4D>;<U0C55>;<U0C56>;<U0C82>;<U0C83>;/
1812        <U0CBE>..<U0CC4>;<U0CC6>..<U0CC8>;<U0CCA>..<U0CCD>;<U0CD5>;<U0CD6>;/
1813        <U0D02>;<U0D03>;<U0D3E>..<U0D43>;<U0D46>..<U0D48>;<U0D4A>..<U0D4D>;<U0D57>;/
1814        <U0E31>;<U0E34>..<U0E3A>;<U0E47>..<U0E4E>;<U0EB1>;<U0EB4>..<U0EB9>;/
1815        <U0EBB>;<U0EBC>;<U0EC8>..<U0ECD>;<U0F18>;<U0F19>;<U0F35>;<U0F37>;<U0F39>;/
1816        <U0F3E>;<U0F3F>;<U0F71>..<U0F84>;<U0F86>..<U0F89>;<U0F8B>;<U0F90>..<U0F95>;/
1817        <U0F97>;<U0F99>..<U0FAD>;<U0FB1>..<U0FB7>;<U0FB9>;<U302A>..<U302F>;/
1818        <U3099>;<U309A>;<UFB1E>
1819    %
1820    % The "combining_level3" class reflects ISO/IEC 10646-1 annex B.2
1821    % That is, combining characters of level 3.
1822    class            "combining_level3";        /
1823        <U0300>..<U036F>;<U20D0>..<U20FF>;<U1100>..<U11FF>;<UFE20>..<UFE2F>;/
1824        <U0483>..<U0486>;<U0591>..<U05A1>;<U05A3>..<U05AE>;<U05C4>;/
1825        <U05AF>;<U093C>;<U0953>;<U0954>;<U09BC>;<U09D7>;<U0A3C>;/
1826        <U0A70>;<U0A71>;<U0ABC>;<U0B3C>;<U0B56>;<U0B57>;<U0BD7>;<U0C55>;<U0C56>;/
1827        <U0CD5>;<U0CD6>;<U0D57>;<U0F39>;<U302A>..<U302F>;<U3099>;<U309A>
1828    %
1829
1830    END LC_CTYPE
```

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 shall assign such values, but that ordering of strings using the resultant collation definition in the FDCC-set shall behave as if such assignment is done and used in the collation process. The collation sequence definition is used by regular expressions, pattern matching, and 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 shall be 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.

1856   (5)   Many-to-Many substitution. A string of one or more characters is substituted by
1857   another string (or an empty string, i.e., the character or characters shall be ignored
1858   for collation purposes).
1859   (6)   Equivalence class definition. Two or more collating elements have the same
1860   collation value (primary weight).
1861   (7)   Ordering by weights. When two strings are compared to determine their relative
1862   order, the two strings are first broken up into a series of collating elements, and
1863   each successive pair of elements are compared according to the relative primary
1864   weights for the elements. If equal, and more than one weight has been assigned,
1865   then the pairs of collating elements are recompared according to the relative
1866   subsequent weights, until either a pair of collating elements compare unequal or the
1867   weights are exhausted.
1868   (8)   Easy reordering of characters. ISO/IEC 14651 has a template for collation
1869   specification that with just a few modifications can be culturally correct for a
1870   specific culture. Here the "reorder-after" keyword gives a convenient way to
1871   modify a FDCC-set template.
1872   (9)   Easy reordering of sections. The template in ISO/IEC 14651 gives an ordering of
1873   the sections that may not be culturally acceptable in certain cultures. The keyword
1874   "reorder-section-after" gives a convenient way to modify the order of sections in a
1875   FDCC-set template.
1876
1877   The following keywords shall be recognized in a collation sequence definition. Some of
1878   them are described in detail in the following subclauses.
1879

1880	**copy**	Specify the name of an existing FDCC-set to be used
1881		as the source for the definition of this category. If
1882		this keyword is specified, only the "reorder-after",
1883		"reorder-end", "reorder-sections-after" and "reorder-
1884		sections-end" keywords may also be specified. The
1885		FDCC-set shall be copied in source form.
1886	**coll_weight_max**	Define as a decimal number the number of collation
1887		levels that an interpreting system needs to support
1888		for this FDCC-set, this value is elsewhere referred as
1889		the COLL_WEIGHT_MAX limit. An interpreting
1890		system shall cater for up to 7 collating levels.
1891	**section-symbol**	Define a section symbol representing a set of
1892		collation order statements. The section is defined
1893		with the "order_start" keyword until the next
1894		"order_start" or "order_end" keyword. This keyword
1895		is optional.
1896	**collating-element**	Define a collating-element symbol representing a
1897		multicharacter collating element. This keyword is
1898		optional.
1899	**collating-symbol**	Define one or more collating symbols for use in
1900		collation order statements. This keyword is optional.
1901	**symbol-equivalence**	Define a collating-symbol to be equivalent to another
1902		defined collating-symbol.
1903	**order_start**	Define collation rules. This statement is followed by
1904		one or more collation order statements, assigning
1905		character collation values and collation weights to

1906	collating elements.
1907 **order_end**	Specify the end of the collation-order statements.
1908 **reorder-after**	Redefine collating rules. Specify after which
1909	collating element the redefinition of collation order
1910	shall take order. This statement is followed by one or
1911	more collation order statements, reassigning character
1912	collation values and collation weights to collating
1913	elements.
1914 **reorder-end**	Specify the end of the "reorder-after" collating order
1915	statements.
1916 **reorder-section-after**	Redefine the order of sections. This statement is
1917	followed by one or more section symbols,
1918	reassigning character collation values and collation
1919	weights to collating elements.
1920 **reorder-section-end**	Specify the end of the "reorder-sections" section
1921	order statements.

1922

4.4.1   Collation statements

1924

1925 The "order_start" and "replace-after" keywords shall be followed by collating statements.
1926 The syntax for the collating statements is

1927

1928   "%s %s;%s;...;%s\n",<collating-identifier>,<weight>,<weight>,...

1929

1930 Each <collating-identifier> shall consist of either a character (in any of the forms defined
1931 in 4.1.1), a <collating-element>, a <collating-symbol>, an ellipsis, or the special symbol
1932 "UNDEFINED". The weights for each of the collation elements determines the character
1933 collation sequence - such that each collation statement does not need to be in collation
1934 order, and weights could be rearranged via for example the "replace-after" keyword. No
1935 character has any specific predetermined placement in the collation sequence. The order in
1936 which collating elements are specified determines the character collation sequence, such
1937 that each collating element shall compare less than the elements following it.

1938

1939 A <collating-element> shall be used to specify multicharacter collating elements, and
1940 indicates that the character sequence specified via the <collating-element> is to be collated
1941 as a unit and in the relative order specified by its place.

1942

1943 A <collating-symbol> shall be used to define a position in the relative order for use in
1944 weights.

1945

1946 The absolute ellipsis symbol ("...") specifies that a sequence of characters shall collate
1947 according to their encoded character values. It shall be interpreted as indicating that all
1948 characters with a coded character set value higher than the value of the character in the
1949 preceding line, and lower than the coded character set value for the character in the
1950 following line, in the current coded character set, shall be placed in the character collation
1951 order between the previous and the following character in ascending order according to
1952 their coded character set values. An initial ellipsis shall be interpreted as if the preceding
1953 line specified the <NUL> character, and a trailing ellipsis as if the following line specified
1954 the highest coded character set value in the current coded character set. An ellipsis shall
1955 be 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 ellipsises ("`..`" or "`....`") specifies a sequence of collating statements. It shall be interpreted as indicating that all characters with symbolic names higher than the symbolic name of the character in the preceding line, and lower than the coded character set value for the character in the following line, shall be placed in the character collation order between the previous and the following character in ascending order.

The symbol "UNDEFINED" shall be interpreted as including all coded character set values not specified explicitly or via the ellipsis or one of the symbolic ellipses symbols. Such characters shall be 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 shall issue a warning message and place such characters at the end of the character collation order.

The optional operands for each collation-element shall be 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 shall behave as if, for each weight level, "IGNORE"d elements are removed. Then each successive pair of elements shall be compared according to the relative weights for the elements. If the two strings compare equal, the process shall be repeated for the next weight level, up to the limit "COLL_WEIGHTS_MAX" of the associated FDCC-set.

Weights shall be 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> shall 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 shall be 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 shall by default be assigned unique weights, equal to the relative order of characters. Characters specified via an explicit or implicit "UNDEFINED" special symbol shall by default be assigned the same primary weight (i.e., belong to the same equivalence class). An ellipsis symbol as a weight shall be interpreted to mean that each character in the sequence shall have 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 shall be treated as an error if the collating element is neither an ellipsis nor the special symbol

2006   "UNDEFINED".

2007

2008   The special keyword "IGNORE" as a weight shall indicate that when strings are compared
2009   using the weights at the level where "IGNORE" is specified, the collating element shall be
2010   ignored; i.e., as if the string did not contain the collating element. In regular expressions
2011   and pattern matching, all characters that are "IGNORE"d in their primary weight form an
2012   equivalence class.

2013

2014   A <comment_character> occurring where the delimiter ";" may occur, terminates the
2015   collating statement.

2016

2017   An empty operand shall be interpreted as the collating-element itself.

2018

2019   For example, the collation statement

2020

2021   <a>   <a>;<a>

2022

2023   is equal to

2024

2025   <a>

2026

2027   An ellipsis (absolute or symbolic) can be used as an operand if the collating-element was
2028   an ellipsis, and shall be interpreted as the value of each character defined by the ellipsis.

2029

2030   Example:
2031

```
2032  collating-element <ch> from "<c><h>"
2033  collating-element <Ch> from "<C><h>"
2034  order_start      forward;backward
2035  UNDEFINED        IGNORE;IGNORE
2036  <LOW>
2037  <space>          <LOW>;<space>
2038  ...              <LOW>;
2039  <a>              <a>;<a>
2040  <a'>             <a>;<a'>
2041  <A>              <a>;<A>
2042  <A'>             <a>;<A'>
2043  <ch>             <ch>;<ch>
2044  <Ch>             <ch>;<Ch>
2045  <s>              <s>;<s>
2046  <ss>             "<s><s>";"<ss><ss>"
2047  order_end
```

2048

2049   This example is interpreted as follows:

2050

2051   (1)   The UNDEFINED means that all characters not specified in this definition (explicitly or via the
2052   ellipsis) shall be ignored.
2053   (2)   <LOW> defines the first collating weight, and thus the lowest weight in this example.
2054   (3)   All characters between <space> and <a> shall have the same primary equivalence class <LOW> and
2055   individual secondary weights based on their ordinal encoded values. (The use of absolute ellipses is
2056   depreciated, but used here to illustrate generic use of ellipses. Symbolic ellipses should be used
2057   instead).
2058   (4)   All characters based on the upper or lowercase character "a" belong to the same primary equivalence
2059   class.
2060   (5)   The multicharacter collating element <c><h> is represented by the collating symbol <ch> and belongs
2061   to the same primary equivalence class as the multicharacter collating element <C><h>.
2062   (6)   The <ss> collating element has two weights on the primary level, and it is in the same primary
2063   equivalence class as two consecutive <s>-es; on the secondary level the collating element has two
2064   weights of the equivalence class <ss>.

2065

2066   **4.4.2   "copy" keyword**

2067

2068   This keyword specifies the name of an existing FDCC-set to be used as the source for the

2069   definition of this category. The syntax is

2071   "copy %s\n", <FDCC-set-name>

2073   The <FDCC-set-name> shall consist of one or more characters (in any of the forms
2074   defined in 4.1.1). If this keyword is specified, only the "reorder-after", "reorder-end",
2075   "reorder-sections-after" and "reorder-sections-end" keywords may also be specified. The
2076   FDCC-set shall be copied in source form.

2078   ### 4.4.3   "col_weight_max" keyword

2080   This keyword defines as a decimal number the number of collation levels that an
2081   interpreting system needs to support, this value is elsewhere referred as the
2082   COLL_WEIGHT_MAX limit. The minimum value is 7. The syntax is

2084   "col_weight_max %d\n", <value>

2086   ### 4.4.4   "section-symbol" keyword

2088   This keyword shall be used to define symbols for use in section related statements; such
2089   as the "order_start", and "reorder-sections-after" keywords and section-reordering
2090   statements. The syntax is

2092   "section-symbol %s\n", <section-symbol>

2094   The <section-symbol> shall be a symbolic name, enclosed between angle brackets (< and
2095   >), and shall not duplicate any symbolic name in the current charmap (if any), or any
2096   other symbolic name defined in this collation definition. A <section-symbol> defined via
2097   this keyword is only defined with the LC_COLLATE category.

2099   Example:
2100   section-symbol <LATIN>
2101   section-symbol <ARABIC>

2103   ### 4.4.5   "collating-element" keyword

2105   In addition to the collating elements in the character set, the collating-element keyword
2106   shall be used to define multicharacter collating elements. The syntax is

2108   "collating-element %s from %s\n",<collating-symbol>,<string>

2110   The <collating-symbol> operand shall be a symbolic name, enclosed between angle
2111   brackets (< and >), and shall not duplicate any symbolic name in the current charmap or
2112   repertoiremap file (if any), or any other symbolic name defined in this collation definition.
2113   The string operand shall be a string of two or more characters that shall collate as an
2114   entity. A <collating-element> defined via this keyword is only defined within the
2115   LC_COLLATE category.

2117   Example with ISO/IEC 10646:
2118   collating-element <ch> from "<c><h>"
2119   collating-element <e-acute> from "<e><combining-acute>"
2120   collating-element <aa> from "<a><a>"

2121 Note: The problem of comparing a fully composed character of ISO/IEC 10646 with a
2122 decomposed representation of the same text is normally handled by the two strings
2123 comparing equal up to level 3 (the case level) of ISO/IEC 14651, but distinguishing the
2124 two at the 4th level.
2125
2126 **4.4.6 "collating-symbol" keyword**
2127
2128 This keyword shall be used to define symbols for use in collation sequence statements;
2129 e.g., between the order_start and the order_end keywords. The syntax is
2130
2131   "collating-symbol %s;%s;...%s\n", <collating-symbol>, <collating-symbol> ...
2132
2133 The <collating-symbol> shall be a symbolic name, enclosed between angle brackets (< and
2134 >), and shall not duplicate any symbolic name in the current charmap (if any), or any
2135 other symbolic name defined in this collation definition. A <collating-symbol> defined via
2136 this keyword is only defined with the LC_COLLATE category. More than one <collating-
2137 symbol> may be defined with one "collating-symbol" keyword, and symbolic ellipses may
2138 be used.
2139
2140   Example:
2141   collating-symbol <CAPITAL>
2142   collating-symbol <HIGH>
2143
2144 **4.4.7 "symbol-equivalence" keyword**
2145
2146 This keyword shall be used to define symbols for use in collation sequence statements;
2147 and assign the same weight as another defined symbol. The syntax is
2148
2149   "symbol-equivalence %s %s\n", <collating-symbol-1>, <collating-symbol-2>
2150
2151 The <collating-symbol-1> and <collating-symbol-2> shall be symbolic names, enclosed
2152 between angle brackets (< and >). <collating-symbol-1> shall not duplicate any symbolic
2153 name in the current charmap (if any), or any other symbolic name defined in this collation
2154 definition. <collating-symbol-2> is defined elsewhere in the LC_COLLATE category as a
2155 collating-symbol. The use of <collating-symbol-2> shall be equivalent to using the
2156 <collating-symbol-2> in the LC_COLLATE category. A <collating-symbol-1> defined via
2157 this keyword is only defined with the LC_COLLATE category.
2158
2159   Example
2160   collating-symbol <CAP>
2161   symbol-equivalence <CAPITAL> <CAP>
2162
2163 **4.4.8 "order_start" keyword**
2164
2165 The "order_start" keyword shall precede collation order entries and also defines the
2166 number of weights for this collation sequence definition, the collation section name and
2167 other collation rules.
2168
2169 The syntax of the "order_start" keyword has two forms:
2170
2171   "order_start %s;%s;...;%s\n", <sort-rules>, <sort-rules> ...

2172   and
2173   "order_start %s;%s;...;%s\n", <section-symbol>, <sort-rules>, <sort-rules> ...
2174

2175   The operands to the order_start keyword are optional. If present, the operands define rules
2176   to be applied when strings are compared. The first operand may be a <section-symbol>
2177   surrounded by "<" and ">" and the set of collating statements following the "order_start"
2178   keyword until the "order_end" keyword are identified with this <section_symbol> or
2179   another "order_start" keyword is encountered. The remaining number of operands define
2180   how many weights each element is assigned; if no operands are present, one forward
2181   operand is assumed. If present, the first operand defines rules to be applied when
2182   comparing strings using the first (primary) weight; the second when comparing strings
2183   using the second weight, and so on. Operands shall be separated by semicolons (;). Each
2184   operand shall consist of one or more collation directives, separated by commas (,). If the
2185   number or operands exceeds the (COLL_WEIGHTS_MAX) limit, a utility parsing the
2186   FDCC-set description shall issue a warning message. The following directives shall be
2187   supported:
2188

2189   **forward**   Specifies that the direction of scanning a part of a string at a given point in a
2190   string is done towards the logical end of the whole string for this weight level.
2191   **backward** Specifies that the direction of scanning a part of a string at a given point in a
2192   string is done towards the logical beginning of the whole string for this weight
2193   level.
2194   **position** Specifies that comparison operations for the weight level will consider the
2195   relative position of non-"IGNORE"d elements in the strings. The string
2196   containing a non-"IGNORE"d element after the fewest IGNOREd collating
2197   elements from the start of the compare shall collate first. If both strings
2198   contain a non-"IGNORE"d character in the same relative position, the collating
2199   values assigned to the elements shall determine the ordering. In case of
2200   equality, subsequent non-IGNOREd characters shall be considered in the same
2201   manner.
2202

2203   The directives "forward" and "backward", and "backward" and "position", are mutually
2204   exclusive at a given level.
2205

2206   Examples:
2207   order_start forward;backward
2208   order_start <CYRILLIC>;forward;forward
2209

2210   If no operands are specified, a single forward operand shall be assumed.
2211

2212

2213   **4.4.9   "order_end" keyword**
2214

2215   The collating order entries shall be terminated with an order_end keyword.
2216

2217   **4.4.10   "reorder-after" keyword**
2218

2219   The "reorder-after" keyword shall be used to specify a modification to a copied collation
2220   specification of an existing FDCC-set. There can be more than one "reorder-after"
2221   statement in a collating specification. The syntax shall be:
2222

2223   "reorder-after %s\n",<collating-symbol>

2224

2225 The <collating-symbol> operand shall be a symbolic name, enclosed between angle

2226 brackets, and shall be present in the source FDCC-set copied via the "copy" keyword.

2227 The "reorder-after" statement is followed by one or more collation statements as described

2228 in the "Collating Order" clause (4.4.5), with the exception that the ellipsis symbol (...)

2229 shall not be used.

2230

2231 Each collation statement reassigns character collation values and collation weights to

2232 collating elements existing in the copied collation specification, by removing the collating

2233 statement from the copied specification, and inserting the collating element in the collating

2234 sequence with the new collation weights after the preceding collating element of the

2235 "reorder-after" specification, the first collating element in the collation sequence being the

2236 <collating-symbol> specified on the "reorder-after" statement.

2237

2238 A "reorder-after" specification is terminated by another "reorder-after" specification or the

2239 "reorder-end" statement.

2240

2241 **4.4.10.1 Example of "reorder-after"**

2242

2243   reorder-after <y8>

2244   <U:>   <Y>;<U:>;<CAPITAL>

2245   <u:>   <Y>;<U:>;<SMALL>

2246   reorder-after <z8>

2247   <AE>   <AE>;<NONE>;<CAPITAL>

2248   <ae>   <AE>;<NONE>;<SMALL>

2249   <A:>   <AE>;<DIAERESIS>;<CAPITAL>

2250   <a:>   <AE>;<DIAERESIS>;<SMALL>

2251   <O/>   <O/>;<NONE>;<CAPITAL>

2252   <o/>   <O/>;<NONE>;<SMALL>

2253   <AA>   <AA>;<NONE>;<CAPITAL>

2254   <aa>   <AA>;<NONE>;<SMALL>

2255   reorder-end

2256

2257 The example is interpreted as follows (using the "i18nrep" repertoiremap):

2258

2259 1.   The collating element <U:> is removed from the copied collating sequence and inserted after <y8> in the

2260   collating sequence with the new weights. The collating element <u:> is removed from the copied collating

2261   sequence and inserted in the resulting collation sequence after <U:> with the new weights. <y8> is used to

2262   indicate the last entry of the <y> letters.

2263

2264 2.   The second "reorder-after" statement terminates the first list of reordering collation identifier entries, and

2265   initiates a second list, rearranging the order and weights for the <AE>, <ae>, <A:>, <a:>, <O/>, and <o/>

2266   collating elements after the <z8> collating symbol in the copied specification. <z8> is used to indicate the

2267   last entry of the <z> letters.

2268

2269 3.   The "reorder-end" statement terminates the second list of reordering entries.

2270

2271 4.   Thus for the original sequence

2272

2273   ... (U u Ü ü) V v W w X x Y y Z z

2274

2275   this example reordering gives

2276

2277   ... U u V v W w X x (Y y Ü ü) Z z (Æ æ Ä ä) Ø ø Å å

2278

2279   where the parenthesis indicate ordering with the same weight on the first level for multiple upper/lowercase

2280   pairs.
2281

4.4.11   "reorder-end" keyword

2283

2284 The "reorder-end" keyword shall specify the end of a list of collating statements, initiated
2285 by the "reorder-after" keyword.

2286

4.4.12   "reorder-sections-after" keyword

2288

2289 The "reorder-sections-after" keyword shall be used to specify a modification to a copied
2290 collation specification of an existing FDCC-set. The "reorder-sections-after" statement is
2291 followed by one or more statements consisting of section reordering statements.

2292

4.4.12.1   section reordering statements

2294

2295 The section reordering statements rearranges the set of collating entries and changes
2296 sorting rules for the set of collating entries identified by a section symbol in a preceding
2297 "order_start" statement. Each section reorder statement has the syntax:

2298

2299   "%s %s;...%s\n", <section-symbol>, <sort-rules>, <sort-rules> ...

2300

2301 The <section-symbol> identifies the set of collating entries, and shall be defined via a
2302 "section-symbol" keyword.

2303

2304 The <sort-rules> are as described for the "order_start" keyword. Specified <sort-rules>
2305 replace the specification for the ordering of the section given on the "order_start"
2306 statement identified by the <section-symbol>. The <sort-rules> are optional and <sort-
2307 rules> not to be changed may be given by empty specifications.

2308

2309 The order of the section reordering statements rearranges the assignment of collation
2310 entries for the sets of collation entries identified by the <section-symbols> to the order
2311 that the <section-symbols> occur after the "reorder-sections-after" statement.

2312

2313 The section reordering statements are terminated by a "reorder-sections-end" statement.

2314

4.4.12.2   Example of section reordering

```
copy "i18n"
reorder-sections-after <DIGITS>
<ARABIC>
<LATIN> forward;backward;forward;forward,position
reorder-sections-end
```

2322

2323 This example is interpreted as follows: The LC_COLLATE category of the "i18n" FDCC-set is copied. Then a
2324 reordering of all collating statements for the sections <ARABIC> and <LATIN> is done, leaving the rest of the
2325 sections as they were in the "i18n" FDCC-set. The <ARABIC> section is placed immediately after the <DIGITS>
2326 section, and the <LATIN> section immediately following the <ARABIC> section. The ordering rules are kept as
2327 they were in the "i18n" FDCC-set, while the <LATIN> section gets new ordering rules as indicated. The
2328 "reorder-sections-end" keyword terminates the section reordering statements.

2329

4.4.13   "reorder-sections-end" keyword

2331

2332 The "reorder-sections-end" keyword shall specify the end of a list of section symbols,
2333 initiated by the "reorder-sections-after" keyword.

2334   **4.4.14   "i18n" LC_COLLATE category**
2335
2336   The "i18n" LC_COLLATE category is defined as the following, which includes the
2337   tailorable template in ISO/IEC 14651.
2338

```
2339    LC_COLLATE
2340
2341    % Case collating symbols
2342    collating-symbol <RES-1>
2343    collating-symbol <BLK>
2344    collating-symbol <MIN>      % SMALL
2345    collating-symbol <WIDE>     % WIDE
2346    collating-symbol <COMPAT>
2347    collating-symbol <FONT>
2348    collating-symbol <CIRCLE>
2349    collating-symbol <RES-2>
2350    collating-symbol <CAP>      % CAPITAL
2351    collating-symbol <WIDECAP>
2352    collating-symbol <COMPATCAP>
2353    collating-symbol <FONTCAP>
2354    collating-symbol <CIRCLECAP>
2355    collating-symbol <HIRA-SMALL>
2356    collating-symbol <HIRA>
2357    collating-symbol <SMALL>
2358    collating-symbol <SMALL-NARROW>
2359    collating-symbol <KATA>
2360    collating-symbol <NARROW>
2361    collating-symbol <CIRCLE-KATA>
2362    collating-symbol <MNN>
2363    collating-symbol <MNS>
2364    collating-symbol <VERTICAL>
2365    % Arabic forms
2366    collating-symbol <AINI>
2367    collating-symbol <AMED>
2368    collating-symbol <AFIN>
2369    collating-symbol <AISO>
2370    %
2371    collating-symbol <NOBREAK>
2372    collating-symbol <SQUARED>
2373    collating-symbol <SQUAREDCAP>
2374    collating-symbol <FRACTION>
2375    collating-symbol <BLANK>
2376    collating-symbol <CAPITAL-SMALL>
2377    collating-symbol <SMALL-CAPITAL>
2378    collating-symbol <BOTH>
2379    % accents
2380    collating-symbol <LOWLINE>  % LOW LINE
2381    collating-symbol <MACRO>    % MACRON
2382    collating-symbol <OBLIK>    % STROKE
2383    collating-symbol <AIGUT>    % ACUTE ACCENT
2384    collating-symbol <GRAVE>    % GRAVE ACCENT
2385    collating-symbol <BREVE>    % BREVE
2386    collating-symbol <CIRCF>    % CIRCUMFLEX ACCENT
2387    collating-symbol <CARON>    % CARON
2388    collating-symbol <CRCLE>    % RING ABOVE
2389    collating-symbol <TREMA>    % DIAERESIS
2390    collating-symbol <2AIGU>    % DOUBLE ACUTE ACCENT
2391    collating-symbol <TILDE>    % TILDE
2392    collating-symbol <POINT>    % DOT ABOVE
2393    collating-symbol <CEDIL>    % CEDILLA
2394    collating-symbol <OGONK>    % OGONEK
2395    collating-symbol <OVERLINE>  % OVERLINE
2396    collating-symbol <CROOK>    % HOOK ABOVE
2397    collating-symbol <TONOS>    % VERTICAL LINE ABOVE
2398    collating-symbol <D030E>    % DOUBLE VERTICAL LINE ABOVE
2399    collating-symbol <2GRAV>    % DOUBLE GRAVE ACCENT
2400    collating-symbol <D0310>    % CANDRABINDU
2401    collating-symbol <BREVR>    % INVERTED BREVE
2402    collating-symbol <D0312>    % TURNED COMMA ABOVE
2403    collating-symbol <PSILI>    % COMMA ABOVE
2404    collating-symbol <DASIA>    % REVERSED COMMA ABOVE
2405    collating-symbol <D0315>    % COMMA ABOVE RIGHT
2406    collating-symbol <D0316>    % GRAVE ACCENT BELOW
2407    collating-symbol <D0317>    % ACUTE ACCENT BELOW
2408    collating-symbol <D0318>    % LEFT TACK BELOW
2409    collating-symbol <D0319>    % RIGHT TACK BELOW
```

```
2410        collating-symbol <D031A> % LEFT ANGLE ABOVE
2411        collating-symbol <HORNU> % HORN
2412        collating-symbol <D031C> % LEFT HALF RING BELOW
2413        collating-symbol <D031D> % UP TACK BELOW
2414        collating-symbol <D031E> % DOWN TACK BELOW
2415        collating-symbol <D031F> % PLUS SIGN BELOW
2416        collating-symbol <D0320> % MINUS SIGN BELOW
2417        collating-symbol <PALCR> % PALATALIZED HOOK BELOW
2418        collating-symbol <RETCR> % RETROFLEX HOOK BELOW
2419        collating-symbol <POINS> % DOT BELOW
2420        collating-symbol <TREMS> % DIAERESIS BELOW
2421        collating-symbol <CRCLS> % RING BELOW
2422        collating-symbol <COMMS> % COMMA BELOW
2423        collating-symbol <D0329> % VERTICAL LINE BELOW
2424        collating-symbol <D032A> % BRIDGE BELOW
2425        collating-symbol <D032B> % INVERTED DOUBLE ARCH BELOW
2426        collating-symbol <D032C> % CARON BELOW
2427        collating-symbol <CIRCS> % CIRCUMFLEX ACCENT BELOW
2428        collating-symbol <BREVS> % BREVE BELOW
2429        collating-symbol <D032F> % INVERTED BREVE BELOW
2430        collating-symbol <TILDS> % TILDE BELOW
2431        collating-symbol <MACRS> % MACRON BELOW
2432        collating-symbol <D0333> % DOUBLE LOW LINE
2433        collating-symbol <TILDX> % TILDE OVERLAY
2434        collating-symbol <BARRE> % SHORT STROKE OVERLAY
2435        collating-symbol <D0336> % LONG STROKE OVERLAY
2436        collating-symbol <D0337> % SHORT SOLIDUS OVERLAY
2437        collating-symbol <CRCL2> % RIGHT HALF RING BELOW
2438        collating-symbol <D033A> % INVERTED BRIDGE BELOW
2439        collating-symbol <D033B> % SQUARE BELOW
2440        collating-symbol <D033C> % SEAGULL BELOW
2441        collating-symbol <D033D> % X ABOVE
2442        collating-symbol <D033E> % VERTICAL TILDE
2443        collating-symbol <D033F> % DOUBLE OVERLINE
2444        collating-symbol <PERIS> % GREEK PERISPOMENI
2445        collating-symbol <YPOGE> % GREEK YPOGEGRAMMENI
2446        collating-symbol <D0360> % DOUBLE TILDE
2447        collating-symbol <D0361> % DOUBLE INVERTED BREVE
2448        collating-symbol <DFE20> % LIGATURE LEFT HALF
2449        collating-symbol <DFE21> % LIGATURE RIGHT HALF
2450        collating-symbol <DFE22> % DOUBLE TILDE LEFT HALF
2451        collating-symbol <DFE23> % DOUBLE TILDE RIGHT HALF
2452        collating-symbol <D0483> % CYRILLIC TITLO
2453        collating-symbol <D0484> % CYRILLIC PALATALIZATION
2454        collating-symbol <D0485> % CYRILLIC DASIA PNEUMATA
2455        collating-symbol <D0486> % CYRILLIC PSILI PNEUMATA
2456        collating-symbol <SHEVA> % HEBREW POINT SHEVA
2457        collating-symbol <HTFSG> % HEBREW POINT HATAF SEGOL
2458        collating-symbol <HTFPT> % HEBREW POINT HATAF PATAH
2459        collating-symbol <HTFQM> % HEBREW POINT HATAF QAMATS
2460        collating-symbol <HIRIQ> % HEBREW POINT HIRIQ
2461        collating-symbol <TSERE> % HEBREW POINT TSERE
2462        collating-symbol <SEGOL> % HEBREW POINT SEGOL
2463        collating-symbol <PATAH> % HEBREW POINT PATAH
2464        collating-symbol <QAMAT> % HEBREW POINT QAMATS
2465        collating-symbol <HOLAM> % HEBREW POINT HOLAM
2466        collating-symbol <QUBUT> % HEBREW POINT QUBUTS
2467        collating-symbol <DAGES> % HEBREW POINT DAGESH OR MAPIQ
2468        collating-symbol <RAPHE> % HEBREW POINT RAFE
2469        collating-symbol <SHINP> % HEBREW POINT SHIN DOT
2470        collating-symbol <SINPT> % HEBREW POINT SIN DOT
2471        collating-symbol <VARIKA>  % HEBREW POINT JUDEO-SPANISH VARIKA
2472        collating-symbol <FATHATAN>  % ARABIC FATHATAN
2473        collating-symbol <DAMMATAN>  % ARABIC DAMMATAN
2474        collating-symbol <KASRATAN>  % ARABIC KASRATAN
2475        collating-symbol <FATHA> % ARABIC FATHA
2476        collating-symbol <DAMMA> % ARABIC DAMMA
2477        collating-symbol <KASRA> % ARABIC KASRA
2478        collating-symbol <SHADDA>  % ARABIC SHADDA
2479        collating-symbol <SUKUN> % ARABIC SUKUN
2480        collating-symbol <SUPERALEF>  % ARABIC LETTER SUPERSCRIPT ALEF
2481        collating-symbol <D06D6> % ARABIC SMALL HIGH LIGATURE SAD WITH LAM WITH ALEF MAKSURA
2482        collating-symbol <D06D7> % ARABIC SMALL HIGH LIGATURE QAF WITH LAM WITH ALEF MAKSURA
2483        collating-symbol <D06D8> % ARABIC SMALL HIGH MEEM INITIAL FORM
2484        collating-symbol <D06D9> % ARABIC SMALL HIGH LAM ALEF
2485        collating-symbol <D06DA> % ARABIC SMALL HIGH JEEM
2486        collating-symbol <D06DB> % ARABIC SMALL HIGH THREE DOTS
2487        collating-symbol <D06DC> % ARABIC SMALL HIGH SEEN
```

```
2488      collating-symbol <D06E1>   % ARABIC SMALL HIGH DOTLESS HEAD OF KHAH
2489      collating-symbol <D06E2>   % ARABIC SMALL HIGH MEEM ISOLATED FORM
2490      collating-symbol <D06E3>   % ARABIC SMALL LOW SEEN
2491      collating-symbol <AMADD>   % ARABIC SMALL HIGH MADDA
2492      collating-symbol <D06E7>   % ARABIC SMALL HIGH YEH
2493      collating-symbol <D06E8>   % ARABIC SMALL HIGH NOON
2494      collating-symbol <D06ED>   % ARABIC SMALL LOW MEEM
2495      collating-symbol <D093C>   % DEVANAGARI SIGN NUKTA
2496      collating-symbol <D0951>   % DEVANAGARI STRESS SIGN UDATTA
2497      collating-symbol <D0952>   % DEVANAGARI STRESS SIGN ANUDATTA
2498      collating-symbol <D0953>   % DEVANAGARI GRAVE ACCENT
2499      collating-symbol <D0954>   % DEVANAGARI ACUTE ACCENT
2500      collating-symbol <D09BC>   % BENGALI SIGN NUKTA
2501      collating-symbol <D0A3C>   % GURMUKHI SIGN NUKTA
2502      collating-symbol <D0ABC>   % GUJARATI SIGN NUKTA
2503      collating-symbol <D0B3C>   % ORIYA SIGN NUKTA
2504      collating-symbol <D0E48>   % THAI CHARACTER MAI EK
2505      collating-symbol <D0E49>   % THAI CHARACTER MAI THO
2506      collating-symbol <D0E4A>   % THAI CHARACTER MAI TRI
2507      collating-symbol <D0E4B>   % THAI CHARACTER MAI CHATTAWA
2508      collating-symbol <D0EC8>   % LAO TONE MAI EK
2509      collating-symbol <D0EC9>   % LAO TONE MAI THO
2510      collating-symbol <D0ECA>   % LAO TONE MAI TI
2511      collating-symbol <D0ECB>   % LAO TONE MAI CATAWA
2512      collating-symbol <D0F39>   % TIBETAN MARK TSA -PHRU
2513      collating-symbol <D0F3E>   % TIBETAN SIGN YAR TSHES
2514      collating-symbol <D0F3F>   % TIBETAN SIGN MAR TSHES
2515      collating-symbol <D302A>   % IDEOGRAPHIC LEVEL TONE MARK
2516      collating-symbol <D302B>   % IDEOGRAPHIC RISING TONE MARK
2517      collating-symbol <D302C>   % IDEOGRAPHIC DEPARTING TONE MARK
2518      collating-symbol <D302D>   % IDEOGRAPHIC ENTERING TONE MARK
2519      collating-symbol <D302E>   % HANGUL SINGLE DOT TONE MARK
2520      collating-symbol <D302F>   % HANGUL DOUBLE DOT TONE MARK
2521      collating-symbol <KNVCE>   % KATAKANA-HIRAGANA VOICED SOUND MARK
2522      collating-symbol <KNSMV>   % KATAKANA-HIRAGANA SEMI-VOICED SOUND MARK
2523      collating-symbol <D20D0>   % LEFT HARPOON ABOVE
2524      collating-symbol <D20D1>   % RIGHT HARPOON ABOVE
2525      collating-symbol <D20D2>   % LONG VERTICAL LINE OVERLAY
2526      collating-symbol <D20D3>   % SHORT VERTICAL LINE OVERLAY
2527      collating-symbol <D20D4>   % ANTICLOCKWISE ARROW ABOVE
2528      collating-symbol <D20D5>   % CLOCKWISE ARROW ABOVE
2529      collating-symbol <D20D6>   % LEFT ARROW ABOVE
2530      collating-symbol <D20D7>   % RIGHT ARROW ABOVE
2531      collating-symbol <D20D8>   % RING OVERLAY
2532      collating-symbol <D20D9>   % CLOCKWISE RING OVERLAY
2533      collating-symbol <D20DA>   % ANTICLOCKWISE RING OVERLAY
2534      collating-symbol <D20DB>   % THREE DOTS ABOVE
2535      collating-symbol <D20DC>   % FOUR DOTS ABOVE
2536      collating-symbol <D20DD>   % ENCLOSING CIRCLE
2537      collating-symbol <D20DE>   % ENCLOSING SQUARE
2538      collating-symbol <D20DF>   % ENCLOSING DIAMOND
2539      collating-symbol <D20E0>   % ENCLOSING CIRCLE BACKSLASH
2540      collating-symbol <D20E1>   % LEFT RIGHT ARROW ABOVE
2541      collating-symbol <NEGATIVE>
2542      collating-symbol <SANSSERIF>
2543      collating-symbol <NEGSANSSERIF>
2544      collating-symbol <ARABIC>
2545      collating-symbol <EXTARABIC>
2546      collating-symbol <NAGAR>
2547      collating-symbol <BENGL>
2548      collating-symbol <BENGALINUMERATOR>
2549      collating-symbol <GURMU>
2550      collating-symbol <GUJAR>
2551      collating-symbol <ORIYA>
2552      collating-symbol <TAMIL>
2553      collating-symbol <TELGU>
2554      collating-symbol <KNNDA>
2555      collating-symbol <MALAY>
2556      collating-symbol <SINHALA>
2557      collating-symbol <THAII>
2558      collating-symbol <LAAOO>
2559      collating-symbol <BODKA>
2560      collating-symbol <CJKVS>
2561      collating-symbol <S0200>..<S1100> % 0x0200..0x1100
2562
2563      collating-symbol <S4E00>..<S9FA5> % Symbols for Han
2564
2565      collating-symbol <SAC00>..<SD7A3> % Symbols for Hangul
2566
```

```
2567    collating-symbol <SFA0E>..<SFA29> % Symbols for Compatibility Han
2568
2569    % equivalences
2570    symbol-equivalence <NONE>                <BLANK>
2571    symbol-equivalence <CAPITAL>             <CAP>
2572    symbol-equivalence <MACRON>              <MACRO>
2573    symbol-equivalence <STROKE>              <OBLIK>
2574    symbol-equivalence <ACUTE>               <AIGUT>
2575    symbol-equivalence <CIRCUMFLEX>          <CIRCF>
2576    symbol-equivalence <RING>                <CRCLE>
2577    symbol-equivalence <DIAERESIS>           <TREMA>
2578    symbol-equivalence <DOT>                 <POINT>
2579    symbol-equivalence <CEDILLA>             <CEDIL>
2580    symbol-equivalence <OGONEK>              <OGONK>
2581    symbol-equivalence <HOOK>                <CROOK>
2582    symbol-equivalence <HORN>                <HORNU>
2583    symbol-equivalence <DOT-BELOW>           <POINS>
2584
2585    order_start <Latin>;forward;backward;forward;forward,position
2586
2587    % Copy the template from ISO/IEC 14651
2588    copy "iso14651_t1"
2589
2590    order_end
2591
2592    END LC_COLLATE
2593
```

4.5   LC_MONETARY

The LC_MONETARY category defines the rules and symbols that shall be used to format
monetary numeric information. The operands are strings. For some keywords, the strings
can contain only integers. More than one set of monetary values may be provided, and for
each set a period of validity and conversion rate may be given. Keywords that are not
provided, string values set to the empty string "", or integer keywords set to -1, shall be
used to indicate that the value is unspecified, and then no default is taken. The following
keywords shall be defined:

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 shall be specified.
valid_from	One or more strings separated by semicolons, representing a Gregorian date in the form "YYYYMMDD" according to ISO 8601, specifying the beginning date (inclusive from the beginning of day local time) of the validity of a currency. The position of the string in the list corresponds to the position of operands in other keywords in the LC_MONETARY category. The currencies should be ordered in terms of validity dates, and for each validity period with the currency that the amounts are stored in first. If not specified, it is taken to be the beginning of time.
valid_to	One or more strings separated by semicolons, representing a Gregorian date in the form "YYYYMMDD" according to ISO 8601, specifying the end date (inclusive to the end of day local time) of the validity of a currency. If not specified, it is taken to be the end of time.
conversion_rate	one or more pairs of integers separated by a <semicolon> specifying the fixed conversion rate between the current currency (determined by the parameter number) and the first currency that is valid, determined by a date provided by the

2626		application. If the currency is not the first valid currency for
2627		the period in question, the first integer is for multiplying the
2628		first valid currency, and the second for dividing this result to
2629		get the amount in the current currency. The currency to be
2630		the current currency is selected by the application from the
2631		date applicable and the currency number (first, second, third
2632		etc valid currency at that date); and whether domestic or
2633		international formatting is used is also determined by the
2634		application. Each pair of integers are separated by a <slash>.
2635		The default value is "1/100". This keyword is optional.
2636	**int_curr_symbol**	One or more strings separated by semicolons that shall be
2637		used as the international currency symbols. Each operand
2638		shall be a four character string, with the first three characters
2639		containing the alphabetic international currency symbol in
2640		accordance with those specified in ISO 4217, *Codes for the*
2641		*representation of currencies and funds*. The fourth character
2642		shall be the character used to separate the international
2643		currency symbol from the monetary quantity. The keyword
2644		shall be specified, unless the "copy" keyword is used.
2645	**currency_symbol**	One or more strings separated by semicolons that shall be
2646		used as the local currency symbol.
2647	**mon_decimal_point**	The operand is a string containing the symbol that shall be
2648		used as the decimal delimiter in monetary formatted
2649		quantities. In contexts where other standards limit the
2650		"mon_decimal_point" to a single byte, the result of
2651		specifying a multibyte operand is unspecified. The keyword
2652		shall be specified, unless the "copy" keyword is used.
2653	**mon_thousands_sep**	The operand is a string containing the symbol that shall be
2654		used as a separator for groups of digits to the left of the
2655		decimal delimiter in formatted monetary quantities. In
2656		contexts where other standards limit the
2657		"mon_thousands_sep" to a single byte, the result of speci-
2658		fying a multibyte operand is unspecified. The keyword shall
2659		be specified, unless the "copy" keyword is used.
2660	**mon_grouping**	Define the size of each group of digits in formatted
2661		monetary quantities. The operand is a sequence of integers
2662		separated by semicolons. Each integer specifies the number
2663		of digits in each group, with the initial integer defining the
2664		size of the group immediately preceding the decimal
2665		delimiter, and the following integers defining the preceding
2666		groups. If the last integer is not -1, then the size of the
2667		previous group (if any) shall be repeatedly used for the
2668		remainder of the digits. If the last integer is -1, then no
2669		further grouping shall be performed. The keyword shall be
2670		specified, unless the "copy" keyword is used.
2671	**positive_sign**	A string that shall be used to indicate a nonnegative-valued
2672		formatted monetary quantity. The keyword shall be specified,
2673		unless the "copy" keyword is used.
2674	**negative_sign**	A string that shall be used to indicate a negative-valued
2675		formatted monetary quantity. The keyword shall be specified,

2676		unless the "copy" keyword is used.
2677	**int_frac_digits**	One or more integers separated by semicolons, representing
2678		the number of fractional digits (those to the right of the
2679		decimal delimiter) to be written in a formatted monetary
2680		quantity using int_curr_symbol. The keyword shall be
2681		specified, unless the "copy" keyword is used.
2682	**frac_digits**	One or more integers separated by semicolons, representing
2683		the number of fractional digits (those to the right of the
2684		decimal delimiter) to be written in a formatted monetary
2685		quantity using "currency_symbol". The keyword shall be
2686		specified, unless the "copy" keyword is used.
2687	**p_cs_precedes**	One or more integers separated by semicolons, set to 1 if the
2688		"currency_symbol" precedes the value for a nonnegative
2689		formatted monetary quantity, and set to 0 if the symbol
2690		succeeds the value. The keyword shall be specified, unless
2691		the "copy" keyword is used.
2692	**p_sep_by_space**	One or more integers separated by semicolons, set to 0 if no
2693		space separates the "currency_symbol" from the value for a
2694		nonnegative formatted monetary quantity, set to 1 if a space
2695		separates the symbol from the value, and set to 2 if a space
2696		separates the symbol and the sign string, if adjacent. The
2697		keyword shall be specified, unless the "copy" keyword is
2698		used.
2699	**n_cs_precedes**	One or more integers separated by semicolons, set to 1 if the
2700		"currency_symbol" precedes the value for a negative
2701		formatted monetary quantity, and set to 0 if the symbol
2702		succeeds the value. The keyword shall be specified, unless
2703		the "copy" keyword is used.
2704	**n_sep_by_space**	One or more integers separated by semicolons, set to 0 if no
2705		space separates the "currency_symbol" from the value for a
2706		negative formatted monetary quantity, set to 1 if a space
2707		separates the symbol from the value, and set to 2 if a space
2708		separates the symbol and the sign string, if adjacent. The
2709		keyword shall be specified, unless the "copy" keyword is
2710		used.
2711	**int_p_cs_precedes**	One or more integers separated by semicolons; set to 1 if the
2712		"int_curr_symbol" precedes the value for a nonnegative
2713		formatted monetary quantity, and set to 0 if the symbol
2714		succeeds the value. If not specified, the value of
2715		"p_cs_precedes" is taken.
2716	**int_p_sep_by_space**	One or more integers separated by semicolons; set to 0 if no
2717		space separates the "int_curr_symbol" from the value for a
2718		nonnegative formatted monetary quantity, set to 1 if a space
2719		separates the symbol from the value, and set to 2 if a space
2720		separates the symbol and the sign string, if adjacent. If not
2721		specified, the value of "p_sep_by_space" is taken.
2722	**int_n_cs_precedes**	One or more integers separated by semicolons; set to 1 if the
2723		"int_curr_symbol" precedes the value for a negative
2724		formatted monetary quantity, and set to 0 if the symbol
2725		succeeds the value. If not specified, the value of

2726		"n_cs_precedes" is taken.
2727	**int_n_sep_by_space**	One or more integers separated by semicolons; set to 0 if no
2728		space separates the "int_curr_symbol" from the value for a
2729		negative formatted monetary quantity, set to 1 if a space
2730		separates the symbol from the value, and set to 2 if a space
2731		separates the symbol and the sign string, if adjacent. If not
2732		specified, the value of "n_sep_by_space" is taken.
2733	**p_sign_posn**	One or more integers separated by semicolons, set to a value
2734		indicating the positioning of the "positive_sign" for a
2735		nonnegative formatted monetary quantity using the
2736		"currency_symbol". The following integer values shall be
2737		defined:
2738		
2739		0 Parentheses enclose the quantity and the
2740		"currency_symbol".
2741		1 The sign string precedes the quantity and the
2742		"currency_symbol".
2743		2 The sign string succeeds the quantity and the
2744		"currency_symbol".
2745		3 The sign string immediately precedes the
2746		"currency_symbol".
2747		4 The sign string immediately succeeds the
2748		"currency_symbol".
2749		The keyword shall be specified, unless the "copy" keyword
2750		is used.
2751		
2752	**n_sign_posn**	One or more integers separated by semicolons, set to a value
2753		indicating the positioning of the "negative_sign" for a
2754		negative formatted monetary quantity using the
2755		"currency_symbol". The following integer values shall be
2756		defined:
2757		
2758		0 Parentheses enclose the quantity and the
2759		"currency_symbol".
2760		1 The sign string precedes the quantity and the
2761		"currency_symbol".
2762		2 The sign string succeeds the quantity and the
2763		"currency_symbol".
2764		3 The sign string immediately precedes the
2765		"currency_symbol".
2766		4 The sign string immediately succeeds the
2767		"currency_symbol".
2768		The keyword shall be specified, unless the "copy" keyword
2769		is used.
2770		
2771	**int_p_sign_posn**	One or more integers separated by semicolons, set to a value
2772		indicating the positioning of the "positive_sign" for a
2773		nonnegative formatted international monetary quantity. The
2774		following integer values shall be defined:
2775		

2776		0	Parentheses enclose the quantity and the
2777			"int_curr_symbol".
2778		1	The sign string precedes the quantity and the
2779			"int_curr_symbol".
2780		2	The sign string succeeds the quantity and the
2781			"int_curr_symbol".
2782		3	The sign string immediately precedes the
2783			"int_curr_symbol".
2784		4	The sign string immediately succeeds the
2785			"int_curr_symbol".

2786   If no "int_p_sign_posn" is present the value of the
2787   "p_sign_posn" is taken.

2788

2789 **int_n_sign_posn**   One or more integers separated by semicolons, set to a value
2790   indicating the positioning of the "negative_sign" for a
2791   negative formatted international monetary quantity. The
2792   following integer values shall be defined:

2793

2794		0	Parentheses enclose the quantity and the
2795			"int_curr_symbol".
2796		1	The sign string precedes the quantity and the
2797			"int_curr_symbol".
2798		2	The sign string succeeds the quantity and the
2799			"int_curr_symbol".
2800		3	The sign string immediately precedes the
2801			"int_curr_symbol".
2802		4	The sign string immediately succeeds the
2803			"int_curr_symbol".

2804   If no "int_n_sign_posn" is present the value of the
2805   "n_sign_posn" is taken.

2806

2807 The "i18n" FDCC-set is defined as follows for the LC_MONETARY category.

2808

```
2809    LC_MONETARY
2810    % This is the 14652 i18n fdcc-set definition for
2811    % the LC_MONETARY category.
2812    %
2813    int_curr_symbol      ""
2814    currency_symbol      ""
2815    mon_decimal_point    "<,>"
2816    mon_thousands_sep    ""
2817    mon_grouping         -1
2818    positive_sign        ""
2819    negative_sign        ""
2820    int_frac_digits      -1
2821    frac_digits          -1
2822    p_cs_precedes        -1
2823    p_sep_by_space       -1
2824    n_cs_precedes        -1
2825    n_sep_by_space       -1
2826    p_sign_posn          -1
2827    n_sign_posn          -1
2828    %
2829    END LC_MONETARY
```

2830
2831

2832 ## 4.6 LC_NUMERIC

2833

2834 The LC_NUMERIC category defines the rules and symbols that shall be used to format

2835 nonmonetary numeric information. The operands are strings. For some keywords, the
2836 strings only can contain integers. Keywords that are not provided, string values set to the
2837 empty string (""), or integer keywords set to -1, shall be used to indicate that the value is
2838 unspecified. The following keywords shall be defined:
2839

2840 **copy** Specify the name of an existing FDCC-set to be used as the
2841 source for the definition of this category. If this keyword is
2842 specified, no other keyword shall be specified.
2843 **decimal_point** The operand is a string containing the symbol that shall be used
2844 as the decimal delimiter in numeric, nonmonetary formatted
2845 quantities. This keyword cannot be omitted and cannot be set to
2846 the empty string. In contexts where other standards limit the
2847 decimal point to a single byte, the result of specifying a mul-
2848 tibyte operand is unspecified.
2849 **thousands_sep** The operand is a string containing the symbol that shall be used
2850 as a separator for groups of digits to the left of the decimal
2851 delimiter in numeric, nonmonetary formatted monetary quan-
2852 tities. In contexts where other standards limit the
2853 "thousands_sep" to a single byte, the result of specifying a
2854 multibyte operand is unspecified.
2855 **grouping** Define the size of each group of digits in formatted non-
2856 monetary quantities. The operand is a sequence of integers
2857 separated by semicolons. Each integer specifies the number of
2858 digits in each group, with the initial integer defining the size of
2859 the group immediately preceding the decimal delimiter, and the
2860 following integers defining the preceding groups. If the last
2861 integer is not -1, then the size of the previous group (if any)
2862 shall be repeatedly used for the remainder of the digits. If the
2863 last integer is -1, then no further grouping shall be performed.
2864

2865 The "i18n" FDCC-set is for the LC_NUMERIC category:
2866

```
2867 LC_NUMERIC
2868 % This is the 14652 i18n fdcc-set definition for
2869 % the LC_NUMERIC category.
2870 %
2871 decimal_point   "<,>"
2872 thousands_sep   ""
2873 grouping        -1
2874 %
2875 END LC_NUMERIC
```
2876

4.7 LC_TIME

2880 The LC_TIME category defines the rules and symbols that shall be used to format date
2881 and time information. The following keywords shall be defined:
2882

2883 **copy** Specify the name of an existing FDCC-set to be used as the source
2884 for the definition of this category. If this keyword is specified, no
2885 other keyword shall be specified.
2886 **abday** Define the abbreviated weekday names for calendar systems with
2887 weeks of constant length, to be referenced by the %a field descriptor.
2888 The length of the week and a gregorian date for the first weekday is

2889		defined by the "week" keyword. The operand shall consist of
2890		semicolon-separated strings. The first string shall be the abbreviated
2891		name of the day corresponding to the first day of the week (default
2892		Sunday), the second the abbreviated name of the day corresponding
2893		to the second day of the week (default Monday), and so on.
2894	**day**	Define the full weekday names for calendar systems with weeks of
2895		constant length, to be referenced by the %A field descriptor. The
2896		length of the week and a gregorian date for the first weekday is
2897		defined by the "week" keyword. The operand shall consist of
2898		semicolon-separated strings. The first string shall be the full name of
2899		the day corresponding to the first day of the week (default Sunday),
2900		the second the full name of the day corresponding to the second day
2901		of the week (default Monday), and so on.
2902	**week**	Shall be used to define the number of days in a week, and which
2903		weekday is the first weekday (the first weekday has the value 1), and
2904		which week is to be considered the first in a year. The first operand
2905		is an integer specifying the number of days in the week. The second
2906		operand is an integer specifying the Gregorian date in the format
2907		YYYYMMDD. The third operand is an integer specifying the
2908		weekday number to be contained in the first week of the year. If the
2909		keyword is not specified the values are taken as 7, 19971130 (a
2910		Sunday), and 7 (Saturday), respectively. ISO 8601 conforming
2911		applications should use the values 7, 19971201 (a Monday), and 4
2912		(Thursday), respectively. This keyword is optional.
2913	**abmon**	Define the abbreviated month names, to be referenced by the %b
2914		field descriptor. The operand shall consist of twelve or thirteen
2915		semicolon-separated strings. The first string shall be the abbreviated
2916		name of the first month of the year (January), the second the
2917		abbreviated name of the second month, and so on.
2918	**mon**	Define the full month names, to be referenced by the %B field
2919		descriptor. The operand shall consist of twelve or thirteen semicolon-
2920		separated strings. The first string shall be the full name of the first
2921		month of the year (January), the second the full name of the second
2922		month, and so on.
2923	**d_t_fmt**	Define the appropriate date and time representation, to be referenced
2924		by the %c field descriptor. The operand shall consist of a string, and
2925		can contain any combination of characters and field descriptors. In
2926		addition, the string can contain escape sequences defined in Table 3.
2927	**d_fmt**	Define the appropriate date representation, to be referenced by the
2928		%x field descriptor. The operand shall consist of a string, and can
2929		contain any combination of characters and field descriptors. In
2930		addition, the string can contain escape sequences defined in Table 3.
2931	**t_fmt**	Define the appropriate time representation, to be referenced by the
2932		%X field descriptor. The operand shall consist of a string, and can
2933		contain any combination of characters and field descriptors. In
2934		addition, the string can contain escape sequences defined in Table 3.
2935	**am_pm**	Define the appropriate representation of the ante meridiem and post
2936		meridiem strings, to be referenced by the %p field descriptor. The
2937		operand shall consist of two strings, separated by a semicolon. The
2938		first string shall represent the antemeridiem designation, the last

2939		string the postmeridiem designation. The keyword is optional. If
2940		unspecified, the %p field descriptor shall refer to the empty string.
2941	**t_fmt_ampm**	Define the appropriate time representation in the 12-hour clock
2942		format with "am_pm", to be referenced by the %r field descriptor.
2943		The operand shall consist of a string and can contain any
2944		combination of characters and field descriptors. If the string is empty,
2945		the 12-hour format is not supported in the FDCC-set.
2946		
2947	**The following keywords are all optional**	
2948		
2949	**era**	Shall be used to define alternate Eras, corresponding to the %E field
2950		descriptor modifier. The format of the operand is unspecified, but
2951		shall support the definition of the %EC and %Ey field descriptors,
2952		and may also define the "era_year" format (%EY).
2953	**era_year**	Shall be used to define the format of the year in alternate Era format,
2954		corresponding to the %EY field descriptor.
2955	**era_d_fmt**	Shall be used to define the format of the date in alternate Era
2956		notation, corresponding to the %Ex field descriptor.
2957	**alt_digits**	Shall be used to define alternate symbols for digits, corresponding to
2958		the %O field descriptor modifier. The operand shall consist of
2959		semicolon-separated strings. The first string shall be the alternate
2960		symbol corresponding with zero, the second string the symbol
2961		corresponding with one, and so on. Up to 100 alternate symbol
2962		strings can be specified. The %O modifier indicates that the string
2963		corresponding to the value specified via the field descriptor shall be
2964		used instead of the value.
2965	**first_weekday**	Shall be used to define the first day to be displayed, for example in a
2966		calendar display utility. The operand is an integer specifying the day
2967		number (1 = first) according to the information specified with the
2968		"day" keyword. The keyword may be omitted, and then the value 1 is
2969		taken, corresponding to Sunday for a week beginning Sunday, or to
2970		Monday for a week beginning Monday.
2971	**first_workday**	Shall be used to define the first workday as an integer according to
2972		the day numbering specified with the "week" keyword.
2973	**cal_direction**	Shall be used to define the direction of the display of dates, for
2974		example in a calendar display utility. The operand is an integer, and
2975		the following values are defined:
2976		1 left-right from top
2977		2 top-down from left
2978		3 right-left from top
2979		The keyword may be omitted, and then the value 1 is taken.
2980	**timezone**	Shall be used to define a set of timezones, each defined by a string.
2981		In the following the characters <, >, [and] are used as
2982		metacharacters. Only characters with a visible glyph from the
2983		portable character set may be used, except in the <std> and <dst>
2984		fields. The syntax of the string is:
2985		
2986		<std><offset><dst>[<offset>][,<rule>[,<rule>...]]
2987		
2988		where

2989	\<std\> and \<dst\>	Indicates no less than three, nor more than 10
2990		characters that are the designation for the
2991		standard \<std\> or summer \<dst\> time zone.
2992		only \<std\> is required; if \<dst\> is missing, then
2993		summer time does not apply in this category.
2994		Upper- and lowercase letters are explicitly
2995		allowed. Any characters except a leading colon
2996		\<:\> or digits, the comma \<,\>, the minus \<-\>,
2997		the plus \<+\>, and the null character are
2998		permitted to appear in these fields, but their
2999		meaning is unspecified.
3000	\<offset\>	Indicates the value one must add to the local
3001		time to arrive at the Coordinated Universal
3002		Time. The \<offset\> has the form:
3003		
3004		hh[:mm[:ss]]
3005		
3006		The minutes (mm) and seconds (ss) are
3007		optional. The hour (hh) shall be required and
3008		may be a single digit. The \<offset\> following
3009		\<std\> shall be required. If no \<offset\> follows
3010		\<dst\>, summer time is assumed to be one hour
3011		ahead of standard time. One or more digits may
3012		be used; the value is always interpreted as a
3013		decimal number. The hour shall be between
3014		zero and 24, and the minutes (and seconds) - if
3015		present - shall be between zero and 59. If
3016		preceded by a "-", the time zone shall be east
3017		of the Prime Meridian; otherwise it shall be
3018		west of (which may be indicated by an optional
3019		preceding "+").
3020	\<rule\>	Indicates when to change to and back from
3021		summer time. The \<rule\> has the form:
3022		\<date\>[/\<time\>/\<year\>],\<date\>[/\<time
3023		\>/\<year\>]
3024		where the first \<date\> describes when the
3025		change from standard time to summer time
3026		occurs, and the second \<date\> describes when
3027		the change back happens. Each \<time\> field
3028		describes when, in current local time, the
3029		change to the other time is made. The first
3030		\<year\> field defines the beginning of the
3031		validity of this rule, and the second \<year\>
3032		field defines the end of the validity of the rule.
3033		A number of rules may be given.
3034		
3035		The format of \<date\> shall be one of the
3036		following:
3037		
3038		J\<n\> The Julian day \<n\> (1 \<= n

3039   <= 365) Leap years shall not

3040   be counted. That is, in all

3041 years - including leap years - February 28 is day 59 and March 1 is day 60. It is impossible to explicitly refer to the occasional February 29.

3046   <n>   The zero-based Julian day (0 <= n <= 365). Leap years shall be counted and it is possible to refer to February 29.

3051   M<m>.<n>.<d>

3052   the <d>th day (0 <= d <= 7) of week <n> of month <m> (1 <= n <= 5, 1 <= m <= 12, where week 5 means "the last <d> day in month <m>" which may occur in either the fourth or fifth week). Week 1 is the first week in which the <d>th day occurs. Day zero and day seven is Sunday.

3063   The <time> has the same format as <offset> except that no leading sign ("-" or "+") shall be allowed. The default, if <time> is not given, shall be "02:00:00".

3068   The <year> has the format YYYY.

3070   NOTE: This way of specifying the timezone is compatible with the format for the environment variable TZ described in Section 8.1.1 of POSIX.1.

4.7.1   Date Field Descriptors

The LC_TIME category defines the interpretation of a number of field descriptors. The field descriptors are also available in the definitions with the following LC_TIME keywords: "d_t_fmt", "d_fmt", "t_fmt", "t_fmt_ampm", "era", and "era_d_fmt". A field descriptor may not be used with the LC_TIME keywords defining it.

Table 3: Escape sequences for the date field

%a	FDCC-set's abbreviated weekday name.
%A	FDCC-set's full weekday name.
%b	FDCC-set's abbreviated month name.
%B	FDCC-set's full month name.
%c	FDCC-set's appropriate date and time representation.
%C	Century (a year divided by 100 and truncated to integer) as decimal

3089		number (00-99).
3090	%d	Day of the month as a decimal number (01-31).
3091	%D	Date in the format mm/dd/yy.
3092	%e	Day of the month as a decimal number (1-31 in at two-digit field with
3093		leading <space> fill).
3094	%F	The date in the format YYYY-MM-DD (ISO 8601 format).
3095	%g	Week-based year within century, as a decimal number (00-99).
3096	%G	Week-based year with century, as a decimal number (for example 1997).
3097	%h	A synonym for %b.
3098	%H	Hour (24-hour clock), as a decimal number (00-23).
3099	%I	Hour (12-hour clock), as a decimal number (01-12).
3100	%j	Day of the year, as a decimal number (001-366).
3101	%m	Month, as a decimal number (01-13).
3102	%M	Minute, as a decimal number (00-59).
3103	%n	A <newline> character.
3104	%p	FDCC-set's equivalent of either AM or PM.
3105	%r	12-hour clock time (01-12), using the AM/PM notation.
3106	%R	24-hour clock time, in the format "%H:%M".
3107	%S	Seconds, as a decimal number (00-61).
3108	%t	A <tab> character.
3109	%T	24-hour clock time, in the format HH:MM:SS.
3110	%u	Weekday, as a decimal number (1(Monday)-7).
3111	%U	Week number of the year (Sunday as the first day of the week) as a
3112		decimal number (00-53). All days in a new year preceding the first
3113		Sunday shall be considered to be in week 0.
3114	%v	Week number of the year, as a decimal number with two digits including
3115		a possible leading zero, according to "week" keyword.
3116	%V	Week of the year (Monday as the first day of the week), as a decimal
3117		number (01-53). The method for determining the week number shall be as
3118		specified by ISO 8601.
3119	%w	Weekday, as a decimal number (0(Sunday)-6).
3120	%W	Week number of the year (Monday as the first day of the week), as a
3121		decimal number (00-53). All days in a new year preceding the first
3122		Monday shall be considered to be in week 0.
3123	%x	FDCC-set's appropriate date representation.
3124	%X	FDCC-set's appropriate time representation.
3125	%y	Year within century (00-99).
3126	%Y	Year with century, as a decimal number.
3127	%z	The offset from UTC in the ISO 8601 format "-0430" (meaning 4 hours
3128		30 minutes behind UTC, west of Greenwich), or by no characters if no
3129		time zone is determinable.
3130	%Z	Time-zone name, or no characters if no time zone is determinable.
3131	%%	A <percent-sign> character.
3132		

3133   NOTE: %g, %G and %V give values according to the ISO 8601 week-based year. In
3134   this system, weeks begin on a Monday and week 1 of the year is the week that includes
3135   4th January, which is also the week that includes the first Thursday of the year, and is
3136   also the first week that contains at least four days in the year. If the first Monday of the
3137   year is the 2nd, 3rd or 4th, the preceding days are part of the last week of the
3138   preceding year; thus, for Saturday 2nd January 1999, %G is replaced by 1998 and %V

3139   is replaced by 53. If the 29th, 30th or 31st January is a Monday, it and any following
3140   days are part of week 1 of the following year. Thus, for Tuesday 30th December 1997,
3141   %G is replaced by 1998 and %V is replaced by 1.

3142

3143   **4.7.2   Modified Field Descriptors**

3144

3145   Some field descriptors can be modified by the E and O modifier characters to indicate a
3146   different format or specification as specified in the LC_TIME FDCC-set description. If the
3147   corresponding keyword (see "era", "era_year", "era_d_fmt", and "alt_digits") is not
3148   specified for the current FDCC-set, the unmodified field descriptor value shall be used.

3149

3150 %Ec	FDCC-set's alternate date and time representation.
3151 %EC	The name of the base year (period) in the FDCC-set's alternate represen-
3152	tation.
3153 %Ex	FDCC-set's alternate date representation.
3154 %EX	FDCC-set's alternate time representation.
3155 %Ey	Offset from %EC (year only) in the FDCC-set's alternate representation.
3156 %EY	Full alternate year representation.
3157 %Od	Day of month using the FDCC-set's alternate numeric symbols.
3158 %Oe	Day of month using the FDCC-set's alternate numeric symbols.
3159 %Of	Weekday as a decimal number according to alt_day (1 is first day).
3160 %OH	Hour (24-hour clock) using the FDCC-set's alternate numeric symbols.
3161 %OI	Hour (12-hour clock) using the FDCC-set's alternate numeric symbols.
3162 %Om	Month using the FDCC-set's alternate numeric symbols.
3163 %OM	Minutes using the FDCC-set's alternate numeric symbols.
3164 %OS	Seconds using the FDCC-set's alternate numeric symbols.
3165 %Ou	Weekday as a number in the alternate representation of the FDCC-set
3166	(Monday=1).
3167 %OU	Week number of the year (Sunday as the first day of the week) using the
3168	FDCC-set's alternate numeric symbols.
3169 %OV	Week number of the year (Monday as the first day of the week, ISO 8601
3170	rules) using the alternate numeric symbols of the FDCC-set.
3171 %Ow	Weekday as number in the FDCC-set's alternate representation
3172	(Sunday=0).
3173 %OW	Week number of the year (Monday as the first day of the week) using the
3174	FDCC-set's alternate numeric symbols.
3175 %Oy	Year (offset from %C) in alternate representation.

3176

3177   **4.7.3   "i18n" LC_TIME category**

3178

3179   The "i18n" LC_TIME category is (following ISO 8601):

3180
```
3181    LC_TIME
3182    % This is the ISO/IEC 14652 "i18n" definition for
3183    % the LC_TIME category.
3184    %
3185    % Weekday and week numbering according to ISO 8601
3186    abday   "<1>";"<2>";"<3>";"<4>";"<5>";"<6>;<7>"
3187    day     "<1>";"<2>";"<3>";"<4>";"<5>";"<6>;<7>"
3188    week    7;19971201;4
3189    abmon   "<0><1>";"<0><2>";"<0><3>";"<0><4>";"<0><5>";"<0><6>";/
3190            "<0><7>";"<0><8>";"<0><9>";"<1><0>";"<1><1>";"<1><2>"
3191    mon     "<0><1>";"<0><2>";"<0><3>";"<0><4>";"<0><5>";"<0><6>";/
3192            "<0><7>";"<0><8>";"<0><9>";"<1><0>";"<1><1>";"<1><2>"
3193    am_pm   "";""
3194    % Date formats following ISO 8601
```

```
3195    % Appropriate date and time representation (%c)
3196    %        "%F %T"
3197    d_t_fmt "<%><F><SP><%><T>"
3198    %
3199    % Appropriate date representation (%x)   "%F"
3200    d_fmt   "<%><F>"
3201    %
3202    % Appropriate time representation (%X)   "%T"
3203    t_fmt   "<%><T>"
3204    t_fmt_ampm ""
3205    %
3206    END LC_TIME
```

4.8   LC_MESSAGES

The LC_MESSAGES category shall define the format and values for affirmative and negative responses. The operands shall be strings or extended regular expressions to specify which response strings that should be considered matches; see ISO/IEC 9945-2:1993 clause 2.8.4 for a definition of extended regular expressions. The following keywords shall be defined:

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 shall be specified.

yesexpr   The operand shall consist of an extended regular expression that describes the acceptable affirmative response to a question expecting an affirmative or negative response.

noexpr   The operand shall consist of an extended regular expression that describes the acceptable negative response to a question expecting an affirmative or negative response.

The "i18n" LC_MESSAGES category is:

```
LC_MESSAGES
% This is the ISO/IEC 14652 "i18n" definition for
% the LC_MESSAGES category.
%
yesexpr "<U005B><+><1><U005D>"
noexpr  "<U005B><-><0><U005D>"
END LC_MESSAGES
```

4.9   LC_PAPER

The LC_PAPER category defines the default size of paper used for documents. The following keywords shall be defined:

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 shall be specified.

height   Shall be used to specify the vertical dimension of the paper. The operand is an integer and the value is the height measured in millimetres.

width   Shall be used to specify the horizontal dimension of the paper. The operand is an integer and the value is the width measured in millimetres.

NOTE: If the height is greater than the width, it is called to be in portrait position, else it is called to be in landscape position.

The "i18n" LC_PAPER category is:

```
LC_PAPER
% This is the ISO/IEC 14652 "i18n" definition for
% the LC_PAPER category.
%
height  297
width   210
END LC_PAPER
```

4.10 LC_NAME

The LC_NAME category defines formats to be used in addressing a person, e.g. in a postal address or in a letter. The following keywords shall be defined:

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 shall be specified.
name_fmt	Define the appropriate representation of a person's name and title. The operand shall consist of a string, and can contain any combination of characters and field descriptors. In addition, the string can contain escape sequences defined below.
name_gen	The operand is a string defining a salutation valid for all persons, example: the Japanese "-sama" salutation in a letter.
name_miss	The operand is a string defining a salutation valid for unmarried females.
name_mr	The operand is a string defining a salutation valid for males.
name_mrs	The operand is a string defining a salutation valid for married females.
name_ms	The operand is a string defining a salutation valid for all females.

　　NOTE: There are a number of variations for addressing a person among the cultures. Middle names are not used in many countries and even the family name is not used in some countries. The specification below should be regarded as a starting point for this problem.

The LC_NAME category defines the interpretation of a number of escape sequences. The escape sequences are also available in the definitions with the following LC_NAME keywords: "name_fmt".

Escape sequences for the "name_fmt" keyword:

%f	Family names.
%F	Family names in uppercase.
%g	First given name.
%G	First given initial.
%l	First given name with latin letters.
%o	Other shorter name, eg. "Bill".
%m	Middle names.
%M	Middle initial.
%p	Profession.
%s	Salutation, such as "Doctor"
%S	Abbreviated salutation, such as "Mr." or "Dr."
%d	Salutation, using the FDCC-sets conventions, with 1 for the name_gen, 2 for name_mr, 3 for name_mrs, 4 for name_miss, 5 for name_ms. The

3305	vaule may be stored in the database with the person information.
3306	%t If the preceding escape sequence resulted in an empty string, then the
3307	empty string, else a <space>.

3309 Each escape sequence may have an <R> after the <%> to specify that the information is
3310 taken from a Romanized version string of the entity.

3312 The "i18n" LC_NAME category is:

```
3314   LC_NAME
3315   % This is the ISO/IEC 14652 "i18n" definition for
3316   % the LC_NAME category.
3317   %
3318   name_fmt    "<%><p><%><t><%><g><%><t><%><m><%><t><%><f>"
3319   END LC_NAME
```

4.11   LC_ADDRESS

The LC_ADDRESS category defines formats to be used in specifying a location like a
person's living or office, for use in a postal address or in a letter, and other items related
to geography. All keywords are optional. The following keywords shall be recognized:

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 shall be specified.
postal_fmt	Define the appropriate representation of a postal address such as street and city. The proper formatting of a person's name and title is done with the "name_fmt" keyword of the LC_NAME category. The operand shall consist of a string, and can contain any combination of characters and field descriptors. In addition, the string can contain escape sequences defined below.
country_name	The operand is a string with the name of the country in the language of the FDCC-set.
country_post	The operand is a string with the abbreviation of the country, used for postal addresses, for example by CEPT-MAILCODE.
country_ab2	The operand is a string with the two-letter abbreviation of the country, according to ISO 3166.
country_ab3	The operand is a string with the three-letter abbreviation of the country, according to ISO 3166.
country_num	The operand is an integer with the three-digit number of the country, according to ISO 3166.
country_car	The operand is a string with the abbreviation of the country, used for motor vehicles and traffic, according to the Genève convention 1949:68.
country_isbn	The operand is a string with the abbreviation of the country, used for book numbering (ISBN), according to ISO 2108. ISBN numbers are allocated according to country.
lang_name	The operand is a string with the name of the language in the language of the FDCC-set.
lang_ab	The operand is a string with the two-letter abbreviation of the language, according to ISO 639.
lang_term	The operand is a string with the three-letter abbreviation of the language for terminology use, according to ISO 639-2.

3358	**lang_lib**	The operand is a string with the three-letter abbreviation of the
3359		language for library use, according to ISO 639-2. If not specified, the
3360		value of the "lang_term" keyword is taken.

3362 The LC_ADDRESS category defines the interpretation of a number of escape sequences.
3363 The escape sequences are also available in the definitions with the following
3364 LC_ADDRESS keywords: "postal_fmt".

3366 Escape sequences for the "postal_fmt" keyword:

3368	%a	C/O address.
3369	%f	Firm name.
3370	%d	department name.
3371	%b	Building name.
3372	%s	street or block (eg. Japanese) name.
3373	%h	house number or designation.
3374	%N	if any graphical characters have been specified then an end of line is
3375		made.
3376	%t	if the preceding escape sequence resulted in an empty string, then the
3377		empty string, else a <space>.
3378	%r	room number, door designation.
3379	%e	floor number.
3380	%C	country designation, from the <country_post> keyword.
3381	%z	zip number, postal code.
3382	%T	town, city.
3383	%S	state, province, or prefecture.
3384	%c	country.

3386 Each escape sequence may have an <R> after the <%> to specify that the information is
3387 taken from a Romanized version string of the entity.

3389   NOTE: There are a number of variations for specifying a location among the cultures.
3390   Some of the information, like the middle names, or even the family name, is not used
3391   in some cultures. The specification here should be regarded as a start point for this
3392   problem.

3394 The "i18n" LC_ADDRESS category is:

```
3396 LC_ADDRESS
3397 % This is the ISO/IEC 14652 "i18n" definition for
3398 % the LC_ADDRESS category.
3399 %
3400 postal_fmt    "<%><a><%><N><%><f><%><N><%><d><%><N><%><b><%><N>/
3401 <%><s><SP><%><h><SP><%><e><SP><%><r><%><N>/
3402 <%><C><-><%><z><SP><%><T><%><N><%><c><%><N>"
3403 END LC_ADDRESS
```

4.12   LC_TELEPHONE

3408 The LC_TELEPHONE category defines formats to be used with telephone services. All
3409 keywords are optional. The following keywords shall be defined:

3411	**copy**	Specify the name of an existing FDCC-set to be used as the source

3412		for the definition of this category. If this keyword is specified, no
3413		other keyword shall be specified.
3414	**tel_int_fmt**	Define the appropriate representation of a telephone number for
3415		international use. The operand shall consist of a string, and can
3416		contain any combination of characters and field descriptors. In
3417		addition, the string can contain escape sequences defined below.
3418	**tel_dom_fmt**	Define the appropriate representation of a telephone number for
3419		domestic use. The operand shall consist of a string, and can contain
3420		any combination of characters and field descriptors. In addition, the
3421		string can contain escape sequences defined below.
3422	**int_select**	The operand is a string with the digits used to call international
3423		telephone numbers.
3424	**int_prefix**	The operand is a string with the prefix used from other countries to
3425		call the area

3426

3427 The LC_TELEPHONE category defines the interpretation of a number of escape
3428 sequences. The escape sequences are also available in the definitions with the following
3429 LC_TELEPHONE keywords: "tel_int_fmt" and "tel_dom_fmt".

3430

3431	%a	area code without prefix (prefix is often <0>).
3432	%A	area code including prefix (prefix is often <0>).
3433	%l	local number.
3434	%c	country code
3435	%C	alternative carrier service code used for dialling abroad

3436

3437 The "i18n" LC_TELEPHONE category is:

3438
```
3439   LC_TELEPHONE
3440   % This is the ISO/IEC 14652 "i18n" definition for
3441   % the LC_TELEPHONE category.
3442   %
3443   tel_int_fmt    "<+><%><c><SP><%><a><SP><%><l>"
3444   END LC_TELEPHONE
```
3445

3446

3447   5. CHARMAP

3448

3449 A character set description may exist for each coded character set supported by an
3450 application. This text is referred elsewhere in this Technical Report as a charmap.

3451

3452 A conforming charmap to be used with a FDCC-set shall support the portable character set
3453 specified in Table 1.

3454

3455 Conforming charmaps shall specify certain character and character set attributes, as
3456 defined in 5.1.

3457

3458   5.1   Character Set Description Text

3459

3460 The character set description text (charmap) describes the mapping between symbolic
3461 character names and actual encoding of a coded character set. It is used to bind the
3462 symbolic character names in a FDCC-set to an actual encoding, so an application can
3463 process data in this encoding.

3464

3465 The following declarations can precede the character definitions. Each shall consist of the
3466 symbol shown in the following list, starting in column 1, including the surrounding
3467 brackets, followed by one of more "blank"s, followed by the value to be assigned to the
3468 symbol. If any of the declarations are included, they shall be specified in the order shown
3469 in the following list:
3470
3471 **<code_set_name>**   The name of the coded character set for which the character set
3472   description text is defined. The characters of the name shall be
3473   taken from the set of characters with visible glyphs defined in
3474   Table 1.
3475
3476 **<mb_cur_max>**   The maximum number of bytes in a multibyte character. This
3477   shall default to 1.
3478
3479 **<mb_cur_min>**   An unsigned positive integer value that shall define the
3480   minimum number of bytes in a character for the encoded
3481   character set. The value shall be less or equal to "mb_cur_max".
3482   If not specified, the minimum number shall be equal to
3483   "mb_cur_max".
3484
3485 **<escape_char>**   The escape character used to indicate that the characters
3486   following shall be interpreted in a special way, as defined later
3487   in this subclause. This shall default to backslash (\). The
3488   character slash (/) is used in all the following text and examples,
3489   unless otherwise noted.
3490
3491 **<comment_char>**   The character that when placed in column 1 of a charmap line, is
3492   used to indicate that the line shall be ignored. The default
3493   character shall be the number sign (#). The character percent-
3494   sign (%) is used in all the following text and examples, unless
3495   otherwise noted.
3496
3497 **<repertoiremap>**   The name of the repertoiremap used to define the symbolic
3498   character names in the charmap. The characters of the name
3499   shall be taken from the set of characters with visible glyphs
3500   defined in Table 1.
3501
3502 **<escseq>**   defines the escape sequences for ISO 2022 shifting for the coded
3503   character set defined by the charmap. The semicolon-separated
3504   operands are all strings with characters taken from the set of
3505   characters with visible glyphs defined in table 1. The first
3506   operand defines the g-set or c-set to be defined, and the
3507   following values are defined: c0, c1, g0, g1, g2, g3. The second
3508   operand defines what range of characters in the charmap is
3509   affected, and the values defined are: c0, c1, g0, g1. The third
3510   operand is the escape sequence that is defined.
3511
3512 **<addset>**   the name of the charmap to be added the current coded character
3513   set and to be selected by the escape sequences defined by
3514   <escseq> of the added charmap.

3515	**\<include\>**	include the encoding of another charmap in the current charmap.
3516		The semicolon-separated operands are all strings with characters
3517		taken from the set of characters with visible glyphs defined in
3518		table 1. The first operand defines the g-set or c-set to be defined
3519		in the current charmap, and the following values are defined: c0,
3520		c1, g0, g1, g2, g3. The second operand defines a range of
3521		characters in the referenced charmap, and the values defined are:
3522		c0, c1, g0, g1. The third operand is the name of the charmap to
3523		be included. The coded character sets are defined initially for the
3524		encoding, and therefore do not need escape sequences for
3525		identification. If two g0 sets are defined, the second is switched
3526		to using the SHIFT OUT control character, while the first is
3527		shifted to using the SHIFT IN control character.

3528
3529 The character set mapping definitions shall be all the lines immediately following an
3530 identifier line containing the string "CHARMAP" starting in column 1, and preceding a
3531 trailer line containing the string "END CHARMAP" starting in column 1. Empty lines
3532 and lines containing a \<comment_char\> in the first column shall be ignored. Each
3533 noncomment line of the character set mapping definition (i.e., between the "CHARMAP"
3534 and "END CHARMAP" lines of the text) shall be in one of the following syntaxes.

3535
3536
3537   "%s %s %s\n", \<symbolic-name\>,\<encoding\>,\<comments\>

3538
3539   "%s...%s %s %s\n", \<symbolic-name\>,\<symbolic-name\>,\<encoding\>,\<comments\>

3540
3541   "%s....%s %s %s\n", \<symbolic-name\>,\<symbolic-name\>,\<encoding\>,\<comments\>

3542
3543   "%s..%s %s %s\n", \<symbolic-name\>,\<symbolic-name\>,\<encoding\>,\<comments\>

3544
3545 In the first syntax, the line of the character set mapping definition shall start with the
3546 symbolic name, immediately preceded by a \<less-than\> character and immediately
3547 followed by a \<greater-than\> character. Symbolic names shall only contain characters
3548 from the set shown with a visible glyph in Table 1.

3549
3550 The same symbolic name may occur several times, with different values. The first value is
3551 the one used when generating an encoding, while the other values are accepted in
3552 decoding. Symbolic names may be included to identify values that can overlap with each
3553 other or with the values of the symbolic names shown in Table 1. It is possible to specify
3554 symbolic names for which no encoding exists in the encoded character set, by not
3555 specifying a value.

3556
3557 In the second and third syntax (symbolic decimal ellipsis), the line in the character set
3558 mapping defines a range of one or more symbolic names. The difference between the
3559 second and the third syntax is the number of dots in the ellipsis: the second has 3 dots, the
3560 third has 4 dots. In these forms the symbolic names shall consist of zero or more
3561 nonnumeric characters from the set shown with visible glyphs in Table 1, followed by an
3562 integer formed by one or more decimal digits. The characters preceding the integer shall
3563 be identical in the two symbolic names, and the integer formed by the digits in the second
3564 symbolic name shall be identical to or greater than the integer formed by the digits in the

3565 first name. This shall be interpreted as a series of symbolic names formed from the
3566 common part and each of the integers in decimal format between the first and the second
3567 integer, inclusive, and with a length of the symbolic names generated that is equal to the
3568 length of the first (and also the second) symbolic name. As an example,
3569 <j0101>....<j0104> is interpreted as the symbolic names <j0101>, <j0102>, <j0103>, and
3570 <j0104>, in that order.

3571
3572   Note: The rationale to allow both a 3-dot and a 4-dot symbol for symbolic decimal
3573   ellipses is that in the POSIX standard the decimal symbolic ellipses was defined by a 3-
3574   dot symbol for charmaps, while the 3-dot symbol was an absolute ellipses for POSIX
3575   locales, and this International standard specifies a 4-dot symbol for the decimal
3576   symbolic ellipses. The 3-dot symbolic decimal ellipses in charmaps is deprecated.

3577
3578 In the fourth syntax (symbolic hexadecimal ellipsis, with two dots), the line in the
3579 character set mapping defines a range of one or more symbolic names. In this form the
3580 symbolic names shall consist of zero or more nonnumeric characters from the set shown
3581 with visible glyphs in Table 1, followed by an integer formed by one or more hexadecimal
3582 digits, using uppercase letters only for the range "A" to "F". The characters preceding the
3583 hexadecimal integer shall be identical in the two symbolic names, and the integer formed
3584 by the hexadecimal digits in the second symbolic name shall be identical to or greater than
3585 the integer formed by the hexadecimal digits in the first name. This shall be interpreted as
3586 a series of symbolic names formed from the common part and each of the integers in
3587 hexadecimal format using uppercase letters only between the first and the second integer,
3588 inclusive, and with a length of the symbolic names generated that is equal to the length of
3589 the first (and also the second) symbolic name. As an example, <U010E>..<U0111> is
3590 interpreted as the symbolic names <U010E>, <U010F>, <U0110>, and <U0111>, in that
3591 order.

3592
3593 The encoding part shall be expressed as one (for single-byte values) or more concatenated
3594 decimal, octal or hexadecimal constants. Decimal constants shall be represented by two or
3595 three decimal digits, preceded by the escape character and the lowercase letter "d"; for
3596 example /d05, /d97, or /d143. Hexadecimal constants shall be represented by two
3597 hexadecimal digits, preceded by the escape character and the lowercase letter "x"; for
3598 example /x05, /x61, or /x8f. Octal constants shall be represented by two or three octal
3599 digits, preceded by the escape character; for example /05, /141, or /217. In a charmap,
3600 each constant should represent an 8 bit byte for portability reasons. Applications
3601 supporting other byte sizes may allow constants to represent values larger than those that
3602 can be represented in 8 bit bytes, and to allow additional digits in constants. When
3603 constants are concatenated for multibyte character values, they may be of different types,
3604 and interpreted in byte order from the first to the last with the least significant byte of the
3605 multibyte character specified by the last byte. The manner in which these constants are
3606 represented in the character stored in the system is application defined. Omitting bytes
3607 from a multibyte character produces undefined results.

3608
3609 In lines defining ranges of symbolic names, the encoded value is the value for the first
3610 symbolic name in the range (the symbolic name preceding the ellipsis). Subsequent
3611 symbolic names defined by the range shall have encoding values in increasing order. For
3612 example the line

3613
3614   <j0101>....<j0104>   /d129/d254

shall be interpreted as

        \<j0101\>   /d129/d254
        \<j0102\>   /d129/d255
        \<j0103\>   /d130/d000
        \<j0104\>   /d130/d001

The comments parameter is optional.

Example of using ISO 2022 techniques:

The following example defines two coded character sets, a 7-bit and a 14-bit. They are then merged into one encoding. It is an example on how encodings used in Eastern Asia could be specified.

The 7-bit charmap

```
<escape_char> /
<comment_char> %
% The 7bit charmap defines both control and graphic characters
<code_set_name>  "eastern7bit"
<escseq>        "c0";"c0","/x21/x40"
<escseq>        "g0";"g0","/x28/x48"
<escseq>        "g1";"g0","/x29/x48"
<escseq>        "g2";"g0","/x2A/x48"
<escseq>        "g3";"g0","/x2B/x48"

CHARMAP
<tab>          /x08
<newline>       /x0D
<a>            /x61
% more character encodings to be defined here
END CHARMAP
```


The 14-bit charmap

```
<escape_char> /
<comment_char> %
<code_set_name>   "eastern14bit"
<mb_cur_max>    2
<esqseq>        "g0";"g0";"/x24/x40"
<esqseq>        "g1";"g0";"/x24/x29/x40"
<esqseq>        "g2";"g0";"/x24/x2A/x40"
<esqseq>        "g3";"g0";"/x24/x2B/x40"
CHARMAP
<U0365>         /d036/d055   % the character codes are only examples
<U0744>         /d036/d056
% more character encodings to be defined here
END CHARMAP
```


The merged encoding

```
<escape_char> /
<comment_char> %
<code_set_name>    "shift-eastern"
<mb_cur_max>    2
<mb_cur_min>    1
```

```
3674        <include>        "c0";"c0";"eastern7bit"
3675        <include>        "g0";"g0";"eastern7bit"
3676        <include>        "g1";"g0";"eastern14bit"
3677        % This defines the g0 values of "eastern14bit" (without the 8th
3678        % bit set) to be the g1 in this encoding (with the 8th bit set).
3679        %
3680        % So the bytes without the 8th bit set is from the "shift7bit"
3681        % coded character set, while bytes with the 8th bit set are from
3682        % the 14-bit set.
3683
3684   Another merged encoding using the same charmaps:
3685
3686        <escape_char> /
3687        <comment_char> %
3688        <code_set_name>    "EUC-eastern"
3689        <mb_cur_max>       2
3690        <mb_cur_min>       1
3691        <include>        "c0";"c0";"eastern7bit"
3692        <include>        "g0";"g0";"eastern7bit"
3693        <include>        "g0";"g0";"eastern14bit"
3694        % As there are two "g0" sets defined, the first referenced is the
3695        % initial g0 set, while the second can be shifted to via the SHIFT OUT
3696        % control character. The first can then be shifted to by the SHIFT IN
3697        % control character.
3698
3699
```

6   REPERTOIREMAP

FDCC-set and Charmap sources may be specified in a coded character set independent way, using symbolic character names. The relation between the symbolic character names and characters may be specified via a Repertoiremap, which defines the repertoire of characters defined for a FDCC-set, and the symbolic character names and corresponding abstract character (by a reference to ISO/IEC 10646).

The repertoire mapping is defined by specifying the symbolic character name and the ISO/IEC 10646 code position in hexadecimal form (with a preceding 'U') and optionally the long ISO/IEC 10646 character name in the following syntax:

"%s %s %s\n",<symbolic-name>,<10646-short-identifier>,<comments>

The symbolic character name and the ISO/IEC 10646 short identifier are each surrounded by angle brackets <>, and the fields shall be separated by one or more spaces or tabs on a line. If a right angle bracket or an escape character is used within a symbolic name, it shall be preceded by the escape character. Characters not in ISO/IEC 10646 may be referenced by the symbolic character names <P00000000>..<PF8FFFFFF>.

The escape character can be redefined from the default reverse solidus (\) with the first line of the Repertoiremap containing the string "escape_char" followed by one or more spaces or tabs and then the escape character.

Several symbolic character names can refer to the same abstract character, and are then used as synonyms in FDCC-sets and charmaps. The set of <U0000>..<UFFFF> and <U00000000>..<U7FFFFFFF> symbolic names (no lowercase letters) are predefined and refers to the corresponding code points of ISO/IEC 10646 with the same short identifier.

3729   The "i18nrep" repertoiremap is defined to accommodate prior art, such as defined in the
3730   ISO/IEC 9945-2:1993 standard annex G, and used by ISO and IEC member bodies in their
3731   national POSIX locale specifications, and as used in POSIX locales distributed by the
3732   ISO/IEC POSIX working group and X/Open. Many POSIX charmaps registered with
3733   ISO/IEC 15897 use these symbolic names. It also reflects use on the Internet, and many of
3734   the Internet registered charsets are specified using these symbolic names. The "i18nrep"
3735   repertoiremap thus facilitates reuse of both POSIX locale data and POSIX charmaps with
3736   data from this Technical Report. The contents of the "i18nrep" repertoiremap is as follows:

```
3737
3738    escape_char /
3739    <NUL>                       <U0000>  NULL (NUL)
3740    <SOH>                       <U0001>  START OF HEADING (SOH)
3741    <STX>                       <U0002>  START OF TEXT (STX)
3742    <ETX>                       <U0003>  END OF TEXT (ETX)
3743    <EOT>                       <U0004>  END OF TRANSMISSION (EOT)
3744    <ENQ>                       <U0005>  ENQUIRY (ENQ)
3745    <ACK>                       <U0006>  ACKNOWLEDGE (ACK)
3746    <alert>                     <U0007>  BELL (BEL)
3747    <BEL>                       <U0007>  BELL (BEL)
3748    <backspace>                 <U0008>  BACKSPACE (BS)
3749    <tab>                       <U0009>  CHARACTER TABULATION (HT)
3750    <newline>                   <U000A>  LINE FEED (LF)
3751    <vertical-tab>              <U000B>  LINE TABULATION (VT)
3752    <form-feed>                 <U000C>  FORM FEED (FF)
3753    <carriage-return>           <U000D>  CARRIAGE RETURN (CR)
3754    <DLE>                       <U0010>  DATALINK ESCAPE (DLE)
3755    <DC1>                       <U0011>  DEVICE CONTROL ONE (DC1)
3756    <DC2>                       <U0012>  DEVICE CONTROL TWO (DC2)
3757    <DC3>                       <U0013>  DEVICE CONTROL THREE (DC3)
3758    <DC4>                       <U0014>  DEVICE CONTROL FOUR (DC4)
3759    <NAK>                       <U0015>  NEGATIVE ACKNOWLEDGE (NAK)
3760    <SYN>                       <U0016>  SYNCRONOUS IDLE (SYN)
3761    <ETB>                       <U0017>  END OF TRANSMISSION BLOCK (ETB)
3762    <CAN>                       <U0018>  CANCEL (CAN)
3763    <SUB>                       <U001A>  SUBSTITUTE (SUB)
3764    <ESC>                       <U001B>  ESCAPE (ESC)
3765    <IS4>                       <U001C>  FILE SEPARATOR (IS4)
3766    <IS3>                       <U001D>  GROUP SEPARATOR (IS3)
3767    <intro>                     <U001D>  GROUP SEPARATOR (IS3)
3768    <IS2>                       <U001E>  RECORD SEPARATOR (IS2)
3769    <IS1>                       <U001F>  UNIT SEPARATOR (IS1)
3770    <DEL>                       <U007F>  DELETE (DEL)
3771    <space>                     <U0020>  SPACE
3772    <exclamation-mark>          <U0021>  EXCLAMATION MARK
3773    <quotation-mark>            <U0022>  QUOTATION MARK
3774    <number-sign>               <U0023>  NUMBER SIGN
3775    <dollar-sign>               <U0024>  DOLLAR SIGN
3776    <percent-sign>              <U0025>  PERCENT SIGN
3777    <ampersand>                 <U0026>  AMPERSAND
3778    <apostrophe>                <U0027>  APOSTROPHE
3779    <left-parenthesis>          <U0028>  LEFT PARENTHESIS
3780    <right-parenthesis>         <U0029>  RIGHT PARENTHESIS
3781    <asterisk>                  <U002A>  ASTERISK
3782    <plus-sign>                 <U002B>  PLUS SIGN
3783    <comma>                     <U002C>  COMMA
3784    <hyphen>                    <U002D>  HYPHEN-MINUS
3785    <hyphen-minus>              <U002D>  HYPHEN-MINUS
3786    <period>                    <U002E>  FULL STOP
3787    <full-stop>                 <U002E>  FULL STOP
3788    <slash>                     <U002F>  SOLIDUS
3789    <solidus>                   <U002F>  SOLIDUS
3790    <zero>                      <U0030>  DIGIT ZERO
3791    <one>                       <U0031>  DIGIT ONE
3792    <two>                       <U0032>  DIGIT TWO
3793    <three>                     <U0033>  DIGIT THREE
3794    <four>                      <U0034>  DIGIT FOUR
3795    <five>                      <U0035>  DIGIT FIVE
3796    <six>                       <U0036>  DIGIT SIX
3797    <seven>                     <U0037>  DIGIT SEVEN
3798    <eight>                     <U0038>  DIGIT EIGHT
3799    <nine>                      <U0039>  DIGIT NINE
3800    <colon>                     <U003A>  COLON
3801    <semicolon>                 <U003B>  SEMICOLON
3802    <less-than-sign>            <U003C>  LESS-THAN SIGN
3803    <equals-sign>               <U003D>  EQUALS SIGN
3804    <greater-than-sign>         <U003E>  GREATER-THAN SIGN
3805    <question-mark>             <U003F>  QUESTION MARK
3806    <commercial-at>             <U0040>  COMMERCIAL AT
3807    <left-square-bracket>       <U005B>  LEFT SQUARE BRACKET
3808    <backslash>                 <U005C>  REVERSE SOLIDUS
3809    <reverse-solidus>           <U005C>  REVERSE SOLIDUS
3810    <right-square-bracket>      <U005D>  RIGHT SQUARE BRACKET
```

```
3811    <circumflex>                <U005E>  CIRCUMFLEX ACCENT
3812    <circumflex-accent>         <U005E>  CIRCUMFLEX ACCENT
3813    <underscore>                <U005F>  LOW LINE
3814    <low-line>                  <U005F>  LOW LINE
3815    <grave-accent>              <U0060>  GRAVE ACCENT
3816    <left-brace>                <U007B>  LEFT CURLY BRACKET
3817    <left-curly-bracket>        <U007B>  LEFT CURLY BRACKET
3818    <vertical-line>             <U007C>  VERTICAL LINE
3819    <right-brace>               <U007D>  RIGHT CURLY BRACKET
3820    <right-curly-bracket>       <U007D>  RIGHT CURLY BRACKET
3821    <tilde>                     <U007E>  TILDE
3822
3823    <a8>        <U0252>  Weight indicating the position of the last a
3824    <b8>        <U0182>  Weight indicating the position of the last b
3825    <c8>        <U0255>  Weight indicating the position of the last c
3826    <d8>        <U018D>  Weight indicating the position of the last d
3827    <e8>        <U0264>  Weight indicating the position of the last e
3828    <f8>        <U0191>  Weight indicating the position of the last f
3829    <g8>        <U01A2>  Weight indicating the position of the last g
3830    <h8>        <U02BD>  Weight indicating the position of the last h
3831    <i8>        <U0196>  Weight indicating the position of the last i
3832    <j8>        <U0284>  Weight indicating the position of the last j
3833    <k8>        <U029E>  Weight indicating the position of the last k
3834    <l8>        <U028E>  Weight indicating the position of the last l
3835    <m8>        <U0271>  Weight indicating the position of the last m
3836    <n8>        <U014A>  Weight indicating the position of the last n
3837    <o8>        <U0277>  Weight indicating the position of the last o
3838    <p8>        <U0278>  Weight indicating the position of the last p
3839    <q8>        <U0138>  Weight indicating the position of the last q
3840    <r8>        <U02B6>  Weight indicating the position of the last r
3841    <s8>        <U0286>  Weight indicating the position of the last s
3842    <t8>        <U0287>  Weight indicating the position of the last t
3843    <u8>        <U01B1>  Weight indicating the position of the last u
3844    <v8>        <U028C>  Weight indicating the position of the last v
3845    <w8>        <U028D>  Weight indicating the position of the last w
3846    <x8>        <U216B>  Weight indicating the position of the last x
3847    <y8>        <U01B3>  Weight indicating the position of the last y
3848    <z8>        <U0293>  Weight indicating the position of the last z
3849
3850    <NU>        <U0000>  NULL (NUL)
3851    <SH>        <U0001>  START OF HEADING (SOH)
3852    <SX>        <U0002>  START OF TEXT (STX)
3853    <EX>        <U0003>  END OF TEXT (ETX)
3854    <ET>        <U0004>  END OF TRANSMISSION (EOT)
3855    <EQ>        <U0005>  ENQUIRY (ENQ)
3856    <AK>        <U0006>  ACKNOWLEDGE (ACK)
3857    <BL>        <U0007>  BELL (BEL)
3858    <BS>        <U0008>  BACKSPACE (BS)
3859    <HT>        <U0009>  CHARACTER TABULATION (HT)
3860    <LF>        <U000A>  LINE FEED (LF)
3861    <VT>        <U000B>  LINE TABULATION (VT)
3862    <FF>        <U000C>  FORM FEED (FF)
3863    <CR>        <U000D>  CARRIAGE RETURN (CR)
3864    <SO>        <U000E>  SHIFT OUT (SO)
3865    <SI>        <U000F>  SHIFT IN (SI)
3866    <DL>        <U0010>  DATALINK ESCAPE (DLE)
3867    <D1>        <U0011>  DEVICE CONTROL ONE (DC1)
3868    <D2>        <U0012>  DEVICE CONTROL TWO (DC2)
3869    <D3>        <U0013>  DEVICE CONTROL THREE (DC3)
3870    <D4>        <U0014>  DEVICE CONTROL FOUR (DC4)
3871    <NK>        <U0015>  NEGATIVE ACKNOWLEDGE (NAK)
3872    <SY>        <U0016>  SYNCHRONOUS IDLE (SYN)
3873    <EB>        <U0017>  END OF TRANSMISSION BLOCK (ETB)
3874    <CN>        <U0018>  CANCEL (CAN)
3875    <EM>        <U0019>  END OF MEDIUM (EM)
3876    <SB>        <U001A>  SUBSTITUTE (SUB)
3877    <EC>        <U001B>  ESCAPE (ESC)
3878    <FS>        <U001C>  FILE SEPARATOR (IS4)
3879    <GS>        <U001D>  GROUP SEPARATOR (IS3)
3880    <RS>        <U001E>  RECORD SEPARATOR (IS2)
3881    <US>        <U001F>  UNIT SEPARATOR (IS1)
3882    <DT>        <U007F>  DELETE (DEL)
3883    <PA>        <U0080>  PADDING CHARACTER (PAD)
3884    <HO>        <U0081>  HIGH OCTET PRESET (HOP)
3885    <BH>        <U0082>  BREAK PERMITTED HERE (BPH)
3886    <NH>        <U0083>  NO BREAK HERE (NBH)
3887    <IN>        <U0084>  INDEX (IND)
3888    <NL>        <U0085>  NEXT LINE (NEL)
3889    <SA>        <U0086>  START OF SELECTED AREA (SSA)
3890    <ES>        <U0087>  END OF SELECTED AREA (ESA)
3891    <HS>        <U0088>  CHARACTER TABULATION SET (HTS)
3892    <HJ>        <U0089>  CHARACTER TABULATION WITH JUSTIFICATION (HTJ)
3893    <VS>        <U008A>  LINE TABULATION SET (VTS)
3894    <PD>        <U008B>  PARTIAL LINE FORWARD (PLD)
3895    <PU>        <U008C>  PARTIAL LINE BACKWARD (PLU)
3896    <RI>        <U008D>  REVERSE LINE FEED (RI)
3897    <S2>        <U008E>  SINGLE-SHIFT TWO (SS2)
3898    <S3>        <U008F>  SINGLE-SHIFT THREE (SS3)
3899    <DC>        <U0090>  DEVICE CONTROL STRING (DCS)
```

```
3900        <P1>        <U0091>     PRIVATE USE ONE (PU1)
3901        <P2>        <U0092>     PRIVATE USE TWO (PU2)
3902        <TS>        <U0093>     SET TRANSMIT STATE (STS)
3903        <CC>        <U0094>     CANCEL CHARACTER (CCH)
3904        <MW>        <U0095>     MESSAGE WAITING (MW)
3905        <SG>        <U0096>     START OF GUARDED AREA (SPA)
3906        <EG>        <U0097>     END OF GUARDED AREA (EPA)
3907        <SS>        <U0098>     START OF STRING (SOS)
3908        <GC>        <U0099>     SINGLE GRAPHIC CHARACTER INTRODUCER (SGCI)
3909        <SC>        <U009A>     SINGLE CHARACTER INTRODUCER (SCI)
3910        <CI>        <U009B>     CONTROL SEQUENCE INTRODUCER (CSI)
3911        <ST>        <U009C>     STRING TERMINATOR (ST)
3912        <OC>        <U009D>     OPERATING SYSTEM COMMAND (OSC)
3913        <PM>        <U009E>     PRIVACY MESSAGE (PM)
3914        <AC>        <U009F>     APPLICATION PROGRAM COMMAND (APC)
3915        <SP>        <U0020>     SPACE
3916        <!>         <U0021>     EXCLAMATION MARK
3917        <">         <U0022>     QUOTATION MARK
3918        <Nb>        <U0023>     NUMBER SIGN
3919        <DO>        <U0024>     DOLLAR SIGN
3920        <%>         <U0025>     PERCENT SIGN
3921        <&>         <U0026>     AMPERSAND
3922        <'>         <U0027>     APOSTROPHE
3923        <(>         <U0028>     LEFT PARENTHESIS
3924        <)>         <U0029>     RIGHT PARENTHESIS
3925        <*>         <U002A>     ASTERISK
3926        <+>         <U002B>     PLUS SIGN
3927        <,>         <U002C>     COMMA
3928        <->         <U002D>     HYPHEN-MINUS
3929        <.>         <U002E>     FULL STOP
3930        <//>        <U002F>     SOLIDUS
3931        <0>         <U0030>     DIGIT ZERO
3932        <1>         <U0031>     DIGIT ONE
3933        <2>         <U0032>     DIGIT TWO
3934        <3>         <U0033>     DIGIT THREE
3935        <4>         <U0034>     DIGIT FOUR
3936        <5>         <U0035>     DIGIT FIVE
3937        <6>         <U0036>     DIGIT SIX
3938        <7>         <U0037>     DIGIT SEVEN
3939        <8>         <U0038>     DIGIT EIGHT
3940        <9>         <U0039>     DIGIT NINE
3941        <:>         <U003A>     COLON
3942        <;>         <U003B>     SEMICOLON
3943        <<>         <U003C>     LESS-THAN SIGN
3944        <=>         <U003D>     EQUALS SIGN
3945        </>>        <U003E>     GREATER-THAN SIGN
3946        <?>         <U003F>     QUESTION MARK
3947        <At>        <U0040>     COMMERCIAL AT
3948        <A>         <U0041>     LATIN CAPITAL LETTER A
3949        <B>         <U0042>     LATIN CAPITAL LETTER B
3950        <C>         <U0043>     LATIN CAPITAL LETTER C
3951        <D>         <U0044>     LATIN CAPITAL LETTER D
3952        <E>         <U0045>     LATIN CAPITAL LETTER E
3953        <F>         <U0046>     LATIN CAPITAL LETTER F
3954        <G>         <U0047>     LATIN CAPITAL LETTER G
3955        <H>         <U0048>     LATIN CAPITAL LETTER H
3956        <I>         <U0049>     LATIN CAPITAL LETTER I
3957        <J>         <U004A>     LATIN CAPITAL LETTER J
3958        <K>         <U004B>     LATIN CAPITAL LETTER K
3959        <L>         <U004C>     LATIN CAPITAL LETTER L
3960        <M>         <U004D>     LATIN CAPITAL LETTER M
3961        <N>         <U004E>     LATIN CAPITAL LETTER N
3962        <O>         <U004F>     LATIN CAPITAL LETTER O
3963        <P>         <U0050>     LATIN CAPITAL LETTER P
3964        <Q>         <U0051>     LATIN CAPITAL LETTER Q
3965        <R>         <U0052>     LATIN CAPITAL LETTER R
3966        <S>         <U0053>     LATIN CAPITAL LETTER S
3967        <T>         <U0054>     LATIN CAPITAL LETTER T
3968        <U>         <U0055>     LATIN CAPITAL LETTER U
3969        <V>         <U0056>     LATIN CAPITAL LETTER V
3970        <W>         <U0057>     LATIN CAPITAL LETTER W
3971        <X>         <U0058>     LATIN CAPITAL LETTER X
3972        <Y>         <U0059>     LATIN CAPITAL LETTER Y
3973        <Z>         <U005A>     LATIN CAPITAL LETTER Z
3974        <<(>        <U005B>     LEFT SQUARE BRACKET
3975        <////>      <U005C>     REVERSE SOLIDUS
3976        <)/>>       <U005D>     RIGHT SQUARE BRACKET
3977        <'/>>       <U005E>     CIRCUMFLEX ACCENT
3978        <_>         <U005F>     LOW LINE
3979        <'!>        <U0060>     GRAVE ACCENT
3980        <a>         <U0061>     LATIN SMALL LETTER A
3981        <b>         <U0062>     LATIN SMALL LETTER B
3982        <c>         <U0063>     LATIN SMALL LETTER C
3983        <d>         <U0064>     LATIN SMALL LETTER D
3984        <e>         <U0065>     LATIN SMALL LETTER E
3985        <f>         <U0066>     LATIN SMALL LETTER F
3986        <g>         <U0067>     LATIN SMALL LETTER G
3987        <h>         <U0068>     LATIN SMALL LETTER H
```

```
3988        <i>        <U0069>    LATIN SMALL LETTER I
3989        <j>        <U006A>    LATIN SMALL LETTER J
3990        <k>        <U006B>    LATIN SMALL LETTER K
3991        <l>        <U006C>    LATIN SMALL LETTER L
3992        <m>        <U006D>    LATIN SMALL LETTER M
3993        <n>        <U006E>    LATIN SMALL LETTER N
3994        <o>        <U006F>    LATIN SMALL LETTER O
3995        <p>        <U0070>    LATIN SMALL LETTER P
3996        <q>        <U0071>    LATIN SMALL LETTER Q
3997        <r>        <U0072>    LATIN SMALL LETTER R
3998        <s>        <U0073>    LATIN SMALL LETTER S
3999        <t>        <U0074>    LATIN SMALL LETTER T
4000        <u>        <U0075>    LATIN SMALL LETTER U
4001        <v>        <U0076>    LATIN SMALL LETTER V
4002        <w>        <U0077>    LATIN SMALL LETTER W
4003        <x>        <U0078>    LATIN SMALL LETTER X
4004        <y>        <U0079>    LATIN SMALL LETTER Y
4005        <z>        <U007A>    LATIN SMALL LETTER Z
4006        <(!>       <U007B>    LEFT CURLY BRACKET
4007        <!!>       <U007C>    VERTICAL LINE
4008        <!)>       <U007D>    RIGHT CURLY BRACKET
4009        <'?>       <U007E>    TILDE
4010        <NS>       <U00A0>    NO-BREAK SPACE
4011        <!I>       <U00A1>    INVERTED EXCLAMATION MARK
4012        <Ct>       <U00A2>    CENT SIGN
4013        <Pd>       <U00A3>    POUND SIGN
4014        <Cu>       <U00A4>    CURRENCY SIGN
4015        <Ye>       <U00A5>    YEN SIGN
4016        <BB>       <U00A6>    BROKEN BAR
4017        <SE>       <U00A7>    SECTION SIGN
4018        <':>       <U00A8>    DIAERESIS
4019        <Co>       <U00A9>    COPYRIGHT SIGN
4020        <-a>       <U00AA>    FEMININE ORDINAL INDICATOR
4021        <<<>       <U00AB>    LEFT-POINTING DOUBLE ANGLE QUOTATION MARK
4022        <NO>       <U00AC>    NOT SIGN
4023        <-->       <U00AD>    SOFT HYPHEN
4024        <Rg>       <U00AE>    REGISTERED SIGN
4025        <'m>       <U00AF>    MACRON
4026        <DG>       <U00B0>    DEGREE SIGN
4027        <+->       <U00B1>    PLUS-MINUS SIGN
4028        <2S>       <U00B2>    SUPERSCRIPT TWO
4029        <3S>       <U00B3>    SUPERSCRIPT THREE
4030        <''>       <U00B4>    ACUTE ACCENT
4031        <My>       <U00B5>    MICRO SIGN
4032        <PI>       <U00B6>    PILCROW SIGN
4033        <.M>       <U00B7>    MIDDLE DOT
4034        <',>       <U00B8>    CEDILLA
4035        <1S>       <U00B9>    SUPERSCRIPT ONE
4036        <-o>       <U00BA>    MASCULINE ORDINAL INDICATOR
4037        </>/>>     <U00BB>    RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK
4038        <14>       <U00BC>    VULGAR FRACTION ONE QUARTER
4039        <12>       <U00BD>    VULGAR FRACTION ONE HALF
4040        <34>       <U00BE>    VULGAR FRACTION THREE QUARTERS
4041        <?I>       <U00BF>    INVERTED QUESTION MARK
4042        <A!>       <U00C0>    LATIN CAPITAL LETTER A WITH GRAVE
4043        <A'>       <U00C1>    LATIN CAPITAL LETTER A WITH ACUTE
4044        <A/>>      <U00C2>    LATIN CAPITAL LETTER A WITH CIRCUMFLEX
4045        <A?>       <U00C3>    LATIN CAPITAL LETTER A WITH TILDE
4046        <A:>       <U00C4>    LATIN CAPITAL LETTER A WITH DIAERESIS
4047        <AA>       <U00C5>    LATIN CAPITAL LETTER A WITH RING ABOVE
4048        <AE>       <U00C6>    LATIN CAPITAL LETTER AE (ash)
4049        <C,>       <U00C7>    LATIN CAPITAL LETTER C WITH CEDILLA
4050        <E!>       <U00C8>    LATIN CAPITAL LETTER E WITH GRAVE
4051        <E'>       <U00C9>    LATIN CAPITAL LETTER E WITH ACUTE
4052        <E/>>      <U00CA>    LATIN CAPITAL LETTER E WITH CIRCUMFLEX
4053        <E:>       <U00CB>    LATIN CAPITAL LETTER E WITH DIAERESIS
4054        <I!>       <U00CC>    LATIN CAPITAL LETTER I WITH GRAVE
4055        <I'>       <U00CD>    LATIN CAPITAL LETTER I WITH ACUTE
4056        <I/>>      <U00CE>    LATIN CAPITAL LETTER I WITH CIRCUMFLEX
4057        <I:>       <U00CF>    LATIN CAPITAL LETTER I WITH DIAERESIS
4058        <D->       <U00D0>    LATIN CAPITAL LETTER ETH (Icelandic)
4059        <N?>       <U00D1>    LATIN CAPITAL LETTER N WITH TILDE
4060        <O!>       <U00D2>    LATIN CAPITAL LETTER O WITH GRAVE
4061        <O'>       <U00D3>    LATIN CAPITAL LETTER O WITH ACUTE
4062        <O/>>      <U00D4>    LATIN CAPITAL LETTER O WITH CIRCUMFLEX
4063        <O?>       <U00D5>    LATIN CAPITAL LETTER O WITH TILDE
4064        <O:>       <U00D6>    LATIN CAPITAL LETTER O WITH DIAERESIS
4065        <*X>       <U00D7>    MULTIPLICATION SIGN
4066        <O//>      <U00D8>    LATIN CAPITAL LETTER O WITH STROKE
4067        <U!>       <U00D9>    LATIN CAPITAL LETTER U WITH GRAVE
4068        <U'>       <U00DA>    LATIN CAPITAL LETTER U WITH ACUTE
4069        <U/>>      <U00DB>    LATIN CAPITAL LETTER U WITH CIRCUMFLEX
4070        <U:>       <U00DC>    LATIN CAPITAL LETTER U WITH DIAERESIS
4071        <Y'>       <U00DD>    LATIN CAPITAL LETTER Y WITH ACUTE
4072        <TH>       <U00DE>    LATIN CAPITAL LETTER THORN (Icelandic)
4073        <ss>       <U00DF>    LATIN SMALL LETTER SHARP S (German)
4074        <a!>       <U00E0>    LATIN SMALL LETTER A WITH GRAVE
4075        <a'>       <U00E1>    LATIN SMALL LETTER A WITH ACUTE
4076        <a/>>      <U00E2>    LATIN SMALL LETTER A WITH CIRCUMFLEX
```

66

```
4077    <a?>        <U00E3>    LATIN SMALL LETTER A WITH TILDE
4078    <a:>        <U00E4>    LATIN SMALL LETTER A WITH DIAERESIS
4079    <aa>        <U00E5>    LATIN SMALL LETTER A WITH RING ABOVE
4080    <ae>        <U00E6>    LATIN SMALL LETTER AE (ash)
4081    <c,>        <U00E7>    LATIN SMALL LETTER C WITH CEDILLA
4082    <e!>        <U00E8>    LATIN SMALL LETTER E WITH GRAVE
4083    <e'>        <U00E9>    LATIN SMALL LETTER E WITH ACUTE
4084    <e/>>       <U00EA>    LATIN SMALL LETTER E WITH CIRCUMFLEX
4085    <e:>        <U00EB>    LATIN SMALL LETTER E WITH DIAERESIS
4086    <i!>        <U00EC>    LATIN SMALL LETTER I WITH GRAVE
4087    <i'>        <U00ED>    LATIN SMALL LETTER I WITH ACUTE
4088    <i/>>       <U00EE>    LATIN SMALL LETTER I WITH CIRCUMFLEX
4089    <i:>        <U00EF>    LATIN SMALL LETTER I WITH DIAERESIS
4090    <d->        <U00F0>    LATIN SMALL LETTER ETH (Icelandic)
4091    <n?>        <U00F1>    LATIN SMALL LETTER N WITH TILDE
4092    <o!>        <U00F2>    LATIN SMALL LETTER O WITH GRAVE
4093    <o'>        <U00F3>    LATIN SMALL LETTER O WITH ACUTE
4094    <o/>>       <U00F4>    LATIN SMALL LETTER O WITH CIRCUMFLEX
4095    <o?>        <U00F5>    LATIN SMALL LETTER O WITH TILDE
4096    <o:>        <U00F6>    LATIN SMALL LETTER O WITH DIAERESIS
4097    <-:>        <U00F7>    DIVISION SIGN
4098    <o//>       <U00F8>    LATIN SMALL LETTER O WITH STROKE
4099    <u!>        <U00F9>    LATIN SMALL LETTER U WITH GRAVE
4100    <u'>        <U00FA>    LATIN SMALL LETTER U WITH ACUTE
4101    <u/>>       <U00FB>    LATIN SMALL LETTER U WITH CIRCUMFLEX
4102    <u:>        <U00FC>    LATIN SMALL LETTER U WITH DIAERESIS
4103    <y'>        <U00FD>    LATIN SMALL LETTER Y WITH ACUTE
4104    <th>        <U00FE>    LATIN SMALL LETTER THORN (Icelandic)
4105    <y:>        <U00FF>    LATIN SMALL LETTER Y WITH DIAERESIS
4106    <A->        <U0100>    LATIN CAPITAL LETTER A WITH MACRON
4107    <a->        <U0101>    LATIN SMALL LETTER A WITH MACRON
4108    <A(>        <U0102>    LATIN CAPITAL LETTER A WITH BREVE
4109    <a(>        <U0103>    LATIN SMALL LETTER A WITH BREVE
4110    <A;>        <U0104>    LATIN CAPITAL LETTER A WITH OGONEK
4111    <a;>        <U0105>    LATIN SMALL LETTER A WITH OGONEK
4112    <C'>        <U0106>    LATIN CAPITAL LETTER C WITH ACUTE
4113    <c'>        <U0107>    LATIN SMALL LETTER C WITH ACUTE
4114    <C/>>       <U0108>    LATIN CAPITAL LETTER C WITH CIRCUMFLEX
4115    <c/>>       <U0109>    LATIN SMALL LETTER C WITH CIRCUMFLEX
4116    <C.>        <U010A>    LATIN CAPITAL LETTER C WITH DOT ABOVE
4117    <c.>        <U010B>    LATIN SMALL LETTER C WITH DOT ABOVE
4118    <C<>        <U010C>    LATIN CAPITAL LETTER C WITH CARON
4119    <c<>        <U010D>    LATIN SMALL LETTER C WITH CARON
4120    <D<>        <U010E>    LATIN CAPITAL LETTER D WITH CARON
4121    <d<>        <U010F>    LATIN SMALL LETTER D WITH CARON
4122    <D//>       <U0110>    LATIN CAPITAL LETTER D WITH STROKE
4123    <d//>       <U0111>    LATIN SMALL LETTER D WITH STROKE
4124    <E->        <U0112>    LATIN CAPITAL LETTER E WITH MACRON
4125    <e->        <U0113>    LATIN SMALL LETTER E WITH MACRON
4126    <E(>        <U0114>    LATIN CAPITAL LETTER E WITH BREVE
4127    <e(>        <U0115>    LATIN SMALL LETTER E WITH BREVE
4128    <E.>        <U0116>    LATIN CAPITAL LETTER E WITH DOT ABOVE
4129    <e.>        <U0117>    LATIN SMALL LETTER E WITH DOT ABOVE
4130    <E;>        <U0118>    LATIN CAPITAL LETTER E WITH OGONEK
4131    <e;>        <U0119>    LATIN SMALL LETTER E WITH OGONEK
4132    <E<>        <U011A>    LATIN CAPITAL LETTER E WITH CARON
4133    <e<>        <U011B>    LATIN SMALL LETTER E WITH CARON
4134    <G/>>       <U011C>    LATIN CAPITAL LETTER G WITH CIRCUMFLEX
4135    <g/>>       <U011D>    LATIN SMALL LETTER G WITH CIRCUMFLEX
4136    <G(>        <U011E>    LATIN CAPITAL LETTER G WITH BREVE
4137    <g(>        <U011F>    LATIN SMALL LETTER G WITH BREVE
4138    <G.>        <U0120>    LATIN CAPITAL LETTER G WITH DOT ABOVE
4139    <g.>        <U0121>    LATIN SMALL LETTER G WITH DOT ABOVE
4140    <G,>        <U0122>    LATIN CAPITAL LETTER G WITH CEDILLA
4141    <g,>        <U0123>    LATIN SMALL LETTER G WITH CEDILLA
4142    <H/>>       <U0124>    LATIN CAPITAL LETTER H WITH CIRCUMFLEX
4143    <h/>>       <U0125>    LATIN SMALL LETTER H WITH CIRCUMFLEX
4144    <H//>       <U0126>    LATIN CAPITAL LETTER H WITH STROKE
4145    <h//>       <U0127>    LATIN SMALL LETTER H WITH STROKE
4146    <I?>        <U0128>    LATIN CAPITAL LETTER I WITH TILDE
4147    <i?>        <U0129>    LATIN SMALL LETTER I WITH TILDE
4148    <I->        <U012A>    LATIN CAPITAL LETTER I WITH MACRON
4149    <i->        <U012B>    LATIN SMALL LETTER I WITH MACRON
4150    <I(>        <U012C>    LATIN CAPITAL LETTER I WITH BREVE
4151    <i(>        <U012D>    LATIN SMALL LETTER I WITH BREVE
4152    <I;>        <U012E>    LATIN CAPITAL LETTER I WITH OGONEK
4153    <i;>        <U012F>    LATIN SMALL LETTER I WITH OGONEK
4154    <I.>        <U0130>    LATIN CAPITAL LETTER I WITH DOT ABOVE
4155    <i.>        <U0131>    LATIN SMALL LETTER DOTLESS I
4156    <IJ>        <U0132>    LATIN CAPITAL LIGATURE IJ
4157    <ij>        <U0133>    LATIN SMALL LIGATURE IJ
4158    <J/>>       <U0134>    LATIN CAPITAL LETTER J WITH CIRCUMFLEX
4159    <j/>>       <U0135>    LATIN SMALL LETTER J WITH CIRCUMFLEX
4160    <K,>        <U0136>    LATIN CAPITAL LETTER K WITH CEDILLA
4161    <k,>        <U0137>    LATIN SMALL LETTER K WITH CEDILLA
4162    <kk>        <U0138>    LATIN SMALL LETTER KRA (Greenlandic)
4163    <L'>        <U0139>    LATIN CAPITAL LETTER L WITH ACUTE
4164    <l'>        <U013A>    LATIN SMALL LETTER L WITH ACUTE
```

```
4165        <L,>        <U013B>    LATIN CAPITAL LETTER L WITH CEDILLA
4166        <l,>        <U013C>    LATIN SMALL LETTER L WITH CEDILLA
4167        <L<>        <U013D>    LATIN CAPITAL LETTER L WITH CARON
4168        <l<>        <U013E>    LATIN SMALL LETTER L WITH CARON
4169        <L.>        <U013F>    LATIN CAPITAL LETTER L WITH MIDDLE DOT
4170        <l.>        <U0140>    LATIN SMALL LETTER L WITH MIDDLE DOT
4171        <L//>       <U0141>    LATIN CAPITAL LETTER L WITH STROKE
4172        <l//>       <U0142>    LATIN SMALL LETTER L WITH STROKE
4173        <N'>        <U0143>    LATIN CAPITAL LETTER N WITH ACUTE
4174        <n'>        <U0144>    LATIN SMALL LETTER N WITH ACUTE
4175        <N,>        <U0145>    LATIN CAPITAL LETTER N WITH CEDILLA
4176        <n,>        <U0146>    LATIN SMALL LETTER N WITH CEDILLA
4177        <N<>        <U0147>    LATIN CAPITAL LETTER N WITH CARON
4178        <n<>        <U0148>    LATIN SMALL LETTER N WITH CARON
4179        <'n>        <U0149>    LATIN SMALL LETTER N PRECEDED BY APOSTROPHE
4180        <NG>        <U014A>    LATIN CAPITAL LETTER ENG (Sami)
4181        <ng>        <U014B>    LATIN SMALL LETTER ENG (Sami)
4182        <O->        <U014C>    LATIN CAPITAL LETTER O WITH MACRON
4183        <o->        <U014D>    LATIN SMALL LETTER O WITH MACRON
4184        <O(>        <U014E>    LATIN CAPITAL LETTER O WITH BREVE
4185        <o(>        <U014F>    LATIN SMALL LETTER O WITH BREVE
4186        <O">        <U0150>    LATIN CAPITAL LETTER O WITH DOUBLE ACUTE
4187        <o">        <U0151>    LATIN SMALL LETTER O WITH DOUBLE ACUTE
4188        <OE>        <U0152>    LATIN CAPITAL LIGATURE OE
4189        <oe>        <U0153>    LATIN SMALL LIGATURE OE
4190        <R'>        <U0154>    LATIN CAPITAL LETTER R WITH ACUTE
4191        <r'>        <U0155>    LATIN SMALL LETTER R WITH ACUTE
4192        <R,>        <U0156>    LATIN CAPITAL LETTER R WITH CEDILLA
4193        <r,>        <U0157>    LATIN SMALL LETTER R WITH CEDILLA
4194        <R<>        <U0158>    LATIN CAPITAL LETTER R WITH CARON
4195        <r<>        <U0159>    LATIN SMALL LETTER R WITH CARON
4196        <S'>        <U015A>    LATIN CAPITAL LETTER S WITH ACUTE
4197        <s'>        <U015B>    LATIN SMALL LETTER S WITH ACUTE
4198        <S/>>       <U015C>    LATIN CAPITAL LETTER S WITH CIRCUMFLEX
4199        <s/>>       <U015D>    LATIN SMALL LETTER S WITH CIRCUMFLEX
4200        <S,>        <U015E>    LATIN CAPITAL LETTER S WITH CEDILLA
4201        <s,>        <U015F>    LATIN SMALL LETTER S WITH CEDILLA
4202        <S<>        <U0160>    LATIN CAPITAL LETTER S WITH CARON
4203        <s<>        <U0161>    LATIN SMALL LETTER S WITH CARON
4204        <T,>        <U0162>    LATIN CAPITAL LETTER T WITH CEDILLA
4205        <t,>        <U0163>    LATIN SMALL LETTER T WITH CEDILLA
4206        <T<>        <U0164>    LATIN CAPITAL LETTER T WITH CARON
4207        <t<>        <U0165>    LATIN SMALL LETTER T WITH CARON
4208        <T//>       <U0166>    LATIN CAPITAL LETTER T WITH STROKE
4209        <t//>       <U0167>    LATIN SMALL LETTER T WITH STROKE
4210        <U?>        <U0168>    LATIN CAPITAL LETTER U WITH TILDE
4211        <u?>        <U0169>    LATIN SMALL LETTER U WITH TILDE
4212        <U->        <U016A>    LATIN CAPITAL LETTER U WITH MACRON
4213        <u->        <U016B>    LATIN SMALL LETTER U WITH MACRON
4214        <U(>        <U016C>    LATIN CAPITAL LETTER U WITH BREVE
4215        <u(>        <U016D>    LATIN SMALL LETTER U WITH BREVE
4216        <U0>        <U016E>    LATIN CAPITAL LETTER U WITH RING ABOVE
4217        <u0>        <U016F>    LATIN SMALL LETTER U WITH RING ABOVE
4218        <U">        <U0170>    LATIN CAPITAL LETTER U WITH DOUBLE ACUTE
4219        <u">        <U0171>    LATIN SMALL LETTER U WITH DOUBLE ACUTE
4220        <U;>        <U0172>    LATIN CAPITAL LETTER U WITH OGONEK
4221        <u;>        <U0173>    LATIN SMALL LETTER U WITH OGONEK
4222        <W/>>       <U0174>    LATIN CAPITAL LETTER W WITH CIRCUMFLEX
4223        <w/>>       <U0175>    LATIN SMALL LETTER W WITH CIRCUMFLEX
4224        <Y/>>       <U0176>    LATIN CAPITAL LETTER Y WITH CIRCUMFLEX
4225        <y/>>       <U0177>    LATIN SMALL LETTER Y WITH CIRCUMFLEX
4226        <Y:>        <U0178>    LATIN CAPITAL LETTER Y WITH DIAERESIS
4227        <Z'>        <U0179>    LATIN CAPITAL LETTER Z WITH ACUTE
4228        <z'>        <U017A>    LATIN SMALL LETTER Z WITH ACUTE
4229        <Z.>        <U017B>    LATIN CAPITAL LETTER Z WITH DOT ABOVE
4230        <z.>        <U017C>    LATIN SMALL LETTER Z WITH DOT ABOVE
4231        <Z<>        <U017D>    LATIN CAPITAL LETTER Z WITH CARON
4232        <z<>        <U017E>    LATIN SMALL LETTER Z WITH CARON
4233        <s1>        <U017F>    LATIN SMALL LETTER LONG S
4234        <b//>       <U0180>    LATIN SMALL LETTER B WITH STROKE
4235        <B2>        <U0181>    LATIN CAPITAL LETTER B WITH HOOK
4236        <C2>        <U0187>    LATIN CAPITAL LETTER C WITH HOOK
4237        <c2>        <U0188>    LATIN SMALL LETTER C WITH HOOK
4238        <F2>        <U0191>    LATIN CAPITAL LETTER F WITH HOOK
4239        <f2>        <U0192>    LATIN SMALL LETTER F WITH HOOK
4240        <K2>        <U0198>    LATIN CAPITAL LETTER K WITH HOOK
4241        <k2>        <U0199>    LATIN SMALL LETTER K WITH HOOK
4242        <O9>        <U01A0>    LATIN CAPITAL LETTER O WITH HORN
4243        <o9>        <U01A1>    LATIN SMALL LETTER O WITH HORN
4244        <OI>        <U01A2>    LATIN CAPITAL LETTER OI
4245        <oi>        <U01A3>    LATIN SMALL LETTER OI
4246        <yr>        <U01A6>    LATIN LETTER YR
4247        <U9>        <U01AF>    LATIN CAPITAL LETTER U WITH HORN
4248        <u9>        <U01B0>    LATIN SMALL LETTER U WITH HORN
4249        <Z//>       <U01B5>    LATIN CAPITAL LETTER Z WITH STROKE
4250        <z//>       <U01B6>    LATIN SMALL LETTER Z WITH STROKE
4251        <ED>        <U01B7>    LATIN CAPITAL LETTER EZH
4252        <DZ<>       <U01C4>    LATIN CAPITAL LETTER DZ WITH CARON
4253        <Dz<>       <U01C5>    LATIN CAPITAL LETTER D WITH SMALL LETTER Z WITH CARON
```

```
4254    <dz<>       <U01C6>    LATIN SMALL LETTER DZ WITH CARON
4255    <LJ3>       <U01C7>    LATIN CAPITAL LETTER LJ
4256    <Lj3>       <U01C8>    LATIN CAPITAL LETTER L WITH SMALL LETTER J
4257    <lj3>       <U01C9>    LATIN SMALL LETTER LJ
4258    <NJ3>       <U01CA>    LATIN CAPITAL LETTER NJ
4259    <Nj3>       <U01CB>    LATIN CAPITAL LETTER N WITH SMALL LETTER J
4260    <nj3>       <U01CC>    LATIN SMALL LETTER NJ
4261    <A<>        <U01CD>    LATIN CAPITAL LETTER A WITH CARON
4262    <a<>        <U01CE>    LATIN SMALL LETTER A WITH CARON
4263    <I<>        <U01CF>    LATIN CAPITAL LETTER I WITH CARON
4264    <i<>        <U01D0>    LATIN SMALL LETTER I WITH CARON
4265    <O<>        <U01D1>    LATIN CAPITAL LETTER O WITH CARON
4266    <o<>        <U01D2>    LATIN SMALL LETTER O WITH CARON
4267    <U<>        <U01D3>    LATIN CAPITAL LETTER U WITH CARON
4268    <u<>        <U01D4>    LATIN SMALL LETTER U WITH CARON
4269    <U:->       <U01D5>    LATIN CAPITAL LETTER U WITH DIAERESIS AND MACRON
4270    <u:->       <U01D6>    LATIN SMALL LETTER U WITH DIAERESIS AND MACRON
4271    <U:'>       <U01D7>    LATIN CAPITAL LETTER U WITH DIAERESIS AND ACUTE
4272    <u:'>       <U01D8>    LATIN SMALL LETTER U WITH DIAERESIS AND ACUTE
4273    <U:<>       <U01D9>    LATIN CAPITAL LETTER U WITH DIAERESIS AND CARON
4274    <u:<>       <U01DA>    LATIN SMALL LETTER U WITH DIAERESIS AND CARON
4275    <U:!>       <U01DB>    LATIN CAPITAL LETTER U WITH DIAERESIS AND GRAVE
4276    <u:!>       <U01DC>    LATIN SMALL LETTER U WITH DIAERESIS AND GRAVE
4277    <e1>        <U01DD>    LATIN SMALL LETTER TURNED E
4278    <A1>        <U01DE>    LATIN CAPITAL LETTER A WITH DIAERESIS AND MACRON
4279    <a1>        <U01DF>    LATIN SMALL LETTER A WITH DIAERESIS AND MACRON
4280    <A7>        <U01E0>    LATIN CAPITAL LETTER A WITH DOT ABOVE AND MACRON
4281    <a7>        <U01E1>    LATIN SMALL LETTER A WITH DOT ABOVE AND MACRON
4282    <A3>        <U01E2>    LATIN CAPITAL LETTER AE WITH MACRON (ash)
4283    <a3>        <U01E3>    LATIN SMALL LETTER AE WITH MACRON (ash)
4284    <G//>       <U01E4>    LATIN CAPITAL LETTER G WITH STROKE
4285    <g//>       <U01E5>    LATIN SMALL LETTER G WITH STROKE
4286    <G<>        <U01E6>    LATIN CAPITAL LETTER G WITH CARON
4287    <g<>        <U01E7>    LATIN SMALL LETTER G WITH CARON
4288    <K<>        <U01E8>    LATIN CAPITAL LETTER K WITH CARON
4289    <k<>        <U01E9>    LATIN SMALL LETTER K WITH CARON
4290    <O;>        <U01EA>    LATIN CAPITAL LETTER O WITH OGONEK
4291    <o;>        <U01EB>    LATIN SMALL LETTER O WITH OGONEK
4292    <O1>        <U01EC>    LATIN CAPITAL LETTER O WITH OGONEK AND MACRON
4293    <o1>        <U01ED>    LATIN SMALL LETTER O WITH OGONEK AND MACRON
4294    <EZ>        <U01EE>    LATIN CAPITAL LETTER EZH WITH CARON
4295    <ez>        <U01EF>    LATIN SMALL LETTER EZH WITH CARON
4296    <j<>        <U01F0>    LATIN SMALL LETTER J WITH CARON
4297    <DZ3>       <U01F1>    LATIN CAPITAL LETTER DZ
4298    <Dz3>       <U01F2>    LATIN CAPITAL LETTER D WITH SMALL LETTER Z
4299    <dz3>       <U01F3>    LATIN SMALL LETTER DZ
4300    <G'>        <U01F4>    LATIN CAPITAL LETTER G WITH ACUTE
4301    <g'>        <U01F5>    LATIN SMALL LETTER G WITH ACUTE
4302    <AA'>       <U01FA>    LATIN CAPITAL LETTER A WITH RING ABOVE AND ACUTE
4303    <aa'>       <U01FB>    LATIN SMALL LETTER A WITH RING ABOVE AND ACUTE
4304    <AE'>       <U01FC>    LATIN CAPITAL LETTER AE WITH ACUTE (ash)
4305    <ae'>       <U01FD>    LATIN SMALL LETTER AE WITH ACUTE (ash)
4306    <O//'>      <U01FE>    LATIN CAPITAL LETTER O WITH STROKE AND ACUTE
4307    <o//'>      <U01FF>    LATIN SMALL LETTER O WITH STROKE AND ACUTE
4308    <A!!>       <U0200>    LATIN CAPITAL LETTER A WITH DOUBLE GRAVE
4309    <a!!>       <U0201>    LATIN SMALL LETTER A WITH DOUBLE GRAVE
4310    <A)>        <U0202>    LATIN CAPITAL LETTER A WITH INVERTED BREVE
4311    <a)>        <U0203>    LATIN SMALL LETTER A WITH INVERTED BREVE
4312    <E!!>       <U0204>    LATIN CAPITAL LETTER E WITH DOUBLE GRAVE
4313    <e!!>       <U0205>    LATIN SMALL LETTER E WITH DOUBLE GRAVE
4314    <E)>        <U0206>    LATIN CAPITAL LETTER E WITH INVERTED BREVE
4315    <e)>        <U0207>    LATIN SMALL LETTER E WITH INVERTED BREVE
4316    <I!!>       <U0208>    LATIN CAPITAL LETTER I WITH DOUBLE GRAVE
4317    <i!!>       <U0209>    LATIN SMALL LETTER I WITH DOUBLE GRAVE
4318    <I)>        <U020A>    LATIN CAPITAL LETTER I WITH INVERTED BREVE
4319    <i)>        <U020B>    LATIN SMALL LETTER I WITH INVERTED BREVE
4320    <O!>        <U020C>    LATIN CAPITAL LETTER O WITH DOUBLE GRAVE
4321    <o!!>       <U020D>    LATIN SMALL LETTER O WITH DOUBLE GRAVE
4322    <O)>        <U020E>    LATIN CAPITAL LETTER O WITH INVERTED BREVE
4323    <o)>        <U020F>    LATIN SMALL LETTER O WITH INVERTED BREVE
4324    <R!>        <U0210>    LATIN CAPITAL LETTER R WITH DOUBLE GRAVE
4325    <r!!>       <U0211>    LATIN SMALL LETTER R WITH DOUBLE GRAVE
4326    <R)>        <U0212>    LATIN CAPITAL LETTER R WITH INVERTED BREVE
4327    <r)>        <U0213>    LATIN SMALL LETTER R WITH INVERTED BREVE
4328    <U!!>       <U0214>    LATIN CAPITAL LETTER U WITH DOUBLE GRAVE
4329    <u!!>       <U0215>    LATIN SMALL LETTER U WITH DOUBLE GRAVE
4330    <U)>        <U0216>    LATIN CAPITAL LETTER U WITH INVERTED BREVE
4331    <u)>        <U0217>    LATIN SMALL LETTER U WITH INVERTED BREVE
4332    <r1>        <U027C>    LATIN SMALL LETTER R WITH LONG LEG
4333    <ed>        <U0292>    LATIN SMALL LETTER EZH
4334    <;S>        <U02BB>    MODIFIER LETTER TURNED COMMA
4335    <1/>>       <U02C6>    MODIFIER LETTER CIRCUMFLEX ACCENT
4336    <'<>        <U02C7>    CARON (Mandarin Chinese third tone)
4337    <1->        <U02C9>    MODIFIER LETTER MACRON (Mandarin Chinese first tone)
4338    <1!>        <U02CB>    MODIFIER LETTER GRAVE ACCENT (Mandarin Chinese fourth tone)
4339    <'(>        <U02D8>    BREVE
4340    <'.>        <U02D9>    DOT ABOVE (Mandarin Chinese light tone)
4341    <'0>        <U02DA>    RING ABOVE
```

```
4342        <';>        <U02DB>        OGONEK
4343        <1?>        <U02DC>        SMALL TILDE
4344        <'">        <U02DD>        DOUBLE ACUTE ACCENT
4345        <'G>        <U0374>        GREEK NUMERAL SIGN (Dexia keraia)
4346        <,G>        <U0375>        GREEK LOWER NUMERAL SIGN (Aristeri keraia)
4347        <j3>        <U037A>        GREEK YPOGEGRAMMENI
4348        <?%>        <U037E>        GREEK QUESTION MARK (Erotimatiko)
4349        <'*>        <U0384>        GREEK TONOS
4350        <'%>        <U0385>        GREEK DIALYTIKA TONOS
4351        <A%>        <U0386>        GREEK CAPITAL LETTER ALPHA WITH TONOS
4352        <.*>        <U0387>        GREEK ANO TELEIA
4353        <E%>        <U0388>        GREEK CAPITAL LETTER EPSILON WITH TONOS
4354        <Y%>        <U0389>        GREEK CAPITAL LETTER ETA WITH TONOS
4355        <I%>        <U038A>        GREEK CAPITAL LETTER IOTA WITH TONOS
4356        <O%>        <U038C>        GREEK CAPITAL LETTER OMICRON WITH TONOS
4357        <U%>        <U038E>        GREEK CAPITAL LETTER UPSILON WITH TONOS
4358        <W%>        <U038F>        GREEK CAPITAL LETTER OMEGA WITH TONOS
4359        <i3>        <U0390>        GREEK SMALL LETTER IOTA WITH DIALYTIKA AND TONOS
4360        <A*>        <U0391>        GREEK CAPITAL LETTER ALPHA
4361        <B*>        <U0392>        GREEK CAPITAL LETTER BETA
4362        <G*>        <U0393>        GREEK CAPITAL LETTER GAMMA
4363        <D*>        <U0394>        GREEK CAPITAL LETTER DELTA
4364        <E*>        <U0395>        GREEK CAPITAL LETTER EPSILON
4365        <Z*>        <U0396>        GREEK CAPITAL LETTER ZETA
4366        <Y*>        <U0397>        GREEK CAPITAL LETTER ETA
4367        <H*>        <U0398>        GREEK CAPITAL LETTER THETA
4368        <I*>        <U0399>        GREEK CAPITAL LETTER IOTA
4369        <K*>        <U039A>        GREEK CAPITAL LETTER KAPPA
4370        <L*>        <U039B>        GREEK CAPITAL LETTER LAMDA
4371        <M*>        <U039C>        GREEK CAPITAL LETTER MU
4372        <N*>        <U039D>        GREEK CAPITAL LETTER NU
4373        <C*>        <U039E>        GREEK CAPITAL LETTER XI
4374        <O*>        <U039F>        GREEK CAPITAL LETTER OMICRON
4375        <P*>        <U03A0>        GREEK CAPITAL LETTER PI
4376        <R*>        <U03A1>        GREEK CAPITAL LETTER RHO
4377        <S*>        <U03A3>        GREEK CAPITAL LETTER SIGMA
4378        <T*>        <U03A4>        GREEK CAPITAL LETTER TAU
4379        <U*>        <U03A5>        GREEK CAPITAL LETTER UPSILON
4380        <F*>        <U03A6>        GREEK CAPITAL LETTER PHI
4381        <X*>        <U03A7>        GREEK CAPITAL LETTER CHI
4382        <Q*>        <U03A8>        GREEK CAPITAL LETTER PSI
4383        <W*>        <U03A9>        GREEK CAPITAL LETTER OMEGA
4384        <J*>        <U03AA>        GREEK CAPITAL LETTER IOTA WITH DIALYTIKA
4385        <V*>        <U03AB>        GREEK CAPITAL LETTER UPSILON WITH DIALYTIKA
4386        <a%>        <U03AC>        GREEK SMALL LETTER ALPHA WITH TONOS
4387        <e%>        <U03AD>        GREEK SMALL LETTER EPSILON WITH TONOS
4388        <y%>        <U03AE>        GREEK SMALL LETTER ETA WITH TONOS
4389        <i%>        <U03AF>        GREEK SMALL LETTER IOTA WITH TONOS
4390        <u3>        <U03B0>        GREEK SMALL LETTER UPSILON WITH DIALYTIKA AND TONOS
4391        <a*>        <U03B1>        GREEK SMALL LETTER ALPHA
4392        <b*>        <U03B2>        GREEK SMALL LETTER BETA
4393        <g*>        <U03B3>        GREEK SMALL LETTER GAMMA
4394        <d*>        <U03B4>        GREEK SMALL LETTER DELTA
4395        <e*>        <U03B5>        GREEK SMALL LETTER EPSILON
4396        <z*>        <U03B6>        GREEK SMALL LETTER ZETA
4397        <y*>        <U03B7>        GREEK SMALL LETTER ETA
4398        <h*>        <U03B8>        GREEK SMALL LETTER THETA
4399        <i*>        <U03B9>        GREEK SMALL LETTER IOTA
4400        <k*>        <U03BA>        GREEK SMALL LETTER KAPPA
4401        <l*>        <U03BB>        GREEK SMALL LETTER LAMDA
4402        <m*>        <U03BC>        GREEK SMALL LETTER MU
4403        <n*>        <U03BD>        GREEK SMALL LETTER NU
4404        <c*>        <U03BE>        GREEK SMALL LETTER XI
4405        <o*>        <U03BF>        GREEK SMALL LETTER OMICRON
4406        <p*>        <U03C0>        GREEK SMALL LETTER PI
4407        <r*>        <U03C1>        GREEK SMALL LETTER RHO
4408        <*s>        <U03C2>        GREEK SMALL LETTER FINAL SIGMA
4409        <s*>        <U03C3>        GREEK SMALL LETTER SIGMA
4410        <t*>        <U03C4>        GREEK SMALL LETTER TAU
4411        <u*>        <U03C5>        GREEK SMALL LETTER UPSILON
4412        <f*>        <U03C6>        GREEK SMALL LETTER PHI
4413        <x*>        <U03C7>        GREEK SMALL LETTER CHI
4414        <q*>        <U03C8>        GREEK SMALL LETTER PSI
4415        <w*>        <U03C9>        GREEK SMALL LETTER OMEGA
4416        <j*>        <U03CA>        GREEK SMALL LETTER IOTA WITH DIALYTIKA
4417        <v*>        <U03CB>        GREEK SMALL LETTER UPSILON WITH DIALYTIKA
4418        <o%>        <U03CC>        GREEK SMALL LETTER OMICRON WITH TONOS
4419        <u%>        <U03CD>        GREEK SMALL LETTER UPSILON WITH TONOS
4420        <w%>        <U03CE>        GREEK SMALL LETTER OMEGA WITH TONOS
4421        <b3>        <U03D0>        GREEK BETA SYMBOL
4422        <T3>        <U03DA>        GREEK LETTER STIGMA
4423        <M3>        <U03DC>        GREEK LETTER DIGAMMA
4424        <K3>        <U03DE>        GREEK LETTER KOPPA
4425        <P3>        <U03E0>        GREEK LETTER SAMPI
4426        <IO>        <U0401>        CYRILLIC CAPITAL LETTER IO
4427        <D%>        <U0402>        CYRILLIC CAPITAL LETTER DJE (Serbocroatian)
4428        <G%>        <U0403>        CYRILLIC CAPITAL LETTER GJE
4429        <IE>        <U0404>        CYRILLIC CAPITAL LETTER UKRAINIAN IE
4430        <DS>        <U0405>        CYRILLIC CAPITAL LETTER DZE
```

70

```
4431       <II>        <U0406>      CYRILLIC CAPITAL LETTER BYELORUSSIAN-UKRAINIAN I
4432       <YI>        <U0407>      CYRILLIC CAPITAL LETTER YI (Ukrainian)
4433       <J%>        <U0408>      CYRILLIC CAPITAL LETTER JE
4434       <LJ>        <U0409>      CYRILLIC CAPITAL LETTER LJE
4435       <NJ>        <U040A>      CYRILLIC CAPITAL LETTER NJE
4436       <Ts>        <U040B>      CYRILLIC CAPITAL LETTER TSHE (Serbocroatian)
4437       <KJ>        <U040C>      CYRILLIC CAPITAL LETTER KJE
4438       <V%>        <U040E>      CYRILLIC CAPITAL LETTER SHORT U (Byelorussian)
4439       <DZ>        <U040F>      CYRILLIC CAPITAL LETTER DZHE
4440       <A=>        <U0410>      CYRILLIC CAPITAL LETTER A
4441       <B=>        <U0411>      CYRILLIC CAPITAL LETTER BE
4442       <V=>        <U0412>      CYRILLIC CAPITAL LETTER VE
4443       <G=>        <U0413>      CYRILLIC CAPITAL LETTER GHE
4444       <D=>        <U0414>      CYRILLIC CAPITAL LETTER DE
4445       <E=>        <U0415>      CYRILLIC CAPITAL LETTER IE
4446       <Z%>        <U0416>      CYRILLIC CAPITAL LETTER ZHE
4447       <Z=>        <U0417>      CYRILLIC CAPITAL LETTER ZE
4448       <I=>        <U0418>      CYRILLIC CAPITAL LETTER I
4449       <J=>        <U0419>      CYRILLIC CAPITAL LETTER SHORT I
4450       <K=>        <U041A>      CYRILLIC CAPITAL LETTER KA
4451       <L=>        <U041B>      CYRILLIC CAPITAL LETTER EL
4452       <M=>        <U041C>      CYRILLIC CAPITAL LETTER EM
4453       <N=>        <U041D>      CYRILLIC CAPITAL LETTER EN
4454       <O=>        <U041E>      CYRILLIC CAPITAL LETTER O
4455       <P=>        <U041F>      CYRILLIC CAPITAL LETTER PE
4456       <R=>        <U0420>      CYRILLIC CAPITAL LETTER ER
4457       <S=>        <U0421>      CYRILLIC CAPITAL LETTER ES
4458       <T=>        <U0422>      CYRILLIC CAPITAL LETTER TE
4459       <U=>        <U0423>      CYRILLIC CAPITAL LETTER U
4460       <F=>        <U0424>      CYRILLIC CAPITAL LETTER EF
4461       <H=>        <U0425>      CYRILLIC CAPITAL LETTER HA
4462       <C=>        <U0426>      CYRILLIC CAPITAL LETTER TSE
4463       <C%>        <U0427>      CYRILLIC CAPITAL LETTER CHE
4464       <S%>        <U0428>      CYRILLIC CAPITAL LETTER SHA
4465       <Sc>        <U0429>      CYRILLIC CAPITAL LETTER SHCHA
4466       <=">        <U042A>      CYRILLIC CAPITAL LETTER HARD SIGN
4467       <Y=>        <U042B>      CYRILLIC CAPITAL LETTER YERU
4468       <%">        <U042C>      CYRILLIC CAPITAL LETTER SOFT SIGN
4469       <JE>        <U042D>      CYRILLIC CAPITAL LETTER E
4470       <JU>        <U042E>      CYRILLIC CAPITAL LETTER YU
4471       <JA>        <U042F>      CYRILLIC CAPITAL LETTER YA
4472       <a=>        <U0430>      CYRILLIC SMALL LETTER A
4473       <b=>        <U0431>      CYRILLIC SMALL LETTER BE
4474       <v=>        <U0432>      CYRILLIC SMALL LETTER VE
4475       <g=>        <U0433>      CYRILLIC SMALL LETTER GHE
4476       <d=>        <U0434>      CYRILLIC SMALL LETTER DE
4477       <e=>        <U0435>      CYRILLIC SMALL LETTER IE
4478       <z%>        <U0436>      CYRILLIC SMALL LETTER ZHE
4479       <z=>        <U0437>      CYRILLIC SMALL LETTER ZE
4480       <i=>        <U0438>      CYRILLIC SMALL LETTER I
4481       <j=>        <U0439>      CYRILLIC SMALL LETTER SHORT I
4482       <k=>        <U043A>      CYRILLIC SMALL LETTER KA
4483       <l=>        <U043B>      CYRILLIC SMALL LETTER EL
4484       <m=>        <U043C>      CYRILLIC SMALL LETTER EM
4485       <n=>        <U043D>      CYRILLIC SMALL LETTER EN
4486       <o=>        <U043E>      CYRILLIC SMALL LETTER O
4487       <p=>        <U043F>      CYRILLIC SMALL LETTER PE
4488       <r=>        <U0440>      CYRILLIC SMALL LETTER ER
4489       <s=>        <U0441>      CYRILLIC SMALL LETTER ES
4490       <t=>        <U0442>      CYRILLIC SMALL LETTER TE
4491       <u=>        <U0443>      CYRILLIC SMALL LETTER U
4492       <f=>        <U0444>      CYRILLIC SMALL LETTER EF
4493       <h=>        <U0445>      CYRILLIC SMALL LETTER HA
4494       <c=>        <U0446>      CYRILLIC SMALL LETTER TSE
4495       <c%>        <U0447>      CYRILLIC SMALL LETTER CHE
4496       <s%>        <U0448>      CYRILLIC SMALL LETTER SHA
4497       <sc>        <U0449>      CYRILLIC SMALL LETTER SHCHA
4498       <='>        <U044A>      CYRILLIC SMALL LETTER HARD SIGN
4499       <y=>        <U044B>      CYRILLIC SMALL LETTER YERU
4500       <%'>        <U044C>      CYRILLIC SMALL LETTER SOFT SIGN
4501       <je>        <U044D>      CYRILLIC SMALL LETTER E
4502       <ju>        <U044E>      CYRILLIC SMALL LETTER YU
4503       <ja>        <U044F>      CYRILLIC SMALL LETTER YA
4504       <io>        <U0451>      CYRILLIC SMALL LETTER IO
4505       <d%>        <U0452>      CYRILLIC SMALL LETTER DJE (Serbocroatian)
4506       <g%>        <U0453>      CYRILLIC SMALL LETTER GJE
4507       <ie>        <U0454>      CYRILLIC SMALL LETTER UKRAINIAN IE
4508       <ds>        <U0455>      CYRILLIC SMALL LETTER DZE
4509       <ii>        <U0456>      CYRILLIC SMALL LETTER BYELORUSSIAN-UKRAINIAN I
4510       <yi>        <U0457>      CYRILLIC SMALL LETTER YI (Ukrainian)
4511       <j%>        <U0458>      CYRILLIC SMALL LETTER JE
4512       <lj>        <U0459>      CYRILLIC SMALL LETTER LJE
4513       <nj>        <U045A>      CYRILLIC SMALL LETTER NJE
4514       <ts>        <U045B>      CYRILLIC SMALL LETTER TSHE (Serbocroatian)
4515       <kj>        <U045C>      CYRILLIC SMALL LETTER KJE
4516       <v%>        <U045E>      CYRILLIC SMALL LETTER SHORT U (Byelorussian)
4517       <dz>        <U045F>      CYRILLIC SMALL LETTER DZHE
4518       <Y3>        <U0462>      CYRILLIC CAPITAL LETTER YAT
```

```
4519        <y3>        <U0463>     CYRILLIC SMALL LETTER YAT
4520        <O3>        <U046A>     CYRILLIC CAPITAL LETTER BIG YUS
4521        <o3>        <U046B>     CYRILLIC SMALL LETTER BIG YUS
4522        <F3>        <U0472>     CYRILLIC CAPITAL LETTER FITA
4523        <f3>        <U0473>     CYRILLIC SMALL LETTER FITA
4524        <V3>        <U0474>     CYRILLIC CAPITAL LETTER IZHITSA
4525        <v3>        <U0475>     CYRILLIC SMALL LETTER IZHITSA
4526        <C3>        <U0480>     CYRILLIC CAPITAL LETTER KOPPA
4527        <c3>        <U0481>     CYRILLIC SMALL LETTER KOPPA
4528        <G3>        <U0490>     CYRILLIC CAPITAL LETTER GHE WITH UPTURN
4529        <g3>        <U0491>     CYRILLIC SMALL LETTER GHE WITH UPTURN
4530        <A+>        <U05D0>     HEBREW LETTER ALEF
4531        <B+>        <U05D1>     HEBREW LETTER BET
4532        <G+>        <U05D2>     HEBREW LETTER GIMEL
4533        <D+>        <U05D3>     HEBREW LETTER DALET
4534        <H+>        <U05D4>     HEBREW LETTER HE
4535        <W+>        <U05D5>     HEBREW LETTER VAV
4536        <Z+>        <U05D6>     HEBREW LETTER ZAYIN
4537        <X+>        <U05D7>     HEBREW LETTER HET
4538        <Tj>        <U05D8>     HEBREW LETTER TET
4539        <J+>        <U05D9>     HEBREW LETTER YOD
4540        <K%>        <U05DA>     HEBREW LETTER FINAL KAF
4541        <K+>        <U05DB>     HEBREW LETTER KAF
4542        <L+>        <U05DC>     HEBREW LETTER LAMED
4543        <M%>        <U05DD>     HEBREW LETTER FINAL MEM
4544        <M+>        <U05DE>     HEBREW LETTER MEM
4545        <N%>        <U05DF>     HEBREW LETTER FINAL NUN
4546        <N+>        <U05E0>     HEBREW LETTER NUN
4547        <S+>        <U05E1>     HEBREW LETTER SAMEKH
4548        <E+>        <U05E2>     HEBREW LETTER AYIN
4549        <P%>        <U05E3>     HEBREW LETTER FINAL PE
4550        <P+>        <U05E4>     HEBREW LETTER PE
4551        <Zj>        <U05E5>     HEBREW LETTER FINAL TSADI
4552        <ZJ>        <U05E6>     HEBREW LETTER TSADI
4553        <Q+>        <U05E7>     HEBREW LETTER QOF
4554        <R+>        <U05E8>     HEBREW LETTER RESH
4555        <Sh>        <U05E9>     HEBREW LETTER SHIN
4556        <T+>        <U05EA>     HEBREW LETTER TAV
4557        <,+>        <U060C>     ARABIC COMMA
4558        <;+>        <U061B>     ARABIC SEMICOLON
4559        <?+>        <U061F>     ARABIC QUESTION MARK
4560        <H'>        <U0621>     ARABIC LETTER HAMZA
4561        <aM>        <U0622>     ARABIC LETTER ALEF WITH MADDA ABOVE
4562        <aH>        <U0623>     ARABIC LETTER ALEF WITH HAMZA ABOVE
4563        <wH>        <U0624>     ARABIC LETTER WAW WITH HAMZA ABOVE
4564        <ah>        <U0625>     ARABIC LETTER ALEF WITH HAMZA BELOW
4565        <yH>        <U0626>     ARABIC LETTER YEH WITH HAMZA ABOVE
4566        <a+>        <U0627>     ARABIC LETTER ALEF
4567        <b+>        <U0628>     ARABIC LETTER BEH
4568        <tm>        <U0629>     ARABIC LETTER TEH MARBUTA
4569        <t+>        <U062A>     ARABIC LETTER TEH
4570        <tk>        <U062B>     ARABIC LETTER THEH
4571        <g+>        <U062C>     ARABIC LETTER JEEM
4572        <hk>        <U062D>     ARABIC LETTER HAH
4573        <x+>        <U062E>     ARABIC LETTER KHAH
4574        <d+>        <U062F>     ARABIC LETTER DAL
4575        <dk>        <U0630>     ARABIC LETTER THAL
4576        <r+>        <U0631>     ARABIC LETTER REH
4577        <z+>        <U0632>     ARABIC LETTER ZAIN
4578        <s+>        <U0633>     ARABIC LETTER SEEN
4579        <sn>        <U0634>     ARABIC LETTER SHEEN
4580        <c+>        <U0635>     ARABIC LETTER SAD
4581        <dd>        <U0636>     ARABIC LETTER DAD
4582        <tj>        <U0637>     ARABIC LETTER TAH
4583        <zH>        <U0638>     ARABIC LETTER ZAH
4584        <e+>        <U0639>     ARABIC LETTER AIN
4585        <i+>        <U063A>     ARABIC LETTER GHAIN
4586        <++>        <U0640>     ARABIC TATWEEL
4587        <f+>        <U0641>     ARABIC LETTER FEH
4588        <q+>        <U0642>     ARABIC LETTER QAF
4589        <k+>        <U0643>     ARABIC LETTER KAF
4590        <l+>        <U0644>     ARABIC LETTER LAM
4591        <m+>        <U0645>     ARABIC LETTER MEEM
4592        <n+>        <U0646>     ARABIC LETTER NOON
4593        <h+>        <U0647>     ARABIC LETTER HEH
4594        <w+>        <U0648>     ARABIC LETTER WAW
4595        <j+>        <U0649>     ARABIC LETTER ALEF MAKSURA
4596        <y+>        <U064A>     ARABIC LETTER YEH
4597        <:+>        <U064B>     ARABIC FATHATAN
4598        <"+>        <U064C>     ARABIC DAMMATAN
4599        <=+>        <U064D>     ARABIC KASRATAN
4600        <//+>       <U064E>     ARABIC FATHA
4601        <'+>        <U064F>     ARABIC DAMMA
4602        <1+>        <U0650>     ARABIC KASRA
4603        <3+>        <U0651>     ARABIC SHADDA
4604        <0+>        <U0652>     ARABIC SUKUN
4605        <0a>        <U0660>     ARABIC-INDIC DIGIT ZERO
4606        <1a>        <U0661>     ARABIC-INDIC DIGIT ONE
4607        <2a>        <U0662>     ARABIC-INDIC DIGIT TWO
```

```
4608        <3a>      <U0663>   ARABIC-INDIC DIGIT THREE
4609        <4a>      <U0664>   ARABIC-INDIC DIGIT FOUR
4610        <5a>      <U0665>   ARABIC-INDIC DIGIT FIVE
4611        <6a>      <U0666>   ARABIC-INDIC DIGIT SIX
4612        <7a>      <U0667>   ARABIC-INDIC DIGIT SEVEN
4613        <8a>      <U0668>   ARABIC-INDIC DIGIT EIGHT
4614        <9a>      <U0669>   ARABIC-INDIC DIGIT NINE
4615        <aS>      <U0670>   ARABIC LETTER SUPERSCRIPT ALEF
4616        <p+>      <U067E>   ARABIC LETTER PEH
4617        <hH>      <U0681>   ARABIC LETTER HAH WITH HAMZA ABOVE
4618        <tc>      <U0686>   ARABIC LETTER TCHEH
4619        <zj>      <U0698>   ARABIC LETTER JEH
4620        <v+>      <U06A4>   ARABIC LETTER VEH
4621        <gf>      <U06AF>   ARABIC LETTER GAF
4622        <A-0>     <U1E00>   LATIN CAPITAL LETTER A WITH RING BELOW
4623        <a-0>     <U1E01>   LATIN SMALL LETTER A WITH RING BELOW
4624        <B.>      <U1E02>   LATIN CAPITAL LETTER B WITH DOT ABOVE
4625        <b.>      <U1E03>   LATIN SMALL LETTER B WITH DOT ABOVE
4626        <B-.>     <U1E04>   LATIN CAPITAL LETTER B WITH DOT BELOW
4627        <b-.>     <U1E05>   LATIN SMALL LETTER B WITH DOT BELOW
4628        <B_>      <U1E06>   LATIN CAPITAL LETTER B WITH LINE BELOW
4629        <b_>      <U1E07>   LATIN SMALL LETTER B WITH LINE BELOW
4630        <C,'>     <U1E08>   LATIN CAPITAL LETTER C WITH CEDILLA AND ACUTE
4631        <c,'>     <U1E09>   LATIN SMALL LETTER C WITH CEDILLA AND ACUTE
4632        <D.>      <U1E0A>   LATIN CAPITAL LETTER D WITH DOT ABOVE
4633        <d.>      <U1E0B>   LATIN SMALL LETTER D WITH DOT ABOVE
4634        <D-.>     <U1E0C>   LATIN CAPITAL LETTER D WITH DOT BELOW
4635        <d-.>     <U1E0D>   LATIN SMALL LETTER D WITH DOT BELOW
4636        <D_>      <U1E0E>   LATIN CAPITAL LETTER D WITH LINE BELOW
4637        <d_>      <U1E0F>   LATIN SMALL LETTER D WITH LINE BELOW
4638        <D,>      <U1E10>   LATIN CAPITAL LETTER D WITH CEDILLA
4639        <d,>      <U1E11>   LATIN SMALL LETTER D WITH CEDILLA
4640        <D-/>>    <U1E12>   LATIN CAPITAL LETTER D WITH CIRCUMFLEX BELOW
4641        <d-/>>    <U1E13>   LATIN SMALL LETTER D WITH CIRCUMFLEX BELOW
4642        <E-!>     <U1E14>   LATIN CAPITAL LETTER E WITH MACRON AND GRAVE
4643        <e-!>     <U1E15>   LATIN SMALL LETTER E WITH MACRON AND GRAVE
4644        <E-'>     <U1E16>   LATIN CAPITAL LETTER E WITH MACRON AND ACUTE
4645        <e-'>     <U1E17>   LATIN SMALL LETTER E WITH MACRON AND ACUTE
4646        <E-/>>    <U1E18>   LATIN CAPITAL LETTER E WITH CIRCUMFLEX BELOW
4647        <e-/>>    <U1E19>   LATIN SMALL LETTER E WITH CIRCUMFLEX BELOW
4648        <E-?>     <U1E1A>   LATIN CAPITAL LETTER E WITH TILDE BELOW
4649        <e-?>     <U1E1B>   LATIN SMALL LETTER E WITH TILDE BELOW
4650        <E,(>     <U1E1C>   LATIN CAPITAL LETTER E WITH CEDILLA AND BREVE
4651        <e,(>     <U1E1D>   LATIN SMALL LETTER E WITH CEDILLA AND BREVE
4652        <F.>      <U1E1E>   LATIN CAPITAL LETTER F WITH DOT ABOVE
4653        <f.>      <U1E1F>   LATIN SMALL LETTER F WITH DOT ABOVE
4654        <G->      <U1E20>   LATIN CAPITAL LETTER G WITH MACRON
4655        <g->      <U1E21>   LATIN SMALL LETTER G WITH MACRON
4656        <H.>      <U1E22>   LATIN CAPITAL LETTER H WITH DOT ABOVE
4657        <h.>      <U1E23>   LATIN SMALL LETTER H WITH DOT ABOVE
4658        <H-.>     <U1E24>   LATIN CAPITAL LETTER H WITH DOT BELOW
4659        <h-.>     <U1E25>   LATIN SMALL LETTER H WITH DOT BELOW
4660        <H:>      <U1E26>   LATIN CAPITAL LETTER H WITH DIAERESIS
4661        <h:>      <U1E27>   LATIN SMALL LETTER H WITH DIAERESIS
4662        <H,>      <U1E28>   LATIN CAPITAL LETTER H WITH CEDILLA
4663        <h,>      <U1E29>   LATIN SMALL LETTER H WITH CEDILLA
4664        <H-(>     <U1E2A>   LATIN CAPITAL LETTER H WITH BREVE BELOW
4665        <h-(>     <U1E2B>   LATIN SMALL LETTER H WITH BREVE BELOW
4666        <I-?>     <U1E2C>   LATIN CAPITAL LETTER I WITH TILDE BELOW
4667        <i-?>     <U1E2D>   LATIN SMALL LETTER I WITH TILDE BELOW
4668        <I:'>     <U1E2E>   LATIN CAPITAL LETTER I WITH DIAERESIS AND ACUTE
4669        <i:'>     <U1E2F>   LATIN SMALL LETTER I WITH DIAERESIS AND ACUTE
4670        <K'>      <U1E30>   LATIN CAPITAL LETTER K WITH ACUTE
4671        <k'>      <U1E31>   LATIN SMALL LETTER K WITH ACUTE
4672        <K-.>     <U1E32>   LATIN CAPITAL LETTER K WITH DOT BELOW
4673        <k-.>     <U1E33>   LATIN SMALL LETTER K WITH DOT BELOW
4674        <K_>      <U1E34>   LATIN CAPITAL LETTER K WITH LINE BELOW
4675        <k_>      <U1E35>   LATIN SMALL LETTER K WITH LINE BELOW
4676        <L-.>     <U1E36>   LATIN CAPITAL LETTER L WITH DOT BELOW
4677        <l-.>     <U1E37>   LATIN SMALL LETTER L WITH DOT BELOW
4678        <L--.>    <U1E38>   LATIN CAPITAL LETTER L WITH DOT BELOW AND MACRON
4679        <l--.>    <U1E39>   LATIN SMALL LETTER L WITH DOT BELOW AND MACRON
4680        <L_>      <U1E3A>   LATIN CAPITAL LETTER L WITH LINE BELOW
4681        <l_>      <U1E3B>   LATIN SMALL LETTER L WITH LINE BELOW
4682        <L-/>>    <U1E3C>   LATIN CAPITAL LETTER L WITH CIRCUMFLEX BELOW
4683        <l-/>>    <U1E3D>   LATIN SMALL LETTER L WITH CIRCUMFLEX BELOW
4684        <M'>      <U1E3E>   LATIN CAPITAL LETTER M WITH ACUTE
4685        <m'>      <U1E3F>   LATIN SMALL LETTER M WITH ACUTE
4686        <M.>      <U1E40>   LATIN CAPITAL LETTER M WITH DOT ABOVE
4687        <m.>      <U1E41>   LATIN SMALL LETTER M WITH DOT ABOVE
4688        <M-.>     <U1E42>   LATIN CAPITAL LETTER M WITH DOT BELOW
4689        <m-.>     <U1E43>   LATIN SMALL LETTER M WITH DOT BELOW
4690        <N.>      <U1E44>   LATIN CAPITAL LETTER N WITH DOT ABOVE
4691        <n.>      <U1E45>   LATIN SMALL LETTER N WITH DOT ABOVE
4692        <N-.>     <U1E46>   LATIN CAPITAL LETTER N WITH DOT BELOW
4693        <n-.>     <U1E47>   LATIN SMALL LETTER N WITH DOT BELOW
4694        <N_>      <U1E48>   LATIN CAPITAL LETTER N WITH LINE BELOW
4695        <n_>      <U1E49>   LATIN SMALL LETTER N WITH LINE BELOW
```

```
4696        <N-/>>      <U1E4A>    LATIN CAPITAL LETTER N WITH CIRCUMFLEX BELOW
4697        <n-/>>      <U1E4B>    LATIN SMALL LETTER N WITH CIRCUMFLEX BELOW
4698        <O?'>       <U1E4C>    LATIN CAPITAL LETTER O WITH TILDE AND ACUTE
4699        <o?'>       <U1E4D>    LATIN SMALL LETTER O WITH TILDE AND ACUTE
4700        <O?:>       <U1E4E>    LATIN CAPITAL LETTER O WITH TILDE AND DIAERESIS
4701        <o?:>       <U1E4F>    LATIN SMALL LETTER O WITH TILDE AND DIAERESIS
4702        <O-!>       <U1E50>    LATIN CAPITAL LETTER O WITH MACRON AND GRAVE
4703        <o-!>       <U1E51>    LATIN SMALL LETTER O WITH MACRON AND GRAVE
4704        <O-'>       <U1E52>    LATIN CAPITAL LETTER O WITH MACRON AND ACUTE
4705        <o-'>       <U1E53>    LATIN SMALL LETTER O WITH MACRON AND ACUTE
4706        <P'>        <U1E54>    LATIN CAPITAL LETTER P WITH ACUTE
4707        <p'>        <U1E55>    LATIN SMALL LETTER P WITH ACUTE
4708        <P.>        <U1E56>    LATIN CAPITAL LETTER P WITH DOT ABOVE
4709        <p.>        <U1E57>    LATIN SMALL LETTER P WITH DOT ABOVE
4710        <R.>        <U1E58>    LATIN CAPITAL LETTER R WITH DOT ABOVE
4711        <r.>        <U1E59>    LATIN SMALL LETTER R WITH DOT ABOVE
4712        <R-.>       <U1E5A>    LATIN CAPITAL LETTER R WITH DOT BELOW
4713        <r-.>       <U1E5B>    LATIN SMALL LETTER R WITH DOT BELOW
4714        <R--.>      <U1E5C>    LATIN CAPITAL LETTER R WITH DOT BELOW AND MACRON
4715        <r--.>      <U1E5D>    LATIN SMALL LETTER R WITH DOT BELOW AND MACRON
4716        <R_>        <U1E5E>    LATIN CAPITAL LETTER R WITH LINE BELOW
4717        <r_>        <U1E5F>    LATIN SMALL LETTER R WITH LINE BELOW
4718        <S.>        <U1E60>    LATIN CAPITAL LETTER S WITH DOT ABOVE
4719        <s.>        <U1E61>    LATIN SMALL LETTER S WITH DOT ABOVE
4720        <S-.>       <U1E62>    LATIN CAPITAL LETTER S WITH DOT BELOW
4721        <s-.>       <U1E63>    LATIN SMALL LETTER S WITH DOT BELOW
4722        <S'.>       <U1E64>    LATIN CAPITAL LETTER S WITH ACUTE AND DOT ABOVE
4723        <s'.>       <U1E65>    LATIN SMALL LETTER S WITH ACUTE AND DOT ABOVE
4724        <S<.>       <U1E66>    LATIN CAPITAL LETTER S WITH CARON AND DOT ABOVE
4725        <s<.>       <U1E67>    LATIN SMALL LETTER S WITH CARON AND DOT ABOVE
4726        <S.-.>      <U1E68>    LATIN CAPITAL LETTER S WITH DOT BELOW AND DOT ABOVE
4727        <s.-.>      <U1E69>    LATIN SMALL LETTER S WITH DOT BELOW AND DOT ABOVE
4728        <T.>        <U1E6A>    LATIN CAPITAL LETTER T WITH DOT ABOVE
4729        <t.>        <U1E6B>    LATIN SMALL LETTER T WITH DOT ABOVE
4730        <T-.>       <U1E6C>    LATIN CAPITAL LETTER T WITH DOT BELOW
4731        <t-.>       <U1E6D>    LATIN SMALL LETTER T WITH DOT BELOW
4732        <T_>        <U1E6E>    LATIN CAPITAL LETTER T WITH LINE BELOW
4733        <t_>        <U1E6F>    LATIN SMALL LETTER T WITH LINE BELOW
4734        <T-/>>      <U1E70>    LATIN CAPITAL LETTER T WITH CIRCUMFLEX BELOW
4735        <t-/>>      <U1E71>    LATIN SMALL LETTER T WITH CIRCUMFLEX BELOW
4736        <U--:>      <U1E72>    LATIN CAPITAL LETTER U WITH DIAERESIS BELOW
4737        <u--:>      <U1E73>    LATIN SMALL LETTER U WITH DIAERESIS BELOW
4738        <U-?>       <U1E74>    LATIN CAPITAL LETTER U WITH TILDE BELOW
4739        <u-?>       <U1E75>    LATIN SMALL LETTER U WITH TILDE BELOW
4740        <U-/>>      <U1E76>    LATIN CAPITAL LETTER U WITH CIRCUMFLEX BELOW
4741        <u-/>>      <U1E77>    LATIN SMALL LETTER U WITH CIRCUMFLEX BELOW
4742        <U?'>       <U1E78>    LATIN CAPITAL LETTER U WITH TILDE AND ACUTE
4743        <u?'>       <U1E79>    LATIN SMALL LETTER U WITH TILDE AND ACUTE
4744        <U-:>       <U1E7A>    LATIN CAPITAL LETTER U WITH MACRON AND DIAERESIS
4745        <u-:>       <U1E7B>    LATIN SMALL LETTER U WITH MACRON AND DIAERESIS
4746        <V?>        <U1E7C>    LATIN CAPITAL LETTER V WITH TILDE
4747        <v?>        <U1E7D>    LATIN SMALL LETTER V WITH TILDE
4748        <V-.>       <U1E7E>    LATIN CAPITAL LETTER V WITH DOT BELOW
4749        <v-.>       <U1E7F>    LATIN SMALL LETTER V WITH DOT BELOW
4750        <W!>        <U1E80>    LATIN CAPITAL LETTER W WITH GRAVE
4751        <w!>        <U1E81>    LATIN SMALL LETTER W WITH GRAVE
4752        <W'>        <U1E82>    LATIN CAPITAL LETTER W WITH ACUTE
4753        <w'>        <U1E83>    LATIN SMALL LETTER W WITH ACUTE
4754        <W:>        <U1E84>    LATIN CAPITAL LETTER W WITH DIAERESIS
4755        <w:>        <U1E85>    LATIN SMALL LETTER W WITH DIAERESIS
4756        <W.>        <U1E86>    LATIN CAPITAL LETTER W WITH DOT ABOVE
4757        <w.>        <U1E87>    LATIN SMALL LETTER W WITH DOT ABOVE
4758        <W-.>       <U1E88>    LATIN CAPITAL LETTER W WITH DOT BELOW
4759        <w-.>       <U1E89>    LATIN SMALL LETTER W WITH DOT BELOW
4760        <X.>        <U1E8A>    LATIN CAPITAL LETTER X WITH DOT ABOVE
4761        <x.>        <U1E8B>    LATIN SMALL LETTER X WITH DOT ABOVE
4762        <X:>        <U1E8C>    LATIN CAPITAL LETTER X WITH DIAERESIS
4763        <x:>        <U1E8D>    LATIN SMALL LETTER X WITH DIAERESIS
4764        <Y.>        <U1E8E>    LATIN CAPITAL LETTER Y WITH DOT ABOVE
4765        <y.>        <U1E8F>    LATIN SMALL LETTER Y WITH DOT ABOVE
4766        <Z/>>       <U1E90>    LATIN CAPITAL LETTER Z WITH CIRCUMFLEX
4767        <z/>>       <U1E91>    LATIN SMALL LETTER Z WITH CIRCUMFLEX
4768        <Z-.>       <U1E92>    LATIN CAPITAL LETTER Z WITH DOT BELOW
4769        <z-.>       <U1E93>    LATIN SMALL LETTER Z WITH DOT BELOW
4770        <Z_>        <U1E94>    LATIN CAPITAL LETTER Z WITH LINE BELOW
4771        <z_>        <U1E95>    LATIN SMALL LETTER Z WITH LINE BELOW
4772        <A-.>       <U1EA0>    LATIN CAPITAL LETTER A WITH DOT BELOW
4773        <a-.>       <U1EA1>    LATIN SMALL LETTER A WITH DOT BELOW
4774        <A2>        <U1EA2>    LATIN CAPITAL LETTER A WITH HOOK ABOVE
4775        <a2>        <U1EA3>    LATIN SMALL LETTER A WITH HOOK ABOVE
4776        <A/>'>      <U1EA4>    LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND ACUTE
4777        <a/>'>      <U1EA5>    LATIN SMALL LETTER A WITH CIRCUMFLEX AND ACUTE
4778        <A/>!>      <U1EA6>    LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND GRAVE
4779        <a/>!>      <U1EA7>    LATIN SMALL LETTER A WITH CIRCUMFLEX AND GRAVE
4780        <A/>2>      <U1EA8>    LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND HOOK ABOVE
4781        <a/>2>      <U1EA9>    LATIN SMALL LETTER A WITH CIRCUMFLEX AND HOOK ABOVE
4782        <A/>?>      <U1EAA>    LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND TILDE
4783        <a/>?>      <U1EAB>    LATIN SMALL LETTER A WITH CIRCUMFLEX AND TILDE
4784        <A/>-.>     <U1EAC>    LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND DOT BELOW
```

74

```
4785    <a/>-.>      <U1EAD>     LATIN SMALL LETTER A WITH CIRCUMFLEX AND DOT BELOW
4786    <A('>        <U1EAE>     LATIN CAPITAL LETTER A WITH BREVE AND ACUTE
4787    <a('>        <U1EAF>     LATIN SMALL LETTER A WITH BREVE AND ACUTE
4788    <A(!>        <U1EB0>     LATIN CAPITAL LETTER A WITH BREVE AND GRAVE
4789    <a(!>        <U1EB1>     LATIN SMALL LETTER A WITH BREVE AND GRAVE
4790    <A(2>        <U1EB2>     LATIN CAPITAL LETTER A WITH BREVE AND HOOK ABOVE
4791    <a(2>        <U1EB3>     LATIN SMALL LETTER A WITH BREVE AND HOOK ABOVE
4792    <A(?>        <U1EB4>     LATIN CAPITAL LETTER A WITH BREVE AND TILDE
4793    <a(?>        <U1EB5>     LATIN SMALL LETTER A WITH BREVE AND TILDE
4794    <A(-.>       <U1EB6>     LATIN CAPITAL LETTER A WITH BREVE AND DOT BELOW
4795    <a(-.>       <U1EB7>     LATIN SMALL LETTER A WITH BREVE AND DOT BELOW
4796    <E-.>        <U1EB8>     LATIN CAPITAL LETTER E WITH DOT BELOW
4797    <e-.>        <U1EB9>     LATIN SMALL LETTER E WITH DOT BELOW
4798    <E2>         <U1EBA>     LATIN CAPITAL LETTER E WITH HOOK ABOVE
4799    <e2>         <U1EBB>     LATIN SMALL LETTER E WITH HOOK ABOVE
4800    <E?>         <U1EBC>     LATIN CAPITAL LETTER E WITH TILDE
4801    <e?>         <U1EBD>     LATIN SMALL LETTER E WITH TILDE
4802    <E/>'>       <U1EBE>     LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND ACUTE
4803    <e/>'>       <U1EBF>     LATIN SMALL LETTER E WITH CIRCUMFLEX AND ACUTE
4804    <E/>!>       <U1EC0>     LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND GRAVE
4805    <e/>!>       <U1EC1>     LATIN SMALL LETTER E WITH CIRCUMFLEX AND GRAVE
4806    <E/>2>       <U1EC2>     LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND HOOK ABOVE
4807    <e/>2>       <U1EC3>     LATIN SMALL LETTER E WITH CIRCUMFLEX AND HOOK ABOVE
4808    <E/>?>       <U1EC4>     LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND TILDE
4809    <e/>?>       <U1EC5>     LATIN SMALL LETTER E WITH CIRCUMFLEX AND TILDE
4810    <E/>-.>      <U1EC6>     LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND DOT BELOW
4811    <e/>-.>      <U1EC7>     LATIN SMALL LETTER E WITH CIRCUMFLEX AND DOT BELOW
4812    <I2>         <U1EC8>     LATIN CAPITAL LETTER I WITH HOOK ABOVE
4813    <i2>         <U1EC9>     LATIN SMALL LETTER I WITH HOOK ABOVE
4814    <I-.>        <U1ECA>     LATIN CAPITAL LETTER I WITH DOT BELOW
4815    <i-.>        <U1ECB>     LATIN SMALL LETTER I WITH DOT BELOW
4816    <O-.>        <U1ECC>     LATIN CAPITAL LETTER O WITH DOT BELOW
4817    <o-.>        <U1ECD>     LATIN SMALL LETTER O WITH DOT BELOW
4818    <O2>         <U1ECE>     LATIN CAPITAL LETTER O WITH HOOK ABOVE
4819    <o2>         <U1ECF>     LATIN SMALL LETTER O WITH HOOK ABOVE
4820    <O/>'>       <U1ED0>     LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND ACUTE
4821    <o/>'>       <U1ED1>     LATIN SMALL LETTER O WITH CIRCUMFLEX AND ACUTE
4822    <O/>!>       <U1ED2>     LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND GRAVE
4823    <o/>!>       <U1ED3>     LATIN SMALL LETTER O WITH CIRCUMFLEX AND GRAVE
4824    <O/>2>       <U1ED4>     LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND HOOK ABOVE
4825    <o/>2>       <U1ED5>     LATIN SMALL LETTER O WITH CIRCUMFLEX AND HOOK ABOVE
4826    <O/>?>       <U1ED6>     LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND TILDE
4827    <o/>?>       <U1ED7>     LATIN SMALL LETTER O WITH CIRCUMFLEX AND TILDE
4828    <O/>-.>      <U1ED8>     LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND DOT BELOW
4829    <o/>-.>      <U1ED9>     LATIN SMALL LETTER O WITH CIRCUMFLEX AND DOT BELOW
4830    <O9'>        <U1EDA>     LATIN CAPITAL LETTER O WITH HORN AND ACUTE
4831    <o9'>        <U1EDB>     LATIN SMALL LETTER O WITH HORN AND ACUTE
4832    <O9!>        <U1EDC>     LATIN CAPITAL LETTER O WITH HORN AND GRAVE
4833    <o9!>        <U1EDD>     LATIN SMALL LETTER O WITH HORN AND GRAVE
4834    <O92>        <U1EDE>     LATIN CAPITAL LETTER O WITH HORN AND HOOK ABOVE
4835    <o92>        <U1EDF>     LATIN SMALL LETTER O WITH HORN AND HOOK ABOVE
4836    <O9?>        <U1EE0>     LATIN CAPITAL LETTER O WITH HORN AND TILDE
4837    <o9?>        <U1EE1>     LATIN SMALL LETTER O WITH HORN AND TILDE
4838    <O9-.>       <U1EE2>     LATIN CAPITAL LETTER O WITH HORN AND DOT BELOW
4839    <o9-.>       <U1EE3>     LATIN SMALL LETTER O WITH HORN AND DOT BELOW
4840    <U-.>        <U1EE4>     LATIN CAPITAL LETTER U WITH DOT BELOW
4841    <u-.>        <U1EE5>     LATIN SMALL LETTER U WITH DOT BELOW
4842    <U2>         <U1EE6>     LATIN CAPITAL LETTER U WITH HOOK ABOVE
4843    <u2>         <U1EE7>     LATIN SMALL LETTER U WITH HOOK ABOVE
4844    <U9'>        <U1EE8>     LATIN CAPITAL LETTER U WITH HORN AND ACUTE
4845    <u9'>        <U1EE9>     LATIN SMALL LETTER U WITH HORN AND ACUTE
4846    <U9!>        <U1EEA>     LATIN CAPITAL LETTER U WITH HORN AND GRAVE
4847    <u9!>        <U1EEB>     LATIN SMALL LETTER U WITH HORN AND GRAVE
4848    <U92>        <U1EEC>     LATIN CAPITAL LETTER U WITH HORN AND HOOK ABOVE
4849    <u92>        <U1EED>     LATIN SMALL LETTER U WITH HORN AND HOOK ABOVE
4850    <U9?>        <U1EEE>     LATIN CAPITAL LETTER U WITH HORN AND TILDE
4851    <u9?>        <U1EEF>     LATIN SMALL LETTER U WITH HORN AND TILDE
4852    <U9-.>       <U1EF0>     LATIN CAPITAL LETTER U WITH HORN AND DOT BELOW
4853    <u9-.>       <U1EF1>     LATIN SMALL LETTER U WITH HORN AND DOT BELOW
4854    <Y!>         <U1EF2>     LATIN CAPITAL LETTER Y WITH GRAVE
4855    <y!>         <U1EF3>     LATIN SMALL LETTER Y WITH GRAVE
4856    <Y-.>        <U1EF4>     LATIN CAPITAL LETTER Y WITH DOT BELOW
4857    <y-.>        <U1EF5>     LATIN SMALL LETTER Y WITH DOT BELOW
4858    <Y2>         <U1EF6>     LATIN CAPITAL LETTER Y WITH HOOK ABOVE
4859    <y2>         <U1EF7>     LATIN SMALL LETTER Y WITH HOOK ABOVE
4860    <Y?>         <U1EF8>     LATIN CAPITAL LETTER Y WITH TILDE
4861    <y?>         <U1EF9>     LATIN SMALL LETTER Y WITH TILDE
4862    <a*,>        <U1F00>     GREEK SMALL LETTER ALPHA WITH PSILI
4863    <a*;>        <U1F01>     GREEK SMALL LETTER ALPHA WITH DASIA
4864    <a*,!>       <U1F02>     GREEK SMALL LETTER ALPHA WITH PSILI AND VARIA
4865    <a*;!>       <U1F03>     GREEK SMALL LETTER ALPHA WITH DASIA AND VARIA
4866    <a*,'>       <U1F04>     GREEK SMALL LETTER ALPHA WITH PSILI AND OXIA
4867    <a*;'>       <U1F05>     GREEK SMALL LETTER ALPHA WITH DASIA AND OXIA
4868    <a*,?>       <U1F06>     GREEK SMALL LETTER ALPHA WITH PSILI AND PERISPOMENI
4869    <a*;?>       <U1F07>     GREEK SMALL LETTER ALPHA WITH DASIA AND PERISPOMENI
4870    <A*,>        <U1F08>     GREEK CAPITAL LETTER ALPHA WITH PSILI
4871    <A*;>        <U1F09>     GREEK CAPITAL LETTER ALPHA WITH DASIA
4872    <A*,!>       <U1F0A>     GREEK CAPITAL LETTER ALPHA WITH PSILI AND VARIA
```

```
4873        <A*;!>     <U1F0B>   GREEK CAPITAL LETTER ALPHA WITH DASIA AND VARIA
4874        <A*,'>     <U1F0C>   GREEK CAPITAL LETTER ALPHA WITH PSILI AND OXIA
4875        <A*;'>     <U1F0D>   GREEK CAPITAL LETTER ALPHA WITH DASIA AND OXIA
4876        <A*,?>     <U1F0E>   GREEK CAPITAL LETTER ALPHA WITH PSILI AND PERISPOMENI
4877        <A*;?>     <U1F0F>   GREEK CAPITAL LETTER ALPHA WITH DASIA AND PERISPOMENI
4878        <e*,>      <U1F10>   GREEK SMALL LETTER EPSILON WITH PSILI
4879        <e*;>      <U1F11>   GREEK SMALL LETTER EPSILON WITH DASIA
4880        <e*,!>     <U1F12>   GREEK SMALL LETTER EPSILON WITH PSILI AND VARIA
4881        <e*;!>     <U1F13>   GREEK SMALL LETTER EPSILON WITH DASIA AND VARIA
4882        <e*,'>     <U1F14>   GREEK SMALL LETTER EPSILON WITH PSILI AND OXIA
4883        <e*;'>     <U1F15>   GREEK SMALL LETTER EPSILON WITH DASIA AND OXIA
4884        <E*,>      <U1F18>   GREEK CAPITAL LETTER EPSILON WITH PSILI
4885        <E*;>      <U1F19>   GREEK CAPITAL LETTER EPSILON WITH DASIA
4886        <E*,!>     <U1F1A>   GREEK CAPITAL LETTER EPSILON WITH PSILI AND VARIA
4887        <E*;!>     <U1F1B>   GREEK CAPITAL LETTER EPSILON WITH DASIA AND VARIA
4888        <E*,'>     <U1F1C>   GREEK CAPITAL LETTER EPSILON WITH PSILI AND OXIA
4889        <E*;'>     <U1F1D>   GREEK CAPITAL LETTER EPSILON WITH DASIA AND OXIA
4890        <y*,>      <U1F20>   GREEK SMALL LETTER ETA WITH PSILI
4891        <y*;>      <U1F21>   GREEK SMALL LETTER ETA WITH DASIA
4892        <y*,!>     <U1F22>   GREEK SMALL LETTER ETA WITH PSILI AND VARIA
4893        <y*;!>     <U1F23>   GREEK SMALL LETTER ETA WITH DASIA AND VARIA
4894        <y*,'>     <U1F24>   GREEK SMALL LETTER ETA WITH PSILI AND OXIA
4895        <y*;'>     <U1F25>   GREEK SMALL LETTER ETA WITH DASIA AND OXIA
4896        <y*,?>     <U1F26>   GREEK SMALL LETTER ETA WITH PSILI AND PERISPOMENI
4897        <y*;?>     <U1F27>   GREEK SMALL LETTER ETA WITH DASIA AND PERISPOMENI
4898        <Y*,>      <U1F28>   GREEK CAPITAL LETTER ETA WITH PSILI
4899        <Y*;>      <U1F29>   GREEK CAPITAL LETTER ETA WITH DASIA
4900        <Y*,!>     <U1F2A>   GREEK CAPITAL LETTER ETA WITH PSILI AND VARIA
4901        <Y*;!>     <U1F2B>   GREEK CAPITAL LETTER ETA WITH DASIA AND VARIA
4902        <Y*,'>     <U1F2C>   GREEK CAPITAL LETTER ETA WITH PSILI AND OXIA
4903        <Y*;'>     <U1F2D>   GREEK CAPITAL LETTER ETA WITH DASIA AND OXIA
4904        <Y*,?>     <U1F2E>   GREEK CAPITAL LETTER ETA WITH PSILI AND PERISPOMENI
4905        <Y*;?>     <U1F2F>   GREEK CAPITAL LETTER ETA WITH DASIA AND PERISPOMENI
4906        <i*,>      <U1F30>   GREEK SMALL LETTER IOTA WITH PSILI
4907        <i*;>      <U1F31>   GREEK SMALL LETTER IOTA WITH DASIA
4908        <i*,!>     <U1F32>   GREEK SMALL LETTER IOTA WITH PSILI AND VARIA
4909        <i*;!>     <U1F33>   GREEK SMALL LETTER IOTA WITH DASIA AND VARIA
4910        <i*,'>     <U1F34>   GREEK SMALL LETTER IOTA WITH PSILI AND OXIA
4911        <i*;'>     <U1F35>   GREEK SMALL LETTER IOTA WITH DASIA AND OXIA
4912        <i*,?>     <U1F36>   GREEK SMALL LETTER IOTA WITH PSILI AND PERISPOMENI
4913        <i*;?>     <U1F37>   GREEK SMALL LETTER IOTA WITH DASIA AND PERISPOMENI
4914        <I*,>      <U1F38>   GREEK CAPITAL LETTER IOTA WITH PSILI
4915        <I*;>      <U1F39>   GREEK CAPITAL LETTER IOTA WITH DASIA
4916        <I*,!>     <U1F3A>   GREEK CAPITAL LETTER IOTA WITH PSILI AND VARIA
4917        <I*;!>     <U1F3B>   GREEK CAPITAL LETTER IOTA WITH DASIA AND VARIA
4918        <I*,'>     <U1F3C>   GREEK CAPITAL LETTER IOTA WITH PSILI AND OXIA
4919        <I*;'>     <U1F3D>   GREEK CAPITAL LETTER IOTA WITH DASIA AND OXIA
4920        <I*,?>     <U1F3E>   GREEK CAPITAL LETTER IOTA WITH PSILI AND PERISPOMENI
4921        <I*;?>     <U1F3F>   GREEK CAPITAL LETTER IOTA WITH DASIA AND PERISPOMENI
4922        <o*,>      <U1F40>   GREEK SMALL LETTER OMICRON WITH PSILI
4923        <o*;>      <U1F41>   GREEK SMALL LETTER OMICRON WITH DASIA
4924        <o*,!>     <U1F42>   GREEK SMALL LETTER OMICRON WITH PSILI AND VARIA
4925        <o*;!>     <U1F43>   GREEK SMALL LETTER OMICRON WITH DASIA AND VARIA
4926        <o*,'>     <U1F44>   GREEK SMALL LETTER OMICRON WITH PSILI AND OXIA
4927        <o*;'>     <U1F45>   GREEK SMALL LETTER OMICRON WITH DASIA AND OXIA
4928        <O*,>      <U1F48>   GREEK CAPITAL LETTER OMICRON WITH PSILI
4929        <O*;>      <U1F49>   GREEK CAPITAL LETTER OMICRON WITH DASIA
4930        <O*,!>     <U1F4A>   GREEK CAPITAL LETTER OMICRON WITH PSILI AND VARIA
4931        <O*;!>     <U1F4B>   GREEK CAPITAL LETTER OMICRON WITH DASIA AND VARIA
4932        <O*,'>     <U1F4C>   GREEK CAPITAL LETTER OMICRON WITH PSILI AND OXIA
4933        <O*;'>     <U1F4D>   GREEK CAPITAL LETTER OMICRON WITH DASIA AND OXIA
4934        <u*,>      <U1F50>   GREEK SMALL LETTER UPSILON WITH PSILI
4935        <u*;>      <U1F51>   GREEK SMALL LETTER UPSILON WITH DASIA
4936        <u*,!>     <U1F52>   GREEK SMALL LETTER UPSILON WITH PSILI AND VARIA
4937        <u*;!>     <U1F53>   GREEK SMALL LETTER UPSILON WITH DASIA AND VARIA
4938        <u*,'>     <U1F54>   GREEK SMALL LETTER UPSILON WITH PSILI AND OXIA
4939        <u*;'>     <U1F55>   GREEK SMALL LETTER UPSILON WITH DASIA AND OXIA
4940        <u*,?>     <U1F56>   GREEK SMALL LETTER UPSILON WITH PSILI AND PERISPOMENI
4941        <u*;?>     <U1F57>   GREEK SMALL LETTER UPSILON WITH DASIA AND PERISPOMENI
4942        <U*;>      <U1F59>   GREEK CAPITAL LETTER UPSILON WITH DASIA
4943        <U*;!>     <U1F5B>   GREEK CAPITAL LETTER UPSILON WITH DASIA AND VARIA
4944        <U*;'>     <U1F5D>   GREEK CAPITAL LETTER UPSILON WITH DASIA AND OXIA
4945        <U*;?>     <U1F5F>   GREEK CAPITAL LETTER UPSILON WITH DASIA AND PERISPOMENI
4946        <w*,>      <U1F60>   GREEK SMALL LETTER OMEGA WITH PSILI
4947        <w*;>      <U1F61>   GREEK SMALL LETTER OMEGA WITH DASIA
4948        <w*,!>     <U1F62>   GREEK SMALL LETTER OMEGA WITH PSILI AND VARIA
4949        <w*;!>     <U1F63>   GREEK SMALL LETTER OMEGA WITH DASIA AND VARIA
4950        <w*,'>     <U1F64>   GREEK SMALL LETTER OMEGA WITH PSILI AND OXIA
4951        <w*;'>     <U1F65>   GREEK SMALL LETTER OMEGA WITH DASIA AND OXIA
4952        <w*,?>     <U1F66>   GREEK SMALL LETTER OMEGA WITH PSILI AND PERISPOMENI
4953        <w*;?>     <U1F67>   GREEK SMALL LETTER OMEGA WITH DASIA AND PERISPOMENI
4954        <W*,>      <U1F68>   GREEK CAPITAL LETTER OMEGA WITH PSILI
4955        <W*;>      <U1F69>   GREEK CAPITAL LETTER OMEGA WITH DASIA
4956        <W*,!>     <U1F6A>   GREEK CAPITAL LETTER OMEGA WITH PSILI AND VARIA
4957        <W*;!>     <U1F6B>   GREEK CAPITAL LETTER OMEGA WITH DASIA AND VARIA
4958        <W*,'>     <U1F6C>   GREEK CAPITAL LETTER OMEGA WITH PSILI AND OXIA
4959        <W*;'>     <U1F6D>   GREEK CAPITAL LETTER OMEGA WITH DASIA AND OXIA
4960        <W*,?>     <U1F6E>   GREEK CAPITAL LETTER OMEGA WITH PSILI AND PERISPOMENI
4961        <W*;?>     <U1F6F>   GREEK CAPITAL LETTER OMEGA WITH DASIA AND PERISPOMENI
```

```
4962   <a*!>      <U1F70>   GREEK SMALL LETTER ALPHA WITH VARIA
4963   <a*'>      <U1F71>   GREEK SMALL LETTER ALPHA WITH OXIA
4964   <e*!>      <U1F72>   GREEK SMALL LETTER EPSILON WITH VARIA
4965   <e*'>      <U1F73>   GREEK SMALL LETTER EPSILON WITH OXIA
4966   <y*!>      <U1F74>   GREEK SMALL LETTER ETA WITH VARIA
4967   <y*'>      <U1F75>   GREEK SMALL LETTER ETA WITH OXIA
4968   <i*!>      <U1F76>   GREEK SMALL LETTER IOTA WITH VARIA
4969   <i*'>      <U1F77>   GREEK SMALL LETTER IOTA WITH OXIA
4970   <o*!>      <U1F78>   GREEK SMALL LETTER OMICRON WITH VARIA
4971   <o*'>      <U1F79>   GREEK SMALL LETTER OMICRON WITH OXIA
4972   <u*!>      <U1F7A>   GREEK SMALL LETTER UPSILON WITH VARIA
4973   <u*'>      <U1F7B>   GREEK SMALL LETTER UPSILON WITH OXIA
4974   <w*!>      <U1F7C>   GREEK SMALL LETTER OMEGA WITH VARIA
4975   <w*'>      <U1F7D>   GREEK SMALL LETTER OMEGA WITH OXIA
4976   <a*,j>     <U1F80>   GREEK SMALL LETTER ALPHA WITH PSILI AND YPOGEGRAMMENI
4977   <a*;j>     <U1F81>   GREEK SMALL LETTER ALPHA WITH DASIA AND YPOGEGRAMMENI
4978   <a*,!j>    <U1F82>   GREEK SMALL LETTER ALPHA WITH PSILI AND VARIA AND YPOGEGRAMMENI
4979   <a*;!j>    <U1F83>   GREEK SMALL LETTER ALPHA WITH DASIA AND VARIA AND YPOGEGRAMMENI
4980   <a*,'j>    <U1F84>   GREEK SMALL LETTER ALPHA WITH PSILI AND OXIA AND YPOGEGRAMMENI
4981   <a*;'j>    <U1F85>   GREEK SMALL LETTER ALPHA WITH DASIA AND OXIA AND YPOGEGRAMMENI
4982   <a*,?j>    <U1F86>   GREEK SMALL LETTER ALPHA WITH PSILI AND PERISPOMENI AND YPOGEGRAMMENI
4983   <a*;?j>    <U1F87>   GREEK SMALL LETTER ALPHA WITH DASIA AND PERISPOMENI AND YPOGEGRAMMENI
4984   <A*,J>     <U1F88>   GREEK CAPITAL LETTER ALPHA WITH PSILI AND PROSGEGRAMMENI
4985   <A*;J>     <U1F89>   GREEK CAPITAL LETTER ALPHA WITH DASIA AND PROSGEGRAMMENI
4986   <A*,!J>    <U1F8A>   GREEK CAPITAL LETTER ALPHA WITH PSILI AND VARIA AND PROSGEGRAMMENI
4987   <A*;!J>    <U1F8B>   GREEK CAPITAL LETTER ALPHA WITH DASIA AND VARIA AND PROSGEGRAMMENI
4988   <A*,'J>    <U1F8C>   GREEK CAPITAL LETTER ALPHA WITH PSILI AND OXIA AND PROSGEGRAMMENI
4989   <A*;'J>    <U1F8D>   GREEK CAPITAL LETTER ALPHA WITH DASIA AND OXIA AND PROSGEGRAMMENI
4990   <A*,?J>    <U1F8E>   GREEK CAPITAL LETTER ALPHA WITH PSILI AND PERISPOMENI AND PROSGEGRAMMENI
4991   <A*;?J>    <U1F8F>   GREEK CAPITAL LETTER ALPHA WITH DASIA AND PERISPOMENI AND PROSGEGRAMMENI
4992   <y*,j>     <U1F90>   GREEK SMALL LETTER ETA WITH PSILI AND YPOGEGRAMMENI
4993   <y*;j>     <U1F91>   GREEK SMALL LETTER ETA WITH DASIA AND YPOGEGRAMMENI
4994   <y*,!j>    <U1F92>   GREEK SMALL LETTER ETA WITH PSILI AND VARIA AND YPOGEGRAMMENI
4995   <y*;!j>    <U1F93>   GREEK SMALL LETTER ETA WITH DASIA AND VARIA AND YPOGEGRAMMENI
4996   <y*,'j>    <U1F94>   GREEK SMALL LETTER ETA WITH PSILI AND OXIA AND YPOGEGRAMMENI
4997   <y*;'j>    <U1F95>   GREEK SMALL LETTER ETA WITH DASIA AND OXIA AND YPOGEGRAMMENI
4998   <y*,?j>    <U1F96>   GREEK SMALL LETTER ETA WITH PSILI AND PERISPOMENI AND YPOGEGRAMMENI
4999   <y*;?j>    <U1F97>   GREEK SMALL LETTER ETA WITH DASIA AND PERISPOMENI AND YPOGEGRAMMENI
5000   <Y*,J>     <U1F98>   GREEK CAPITAL LETTER ETA WITH PSILI AND PROSGEGRAMMENI
5001   <Y*;J>     <U1F99>   GREEK CAPITAL LETTER ETA WITH DASIA AND PROSGEGRAMMENI
5002   <Y*,!J>    <U1F9A>   GREEK CAPITAL LETTER ETA WITH PSILI AND VARIA AND PROSGEGRAMMENI
5003   <Y*;!J>    <U1F9B>   GREEK CAPITAL LETTER ETA WITH DASIA AND VARIA AND PROSGEGRAMMENI
5004   <Y*,'J>    <U1F9C>   GREEK CAPITAL LETTER ETA WITH PSILI AND OXIA AND PROSGEGRAMMENI
5005   <Y*;'J>    <U1F9D>   GREEK CAPITAL LETTER ETA WITH DASIA AND OXIA AND PROSGEGRAMMENI
5006   <Y*,?J>    <U1F9E>   GREEK CAPITAL LETTER ETA WITH PSILI AND PERISPOMENI AND PROSGEGRAMMENI
5007   <Y*;?J>    <U1F9F>   GREEK CAPITAL LETTER ETA WITH DASIA AND PERISPOMENI AND PROSGEGRAMMENI
5008   <w*,j>     <U1FA0>   GREEK SMALL LETTER OMEGA WITH PSILI AND YPOGEGRAMMENI
5009   <w*;j>     <U1FA1>   GREEK SMALL LETTER OMEGA WITH DASIA AND YPOGEGRAMMENI
5010   <w*,!j>    <U1FA2>   GREEK SMALL LETTER OMEGA WITH PSILI AND VARIA AND YPOGEGRAMMENI
5011   <w*;!j>    <U1FA3>   GREEK SMALL LETTER OMEGA WITH DASIA AND VARIA AND YPOGEGRAMMENI
5012   <w*,'j>    <U1FA4>   GREEK SMALL LETTER OMEGA WITH PSILI AND OXIA AND YPOGEGRAMMENI
5013   <w*;'j>    <U1FA5>   GREEK SMALL LETTER OMEGA WITH DASIA AND OXIA AND YPOGEGRAMMENI
5014   <w*,?j>    <U1FA6>   GREEK SMALL LETTER OMEGA WITH PSILI AND PERISPOMENI AND YPOGEGRAMMENI
5015   <w*;?j>    <U1FA7>   GREEK SMALL LETTER OMEGA WITH DASIA AND PERISPOMENI AND YPOGEGRAMMENI
5016   <W*,J>     <U1FA8>   GREEK CAPITAL LETTER OMEGA WITH PSILI AND PROSGEGRAMMENI
5017   <W*;J>     <U1FA9>   GREEK CAPITAL LETTER OMEGA WITH DASIA AND PROSGEGRAMMENI
5018   <W*,!J>    <U1FAA>   GREEK CAPITAL LETTER OMEGA WITH PSILI AND VARIA AND PROSGEGRAMMENI
5019   <W*;!J>    <U1FAB>   GREEK CAPITAL LETTER OMEGA WITH DASIA AND VARIA AND PROSGEGRAMMENI
5020   <W*,'J>    <U1FAC>   GREEK CAPITAL LETTER OMEGA WITH PSILI AND OXIA AND PROSGEGRAMMENI
5021   <W*;'J>    <U1FAD>   GREEK CAPITAL LETTER OMEGA WITH DASIA AND OXIA AND PROSGEGRAMMENI
5022   <W*,?J>    <U1FAE>   GREEK CAPITAL LETTER OMEGA WITH PSILI AND PERISPOMENI AND PROSGEGRAMMENI
5023   <W*;?J>    <U1FAF>   GREEK CAPITAL LETTER OMEGA WITH DASIA AND PERISPOMENI AND PROSGEGRAMMENI
5024   <a*(>      <U1FB0>   GREEK SMALL LETTER ALPHA WITH VRACHY
5025   <a*->      <U1FB1>   GREEK SMALL LETTER ALPHA WITH MACRON
5026   <a*!j>     <U1FB2>   GREEK SMALL LETTER ALPHA WITH VARIA AND YPOGEGRAMMENI
5027   <a*j>      <U1FB3>   GREEK SMALL LETTER ALPHA WITH YPOGEGRAMMENI
5028   <a*'j>     <U1FB4>   GREEK SMALL LETTER ALPHA WITH OXIA AND YPOGEGRAMMENI
5029   <a*?>      <U1FB6>   GREEK SMALL LETTER ALPHA WITH PERISPOMENI
5030   <a*?j>     <U1FB7>   GREEK SMALL LETTER ALPHA WITH PERISPOMENI AND YPOGEGRAMMENI
5031   <A*(>      <U1FB8>   GREEK CAPITAL LETTER ALPHA WITH VRACHY
5032   <A*->      <U1FB9>   GREEK CAPITAL LETTER ALPHA WITH MACRON
5033   <A*!>      <U1FBA>   GREEK CAPITAL LETTER ALPHA WITH VARIA
5034   <A*'>      <U1FBB>   GREEK CAPITAL LETTER ALPHA WITH OXIA
5035   <A*J>      <U1FBC>   GREEK CAPITAL LETTER ALPHA WITH PROSGEGRAMMENI
5036   <)*>       <U1FBD>   GREEK KORONIS
5037   <J3>       <U1FBE>   GREEK PROSGEGRAMMENI
5038   <,,>       <U1FBF>   GREEK PSILI
5039   <?*>       <U1FC0>   GREEK PERISPOMENI
5040   <?:>       <U1FC1>   GREEK DIALYTIKA AND PERISPOMENI
5041   <y*!j>     <U1FC2>   GREEK SMALL LETTER ETA WITH VARIA AND YPOGEGRAMMENI
5042   <y*j>      <U1FC3>   GREEK SMALL LETTER ETA WITH YPOGEGRAMMENI
5043   <y*'j>     <U1FC4>   GREEK SMALL LETTER ETA WITH OXIA AND YPOGEGRAMMENI
5044   <y*?>      <U1FC6>   GREEK SMALL LETTER ETA WITH PERISPOMENI
5045   <y*?j>     <U1FC7>   GREEK SMALL LETTER ETA WITH PERISPOMENI AND YPOGEGRAMMENI
5046   <E*!!>     <U1FC8>   GREEK CAPITAL LETTER EPSILON WITH VARIA
5047   <E*'>      <U1FC9>   GREEK CAPITAL LETTER EPSILON WITH OXIA
5048   <Y*!>      <U1FCA>   GREEK CAPITAL LETTER ETA WITH VARIA
5049   <Y*'>      <U1FCB>   GREEK CAPITAL LETTER ETA WITH OXIA
```

```
5050        <Y*J>        <U1FCC>    GREEK CAPITAL LETTER ETA WITH PROSGEGRAMMENI
5051        <,!>         <U1FCD>    GREEK PSILI AND VARIA
5052        <,'>         <U1FCE>    GREEK PSILI AND OXIA
5053        <?,>         <U1FCF>    GREEK PSILI AND PERISPOMENI
5054        <i*(>        <U1FD0>    GREEK SMALL LETTER IOTA WITH VRACHY
5055        <i*->        <U1FD1>    GREEK SMALL LETTER IOTA WITH MACRON
5056        <i*:!>       <U1FD2>    GREEK SMALL LETTER IOTA WITH DIALYTIKA AND VARIA
5057        <i*:'>       <U1FD3>    GREEK SMALL LETTER IOTA WITH DIALYTIKA AND OXIA
5058        <i*?>        <U1FD6>    GREEK SMALL LETTER IOTA WITH PERISPOMENI
5059        <i*:?>       <U1FD7>    GREEK SMALL LETTER IOTA WITH DIALYTIKA AND PERISPOMENI
5060        <I*(>        <U1FD8>    GREEK CAPITAL LETTER IOTA WITH VRACHY
5061        <I*->        <U1FD9>    GREEK CAPITAL LETTER IOTA WITH MACRON
5062        <I*!>        <U1FDA>    GREEK CAPITAL LETTER IOTA WITH VARIA
5063        <I*'>        <U1FDB>    GREEK CAPITAL LETTER IOTA WITH OXIA
5064        <;!>         <U1FDD>    GREEK DASIA AND VARIA
5065        <;'>         <U1FDE>    GREEK DASIA AND OXIA
5066        <?;>         <U1FDF>    GREEK DASIA AND PERISPOMENI
5067        <u*(>        <U1FE0>    GREEK SMALL LETTER UPSILON WITH VRACHY
5068        <u*->        <U1FE1>    GREEK SMALL LETTER UPSILON WITH MACRON
5069        <u*:!>       <U1FE2>    GREEK SMALL LETTER UPSILON WITH DIALYTIKA AND VARIA
5070        <u*:'>       <U1FE3>    GREEK SMALL LETTER UPSILON WITH DIALYTIKA AND OXIA
5071        <r*,>        <U1FE4>    GREEK SMALL LETTER RHO WITH PSILI
5072        <r*;>        <U1FE5>    GREEK SMALL LETTER RHO WITH DASIA
5073        <u*?>        <U1FE6>    GREEK SMALL LETTER UPSILON WITH PERISPOMENI
5074        <u*:?>       <U1FE7>    GREEK SMALL LETTER UPSILON WITH DIALYTIKA AND PERISPOMENI
5075        <U*(>        <U1FE8>    GREEK CAPITAL LETTER UPSILON WITH VRACHY
5076        <U*->        <U1FE9>    GREEK CAPITAL LETTER UPSILON WITH MACRON
5077        <U*!>        <U1FEA>    GREEK CAPITAL LETTER UPSILON WITH VARIA
5078        <U*'>        <U1FEB>    GREEK CAPITAL LETTER UPSILON WITH OXIA
5079        <R*;>        <U1FEC>    GREEK CAPITAL LETTER RHO WITH DASIA
5080        <!:>         <U1FED>    GREEK DIALYTIKA AND VARIA
5081        <:'>         <U1FEE>    GREEK DIALYTIKA AND OXIA
5082        <!*>         <U1FEF>    GREEK VARIA
5083        <w*!j>       <U1FF2>    GREEK SMALL LETTER OMEGA WITH VARIA AND YPOGEGRAMMENI
5084        <w*j>        <U1FF3>    GREEK SMALL LETTER OMEGA WITH YPOGEGRAMMENI
5085        <w*'j>       <U1FF4>    GREEK SMALL LETTER OMEGA WITH OXIA AND YPOGEGRAMMENI
5086        <w*?>        <U1FF6>    GREEK SMALL LETTER OMEGA WITH PERISPOMENI
5087        <w*?j>       <U1FF7>    GREEK SMALL LETTER OMEGA WITH PERISPOMENI AND YPOGEGRAMMENI
5088        <O*!>        <U1FF8>    GREEK CAPITAL LETTER OMICRON WITH VARIA
5089        <O*'>        <U1FF9>    GREEK CAPITAL LETTER OMICRON WITH OXIA
5090        <W*!>        <U1FFA>    GREEK CAPITAL LETTER OMEGA WITH VARIA
5091        <W*'>        <U1FFB>    GREEK CAPITAL LETTER OMEGA WITH OXIA
5092        <W*J>        <U1FFC>    GREEK CAPITAL LETTER OMEGA WITH PROSGEGRAMMENI
5093        <//*>        <U1FFD>    GREEK OXIA
5094        <;;>         <U1FFE>    GREEK DASIA
5095        <1N>         <U2002>    EN SPACE
5096        <1M>         <U2003>    EM SPACE
5097        <3M>         <U2004>    THREE-PER-EM SPACE
5098        <4M>         <U2005>    FOUR-PER-EM SPACE
5099        <6M>         <U2006>    SIX-PER-EM SPACE
5100        <LR>         <U200E>    LEFT-TO-RIGHT MARK
5101        <RL>         <U200F>    RIGHT-TO-LEFT MARK
5102        <1T>         <U2009>    THIN SPACE
5103        <1H>         <U200A>    HAIR SPACE
5104        <-1>         <U2010>    HYPHEN
5105        <-N>         <U2013>    EN DASH
5106        <-M>         <U2014>    EM DASH
5107        <-3>         <U2015>    HORIZONTAL BAR
5108        <!2>         <U2016>    DOUBLE VERTICAL LINE
5109        <=2>         <U2017>    DOUBLE LOW LINE
5110        <'6>         <U2018>    LEFT SINGLE QUOTATION MARK
5111        <'9>         <U2019>    RIGHT SINGLE QUOTATION MARK
5112        <.9>         <U201A>    SINGLE LOW-9 QUOTATION MARK
5113        <9'>         <U201B>    SINGLE HIGH-REVERSED-9 QUOTATION MARK
5114        <"6>         <U201C>    LEFT DOUBLE QUOTATION MARK
5115        <"9>         <U201D>    RIGHT DOUBLE QUOTATION MARK
5116        <:9>         <U201E>    DOUBLE LOW-9 QUOTATION MARK
5117        <9">         <U201F>    DOUBLE HIGH-REVERSED-9 QUOTATION MARK
5118        <//->        <U2020>    DAGGER
5119        <//=>        <U2021>    DOUBLE DAGGER
5120        <sb>         <U2022>    BULLET
5121        <3b>         <U2023>    TRIANGULAR BULLET
5122        <..>         <U2025>    TWO DOT LEADER
5123        <.3>         <U2026>    HORIZONTAL ELLIPSIS
5124        <.->         <U2027>    HYPHENATION POINT
5125        <linesep>    <U2028>    LINE SEPARATOR
5126        <parsep>     <U2029>    PARAGRAPH SEPARATOR
5127        <%0>         <U2030>    PER MILLE SIGN
5128        <1'>         <U2032>    PRIME
5129        <2'>         <U2033>    DOUBLE PRIME
5130        <3'>         <U2034>    TRIPLE PRIME
5131        <1">         <U2035>    REVERSED PRIME
5132        <2">         <U2036>    REVERSED DOUBLE PRIME
5133        <3">         <U2037>    REVERSED TRIPLE PRIME
5134        <Ca>         <U2038>    CARET
5135        <<1>         <U2039>    SINGLE LEFT-POINTING ANGLE QUOTATION MARK
5136        </>1>        <U203A>    SINGLE RIGHT-POINTING ANGLE QUOTATION MARK
5137        <:X>         <U203B>    REFERENCE MARK
5138        <!*2>        <U203C>    DOUBLE EXCLAMATION MARK
```

```
5139        <'->        <U203E>    OVERLINE
5140        <-b>        <U2043>    HYPHEN BULLET
5141        <//f>       <U2044>    FRACTION SLASH
5142        <0S>        <U2070>    SUPERSCRIPT ZERO
5143        <4S>        <U2074>    SUPERSCRIPT FOUR
5144        <5S>        <U2075>    SUPERSCRIPT FIVE
5145        <6S>        <U2076>    SUPERSCRIPT SIX
5146        <7S>        <U2077>    SUPERSCRIPT SEVEN
5147        <8S>        <U2078>    SUPERSCRIPT EIGHT
5148        <9S>        <U2079>    SUPERSCRIPT NINE
5149        <+S>        <U207A>    SUPERSCRIPT PLUS SIGN
5150        <-S>        <U207B>    SUPERSCRIPT MINUS
5151        <=S>        <U207C>    SUPERSCRIPT EQUALS SIGN
5152        <(S>        <U207D>    SUPERSCRIPT LEFT PARENTHESIS
5153        <)S>        <U207E>    SUPERSCRIPT RIGHT PARENTHESIS
5154        <nS>        <U207F>    SUPERSCRIPT LATIN SMALL LETTER N
5155        <0s>        <U2080>    SUBSCRIPT ZERO
5156        <1s>        <U2081>    SUBSCRIPT ONE
5157        <2s>        <U2082>    SUBSCRIPT TWO
5158        <3s>        <U2083>    SUBSCRIPT THREE
5159        <4s>        <U2084>    SUBSCRIPT FOUR
5160        <5s>        <U2085>    SUBSCRIPT FIVE
5161        <6s>        <U2086>    SUBSCRIPT SIX
5162        <7s>        <U2087>    SUBSCRIPT SEVEN
5163        <8s>        <U2088>    SUBSCRIPT EIGHT
5164        <9s>        <U2089>    SUBSCRIPT NINE
5165        <+s>        <U208A>    SUBSCRIPT PLUS SIGN
5166        <-s>        <U208B>    SUBSCRIPT MINUS
5167        <=s>        <U208C>    SUBSCRIPT EQUALS SIGN
5168        <(s>        <U208D>    SUBSCRIPT LEFT PARENTHESIS
5169        <)s>        <U208E>    SUBSCRIPT RIGHT PARENTHESIS
5170        <Ff>        <U20A3>    FRENCH FRANC SIGN
5171        <Li>        <U20A4>    LIRA SIGN
5172        <Pt>        <U20A7>    PESETA SIGN
5173        <W=>        <U20A9>    WON SIGN
5174        <"7>        <U20D1>    COMBINING RIGHT HARPOON ABOVE
5175        <oC>        <U2103>    DEGREE CELSIUS
5176        <co>        <U2105>    CARE OF
5177        <oF>        <U2109>    DEGREE FAHRENHEIT
5178        <N0>        <U2116>    NUMERO SIGN
5179        <PO>        <U2117>    SOUND RECORDING COPYRIGHT
5180        <Rx>        <U211E>    PRESCRIPTION TAKE
5181        <SM>        <U2120>    SERVICE MARK
5182        <TM>        <U2122>    TRADE MARK SIGN
5183        <Om>        <U2126>    OHM SIGN
5184        <AO>        <U212B>    ANGSTROM SIGN
5185        <Est>       <U212E>    ESTIMATED SYMBOL
5186        <13>        <U2153>    VULGAR FRACTION ONE THIRD
5187        <23>        <U2154>    VULGAR FRACTION TWO THIRDS
5188        <15>        <U2155>    VULGAR FRACTION ONE FIFTH
5189        <25>        <U2156>    VULGAR FRACTION TWO FIFTHS
5190        <35>        <U2157>    VULGAR FRACTION THREE FIFTHS
5191        <45>        <U2158>    VULGAR FRACTION FOUR FIFTHS
5192        <16>        <U2159>    VULGAR FRACTION ONE SIXTH
5193        <56>        <U215A>    VULGAR FRACTION FIVE SIXTHS
5194        <18>        <U215B>    VULGAR FRACTION ONE EIGHTH
5195        <38>        <U215C>    VULGAR FRACTION THREE EIGHTHS
5196        <58>        <U215D>    VULGAR FRACTION FIVE EIGHTHS
5197        <78>        <U215E>    VULGAR FRACTION SEVEN EIGHTHS
5198        <1R>        <U2160>    ROMAN NUMERAL ONE
5199        <2R>        <U2161>    ROMAN NUMERAL TWO
5200        <3R>        <U2162>    ROMAN NUMERAL THREE
5201        <4R>        <U2163>    ROMAN NUMERAL FOUR
5202        <5R>        <U2164>    ROMAN NUMERAL FIVE
5203        <6R>        <U2165>    ROMAN NUMERAL SIX
5204        <7R>        <U2166>    ROMAN NUMERAL SEVEN
5205        <8R>        <U2167>    ROMAN NUMERAL EIGHT
5206        <9R>        <U2168>    ROMAN NUMERAL NINE
5207        <aR>        <U2169>    ROMAN NUMERAL TEN
5208        <bR>        <U216A>    ROMAN NUMERAL ELEVEN
5209        <cR>        <U216B>    ROMAN NUMERAL TWELVE
5210        <50R>       <U216C>    ROMAN NUMERAL FIFTY
5211        <100R>      <U216D>    ROMAN NUMERAL ONE HUNDRED
5212        <500R>      <U216E>    ROMAN NUMERAL FIVE HUNDRED
5213        <1000R>     <U216F>    ROMAN NUMERAL ONE THOUSAND
5214        <1r>        <U2170>    SMALL ROMAN NUMERAL ONE
5215        <2r>        <U2171>    SMALL ROMAN NUMERAL TWO
5216        <3r>        <U2172>    SMALL ROMAN NUMERAL THREE
5217        <4r>        <U2173>    SMALL ROMAN NUMERAL FOUR
5218        <5r>        <U2174>    SMALL ROMAN NUMERAL FIVE
5219        <6r>        <U2175>    SMALL ROMAN NUMERAL SIX
5220        <7r>        <U2176>    SMALL ROMAN NUMERAL SEVEN
5221        <8r>        <U2177>    SMALL ROMAN NUMERAL EIGHT
5222        <9r>        <U2178>    SMALL ROMAN NUMERAL NINE
5223        <ar>        <U2179>    SMALL ROMAN NUMERAL TEN
5224        <br>        <U217A>    SMALL ROMAN NUMERAL ELEVEN
5225        <cr>        <U217B>    SMALL ROMAN NUMERAL TWELVE
5226        <50r>       <U217C>    SMALL ROMAN NUMERAL FIFTY
```

```
5227    <100r>      <U217D>    SMALL ROMAN NUMERAL ONE HUNDRED
5228    <500r>      <U217E>    SMALL ROMAN NUMERAL FIVE HUNDRED
5229    <1000r>     <U217F>    SMALL ROMAN NUMERAL ONE THOUSAND
5230    <1000RCD>   <U2180>    ROMAN NUMERAL ONE THOUSAND C D
5231    <5000R>     <U2181>    ROMAN NUMERAL FIVE THOUSAND
5232    <10000R>    <U2182>    ROMAN NUMERAL TEN THOUSAND
5233    <<->        <U2190>    LEFTWARDS ARROW
5234    <-!>        <U2191>    UPWARDS ARROW
5235    <-/>>       <U2192>    RIGHTWARDS ARROW
5236    <-v>        <U2193>    DOWNWARDS ARROW
5237    <</>>       <U2194>    LEFT RIGHT ARROW
5238    <UD>        <U2195>    UP DOWN ARROW
5239    <<!!>       <U2196>    NORTH WEST ARROW
5240    <//////>>   <U2197>    NORTH EAST ARROW
5241    <!!/>>      <U2198>    SOUTH EAST ARROW
5242    <<////>     <U2199>    SOUTH WEST ARROW
5243    <UD->       <U21A8>    UP DOWN ARROW WITH BASE
5244    </>V        <U21C0>    RIGHTWARDS HARPOON WITH BARB UPWARDS
5245    <<=>        <U21D0>    LEFTWARDS DOUBLE ARROW
5246    <=/>>       <U21D2>    RIGHTWARDS DOUBLE ARROW
5247    <==>        <U21D4>    LEFT RIGHT DOUBLE ARROW
5248    <FA>        <U2200>    FOR ALL
5249    <dP>        <U2202>    PARTIAL DIFFERENTIAL
5250    <TE>        <U2203>    THERE EXISTS
5251    <//0>       <U2205>    EMPTY SET
5252    <DE>        <U2206>    INCREMENT
5253    <NB>        <U2207>    NABLA
5254    <(->        <U2208>    ELEMENT OF
5255    <-)>        <U220B>    CONTAINS AS MEMBER
5256    <FP>        <U220E>    END OF PROOF
5257    <*P>        <U220F>    N-ARY PRODUCT
5258    <+Z>        <U2211>    N-ARY SUMMATION
5259    <-2>        <U2212>    MINUS SIGN
5260    <-+>        <U2213>    MINUS-OR-PLUS SIGN
5261    <.+>        <U2214>    DOT PLUS
5262    <*->        <U2217>    ASTERISK OPERATOR
5263    <Ob>        <U2218>    RING OPERATOR
5264    <Sb>        <U2219>    BULLET OPERATOR
5265    <RT>        <U221A>    SQUARE ROOT
5266    <0(>        <U221D>    PROPORTIONAL TO
5267    <00>        <U221E>    INFINITY
5268    <-L>        <U221F>    RIGHT ANGLE
5269    <-V>        <U2220>    ANGLE
5270    <PP>        <U2225>    PARALLEL TO
5271    <AN>        <U2227>    LOGICAL AND
5272    <OR>        <U2228>    LOGICAL OR
5273    <(U>        <U2229>    INTERSECTION
5274    <)U>        <U222A>    UNION
5275    <In>        <U222B>    INTEGRAL
5276    <DI>        <U222C>    DOUBLE INTEGRAL
5277    <Io>        <U222E>    CONTOUR INTEGRAL
5278    <.:>        <U2234>    THEREFORE
5279    <:.>        <U2235>    BECAUSE
5280    <:R>        <U2236>    RATIO
5281    <::>        <U2237>    PROPORTION
5282    <?1>        <U223C>    TILDE OPERATOR
5283    <CG>        <U223E>    INVERTED LAZY S
5284    <?->        <U2243>    ASYMPTOTICALLY EQUAL TO
5285    <?=>        <U2245>    APPROXIMATELY EQUAL TO
5286    <?2>        <U2248>    ALMOST EQUAL TO
5287    <=?>        <U224C>    ALL EQUAL TO
5288    <HI>        <U2253>    IMAGE OF OR APPROXIMATELY EQUAL TO
5289    <!=>        <U2260>    NOT EQUAL TO
5290    <=3>        <U2261>    IDENTICAL TO
5291    <=<>        <U2264>    LESS-THAN OR EQUAL TO
5292    </>=>       <U2265>    GREATER-THAN OR EQUAL TO
5293    <<*>        <U226A>    MUCH LESS-THAN
5294    <*/>>       <U226B>    MUCH GREATER-THAN
5295    <!<>        <U226E>    NOT LESS-THAN
5296    <!/>>       <U226F>    NOT GREATER-THAN
5297    <(C>        <U2282>    SUBSET OF
5298    <)C>        <U2283>    SUPERSET OF
5299    <(_>        <U2286>    SUBSET OF OR EQUAL TO
5300    <)_>        <U2287>    SUPERSET OF OR EQUAL TO
5301    <0.>        <U2299>    CIRCLED DOT OPERATOR
5302    <02>        <U229A>    CIRCLED RING OPERATOR
5303    <-T>        <U22A5>    UP TACK
5304    <.P>        <U22C5>    DOT OPERATOR
5305    <:3>        <U22EE>    VERTICAL ELLIPSIS
5306    <Eh>        <U2302>    HOUSE
5307    <<7>        <U2308>    LEFT CEILING
5308    </>7>       <U2309>    RIGHT CEILING
5309    <7<>        <U230A>    LEFT FLOOR
5310    <7/>>       <U230B>    RIGHT FLOOR
5311    <NI>        <U2310>    REVERSED NOT SIGN
5312    <(A>        <U2312>    ARC
5313    <TR>        <U2315>    TELEPHONE RECORDER
5314    <88>        <U2318>    PLACE OF INTEREST SIGN
5315    <Iu>        <U2320>    TOP HALF INTEGRAL
```

```
5316        <Il>         <U2321>      BOTTOM HALF INTEGRAL
5317        <<//>        <U2329>      LEFT-POINTING ANGLE BRACKET
5318        <///>>       <U232A>      RIGHT-POINTING ANGLE BRACKET
5319        <Vs>         <U2423>      OPEN BOX
5320        <1h>         <U2440>      OCR HOOK
5321        <3h>         <U2441>      OCR CHAIR
5322        <2h>         <U2442>      OCR FORK
5323        <4h>         <U2443>      OCR INVERTED FORK
5324        <1j>         <U2446>      OCR BRANCH BANK IDENTIFICATION
5325        <2j>         <U2447>      OCR AMOUNT OF CHECK
5326        <3j>         <U2448>      OCR DASH
5327        <4j>         <U2449>      OCR CUSTOMER ACCOUNT NUMBER
5328        <1-o>        <U2460>      CIRCLED DIGIT ONE
5329        <2-o>        <U2461>      CIRCLED DIGIT TWO
5330        <3-o>        <U2462>      CIRCLED DIGIT THREE
5331        <4-o>        <U2463>      CIRCLED DIGIT FOUR
5332        <5-o>        <U2464>      CIRCLED DIGIT FIVE
5333        <6-o>        <U2465>      CIRCLED DIGIT SIX
5334        <7-o>        <U2466>      CIRCLED DIGIT SEVEN
5335        <8-o>        <U2467>      CIRCLED DIGIT EIGHT
5336        <9-o>        <U2468>      CIRCLED DIGIT NINE
5337        <10-o>       <U2469>      CIRCLED NUMBER TEN
5338        <11-o>       <U246A>      CIRCLED NUMBER ELEVEN
5339        <12-o>       <U246B>      CIRCLED NUMBER TWELVE
5340        <13-o>       <U246C>      CIRCLED NUMBER THIRTEEN
5341        <14-o>       <U246D>      CIRCLED NUMBER FOURTEEN
5342        <15-o>       <U246E>      CIRCLED NUMBER FIFTEEN
5343        <16-o>       <U246F>      CIRCLED NUMBER SIXTEEN
5344        <17-o>       <U2470>      CIRCLED NUMBER SEVENTEEN
5345        <18-o>       <U2471>      CIRCLED NUMBER EIGHTEEN
5346        <19-o>       <U2472>      CIRCLED NUMBER NINETEEN
5347        <20-o>       <U2473>      CIRCLED NUMBER TWENTY
5348        <(1)>        <U2474>      PARENTHESIZED DIGIT ONE
5349        <(2)>        <U2475>      PARENTHESIZED DIGIT TWO
5350        <(3)>        <U2476>      PARENTHESIZED DIGIT THREE
5351        <(4)>        <U2477>      PARENTHESIZED DIGIT FOUR
5352        <(5)>        <U2478>      PARENTHESIZED DIGIT FIVE
5353        <(6)>        <U2479>      PARENTHESIZED DIGIT SIX
5354        <(7)>        <U247A>      PARENTHESIZED DIGIT SEVEN
5355        <(8)>        <U247B>      PARENTHESIZED DIGIT EIGHT
5356        <(9)>        <U247C>      PARENTHESIZED DIGIT NINE
5357        <(10)>       <U247D>      PARENTHESIZED NUMBER TEN
5358        <(11)>       <U247E>      PARENTHESIZED NUMBER ELEVEN
5359        <(12)>       <U247F>      PARENTHESIZED NUMBER TWELVE
5360        <(13)>       <U2480>      PARENTHESIZED NUMBER THIRTEEN
5361        <(14)>       <U2481>      PARENTHESIZED NUMBER FOURTEEN
5362        <(15)>       <U2482>      PARENTHESIZED NUMBER FIFTEEN
5363        <(16)>       <U2483>      PARENTHESIZED NUMBER SIXTEEN
5364        <(17)>       <U2484>      PARENTHESIZED NUMBER SEVENTEEN
5365        <(18)>       <U2485>      PARENTHESIZED NUMBER EIGHTEEN
5366        <(19)>       <U2486>      PARENTHESIZED NUMBER NINETEEN
5367        <(20)>       <U2487>      PARENTHESIZED NUMBER TWENTY
5368        <1.>         <U2488>      DIGIT ONE FULL STOP
5369        <2.>         <U2489>      DIGIT TWO FULL STOP
5370        <3.>         <U248A>      DIGIT THREE FULL STOP
5371        <4.>         <U248B>      DIGIT FOUR FULL STOP
5372        <5.>         <U248C>      DIGIT FIVE FULL STOP
5373        <6.>         <U248D>      DIGIT SIX FULL STOP
5374        <7.>         <U248E>      DIGIT SEVEN FULL STOP
5375        <8.>         <U248F>      DIGIT EIGHT FULL STOP
5376        <9.>         <U2490>      DIGIT NINE FULL STOP
5377        <10.>        <U2491>      NUMBER TEN FULL STOP
5378        <11.>        <U2492>      NUMBER ELEVEN FULL STOP
5379        <12.>        <U2493>      NUMBER TWELVE FULL STOP
5380        <13.>        <U2494>      NUMBER THIRTEEN FULL STOP
5381        <14.>        <U2495>      NUMBER FOURTEEN FULL STOP
5382        <15.>        <U2496>      NUMBER FIFTEEN FULL STOP
5383        <16.>        <U2497>      NUMBER SIXTEEN FULL STOP
5384        <17.>        <U2498>      NUMBER SEVENTEEN FULL STOP
5385        <18.>        <U2499>      NUMBER EIGHTEEN FULL STOP
5386        <19.>        <U249A>      NUMBER NINETEEN FULL STOP
5387        <20.>        <U249B>      NUMBER TWENTY FULL STOP
5388        <(a)>        <U249C>      PARENTHESIZED LATIN SMALL LETTER A
5389        <(b)>        <U249D>      PARENTHESIZED LATIN SMALL LETTER B
5390        <(c)>        <U249E>      PARENTHESIZED LATIN SMALL LETTER C
5391        <(d)>        <U249F>      PARENTHESIZED LATIN SMALL LETTER D
5392        <(e)>        <U24A0>      PARENTHESIZED LATIN SMALL LETTER E
5393        <(f)>        <U24A1>      PARENTHESIZED LATIN SMALL LETTER F
5394        <(g)>        <U24A2>      PARENTHESIZED LATIN SMALL LETTER G
5395        <(h)>        <U24A3>      PARENTHESIZED LATIN SMALL LETTER H
5396        <(i)>        <U24A4>      PARENTHESIZED LATIN SMALL LETTER I
5397        <(j)>        <U24A5>      PARENTHESIZED LATIN SMALL LETTER J
5398        <(k)>        <U24A6>      PARENTHESIZED LATIN SMALL LETTER K
5399        <(l)>        <U24A7>      PARENTHESIZED LATIN SMALL LETTER L
5400        <(m)>        <U24A8>      PARENTHESIZED LATIN SMALL LETTER M
5401        <(n)>        <U24A9>      PARENTHESIZED LATIN SMALL LETTER N
5402        <(o)>        <U24AA>      PARENTHESIZED LATIN SMALL LETTER O
5403        <(p)>        <U24AB>      PARENTHESIZED LATIN SMALL LETTER P
```

```
5404        <(q)>        <U24AC>    PARENTHESIZED LATIN SMALL LETTER Q
5405        <(r)>        <U24AD>    PARENTHESIZED LATIN SMALL LETTER R
5406        <(s)>        <U24AE>    PARENTHESIZED LATIN SMALL LETTER S
5407        <(t)>        <U24AF>    PARENTHESIZED LATIN SMALL LETTER T
5408        <(u)>        <U24B0>    PARENTHESIZED LATIN SMALL LETTER U
5409        <(v)>        <U24B1>    PARENTHESIZED LATIN SMALL LETTER V
5410        <(w)>        <U24B2>    PARENTHESIZED LATIN SMALL LETTER W
5411        <(x)>        <U24B3>    PARENTHESIZED LATIN SMALL LETTER X
5412        <(y)>        <U24B4>    PARENTHESIZED LATIN SMALL LETTER Y
5413        <(z)>        <U24B5>    PARENTHESIZED LATIN SMALL LETTER Z
5414        <A-o>        <U24B6>    CIRCLED LATIN CAPITAL LETTER A
5415        <B-o>        <U24B7>    CIRCLED LATIN CAPITAL LETTER B
5416        <C-o>        <U24B8>    CIRCLED LATIN CAPITAL LETTER C
5417        <D-o>        <U24B9>    CIRCLED LATIN CAPITAL LETTER D
5418        <E-o>        <U24BA>    CIRCLED LATIN CAPITAL LETTER E
5419        <F-o>        <U24BB>    CIRCLED LATIN CAPITAL LETTER F
5420        <G-o>        <U24BC>    CIRCLED LATIN CAPITAL LETTER G
5421        <H-o>        <U24BD>    CIRCLED LATIN CAPITAL LETTER H
5422        <I-o>        <U24BE>    CIRCLED LATIN CAPITAL LETTER I
5423        <J-o>        <U24BF>    CIRCLED LATIN CAPITAL LETTER J
5424        <K-o>        <U24C0>    CIRCLED LATIN CAPITAL LETTER K
5425        <L-o>        <U24C1>    CIRCLED LATIN CAPITAL LETTER L
5426        <M-o>        <U24C2>    CIRCLED LATIN CAPITAL LETTER M
5427        <N-o>        <U24C3>    CIRCLED LATIN CAPITAL LETTER N
5428        <O-o>        <U24C4>    CIRCLED LATIN CAPITAL LETTER O
5429        <P-o>        <U24C5>    CIRCLED LATIN CAPITAL LETTER P
5430        <Q-o>        <U24C6>    CIRCLED LATIN CAPITAL LETTER Q
5431        <R-o>        <U24C7>    CIRCLED LATIN CAPITAL LETTER R
5432        <S-o>        <U24C8>    CIRCLED LATIN CAPITAL LETTER S
5433        <T-o>        <U24C9>    CIRCLED LATIN CAPITAL LETTER T
5434        <U-o>        <U24CA>    CIRCLED LATIN CAPITAL LETTER U
5435        <V-o>        <U24CB>    CIRCLED LATIN CAPITAL LETTER V
5436        <W-o>        <U24CC>    CIRCLED LATIN CAPITAL LETTER W
5437        <X-o>        <U24CD>    CIRCLED LATIN CAPITAL LETTER X
5438        <Y-o>        <U24CE>    CIRCLED LATIN CAPITAL LETTER Y
5439        <Z-o>        <U24CF>    CIRCLED LATIN CAPITAL LETTER Z
5440        <a-o>        <U24D0>    CIRCLED LATIN SMALL LETTER A
5441        <b-o>        <U24D1>    CIRCLED LATIN SMALL LETTER B
5442        <c-o>        <U24D2>    CIRCLED LATIN SMALL LETTER C
5443        <d-o>        <U24D3>    CIRCLED LATIN SMALL LETTER D
5444        <e-o>        <U24D4>    CIRCLED LATIN SMALL LETTER E
5445        <f-o>        <U24D5>    CIRCLED LATIN SMALL LETTER F
5446        <g-o>        <U24D6>    CIRCLED LATIN SMALL LETTER G
5447        <h-o>        <U24D7>    CIRCLED LATIN SMALL LETTER H
5448        <i-o>        <U24D8>    CIRCLED LATIN SMALL LETTER I
5449        <j-o>        <U24D9>    CIRCLED LATIN SMALL LETTER J
5450        <k-o>        <U24DA>    CIRCLED LATIN SMALL LETTER K
5451        <l-o>        <U24DB>    CIRCLED LATIN SMALL LETTER L
5452        <m-o>        <U24DC>    CIRCLED LATIN SMALL LETTER M
5453        <n-o>        <U24DD>    CIRCLED LATIN SMALL LETTER N
5454        <o-o>        <U24DE>    CIRCLED LATIN SMALL LETTER O
5455        <p-o>        <U24DF>    CIRCLED LATIN SMALL LETTER P
5456        <q-o>        <U24E0>    CIRCLED LATIN SMALL LETTER Q
5457        <r-o>        <U24E1>    CIRCLED LATIN SMALL LETTER R
5458        <s-o>        <U24E2>    CIRCLED LATIN SMALL LETTER S
5459        <t-o>        <U24E3>    CIRCLED LATIN SMALL LETTER T
5460        <u-o>        <U24E4>    CIRCLED LATIN SMALL LETTER U
5461        <v-o>        <U24E5>    CIRCLED LATIN SMALL LETTER V
5462        <w-o>        <U24E6>    CIRCLED LATIN SMALL LETTER W
5463        <x-o>        <U24E7>    CIRCLED LATIN SMALL LETTER X
5464        <y-o>        <U24E8>    CIRCLED LATIN SMALL LETTER Y
5465        <z-o>        <U24E9>    CIRCLED LATIN SMALL LETTER Z
5466        <0-o>        <U24EA>    CIRCLED DIGIT ZERO
5467        <hh>         <U2500>    BOX DRAWINGS LIGHT HORIZONTAL
5468        <HH->        <U2501>    BOX DRAWINGS HEAVY HORIZONTAL
5469        <vv>         <U2502>    BOX DRAWINGS LIGHT VERTICAL
5470        <VV->        <U2503>    BOX DRAWINGS HEAVY VERTICAL
5471        <3->         <U2504>    BOX DRAWINGS LIGHT TRIPLE DASH HORIZONTAL
5472        <3_>         <U2505>    BOX DRAWINGS HEAVY TRIPLE DASH HORIZONTAL
5473        <3!>         <U2506>    BOX DRAWINGS LIGHT TRIPLE DASH VERTICAL
5474        <3//>        <U2507>    BOX DRAWINGS HEAVY TRIPLE DASH VERTICAL
5475        <4->         <U2508>    BOX DRAWINGS LIGHT QUADRUPLE DASH HORIZONTAL
5476        <4_>         <U2509>    BOX DRAWINGS HEAVY QUADRUPLE DASH HORIZONTAL
5477        <4!>         <U250A>    BOX DRAWINGS LIGHT QUADRUPLE DASH VERTICAL
5478        <4//>        <U250B>    BOX DRAWINGS HEAVY QUADRUPLE DASH VERTICAL
5479        <dr>         <U250C>    BOX DRAWINGS LIGHT DOWN AND RIGHT
5480        <dR->        <U250D>    BOX DRAWINGS DOWN LIGHT AND RIGHT HEAVY
5481        <Dr->        <U250E>    BOX DRAWINGS DOWN HEAVY AND RIGHT LIGHT
5482        <DR->        <U250F>    BOX DRAWINGS HEAVY DOWN AND RIGHT
5483        <dl>         <U2510>    BOX DRAWINGS LIGHT DOWN AND LEFT
5484        <dL->        <U2511>    BOX DRAWINGS DOWN LIGHT AND LEFT HEAVY
5485        <Dl->        <U2512>    BOX DRAWINGS DOWN HEAVY AND LEFT LIGHT
5486        <LD->        <U2513>    BOX DRAWINGS HEAVY DOWN AND LEFT
5487        <ur>         <U2514>    BOX DRAWINGS LIGHT UP AND RIGHT
5488        <uR->        <U2515>    BOX DRAWINGS UP LIGHT AND RIGHT HEAVY
5489        <Ur->        <U2516>    BOX DRAWINGS UP HEAVY AND RIGHT LIGHT
5490        <UR->        <U2517>    BOX DRAWINGS HEAVY UP AND RIGHT
5491        <ul>         <U2518>    BOX DRAWINGS LIGHT UP AND LEFT
5492        <uL->        <U2519>    BOX DRAWINGS UP LIGHT AND LEFT HEAVY
```

```
5493    <Ul->    <U251A>    BOX DRAWINGS UP HEAVY AND LEFT LIGHT
5494    <UL->    <U251B>    BOX DRAWINGS HEAVY UP AND LEFT
5495    <vr>     <U251C>    BOX DRAWINGS LIGHT VERTICAL AND RIGHT
5496    <vR->    <U251D>    BOX DRAWINGS VERTICAL LIGHT AND RIGHT HEAVY
5497    <Udr>    <U251E>    BOX DRAWINGS UP HEAVY AND RIGHT DOWN LIGHT
5498    <uDr>    <U251F>    BOX DRAWINGS DOWN HEAVY AND RIGHT UP LIGHT
5499    <Vr->    <U2520>    BOX DRAWINGS VERTICAL HEAVY AND RIGHT LIGHT
5500    <UdR>    <U2521>    BOX DRAWINGS DOWN LIGHT AND RIGHT UP HEAVY
5501    <uDR>    <U2522>    BOX DRAWINGS UP LIGHT AND RIGHT DOWN HEAVY
5502    <VR->    <U2523>    BOX DRAWINGS HEAVY VERTICAL AND RIGHT
5503    <vl>     <U2524>    BOX DRAWINGS LIGHT VERTICAL AND LEFT
5504    <vL->    <U2525>    BOX DRAWINGS VERTICAL LIGHT AND LEFT HEAVY
5505    <Udl>    <U2526>    BOX DRAWINGS UP HEAVY AND LEFT DOWN LIGHT
5506    <uDl>    <U2527>    BOX DRAWINGS DOWN HEAVY AND LEFT UP LIGHT
5507    <Vl->    <U2528>    BOX DRAWINGS VERTICAL HEAVY AND LEFT LIGHT
5508    <UdL>    <U2529>    BOX DRAWINGS DOWN LIGHT AND LEFT UP HEAVY
5509    <uDL>    <U252A>    BOX DRAWINGS UP LIGHT AND LEFT DOWN HEAVY
5510    <VL->    <U252B>    BOX DRAWINGS HEAVY VERTICAL AND LEFT
5511    <dh>     <U252C>    BOX DRAWINGS LIGHT DOWN AND HORIZONTAL
5512    <dLr>    <U252D>    BOX DRAWINGS LEFT HEAVY AND RIGHT DOWN LIGHT
5513    <dlR>    <U252E>    BOX DRAWINGS RIGHT HEAVY AND LEFT DOWN LIGHT
5514    <dH->    <U252F>    BOX DRAWINGS DOWN LIGHT AND HORIZONTAL HEAVY
5515    <Dh->    <U2530>    BOX DRAWINGS DOWN HEAVY AND HORIZONTAL LIGHT
5516    <DLr>    <U2531>    BOX DRAWINGS RIGHT LIGHT AND LEFT DOWN HEAVY
5517    <DlR>    <U2532>    BOX DRAWINGS LEFT LIGHT AND RIGHT DOWN HEAVY
5518    <DH->    <U2533>    BOX DRAWINGS HEAVY DOWN AND HORIZONTAL
5519    <uh>     <U2534>    BOX DRAWINGS LIGHT UP AND HORIZONTAL
5520    <uLr>    <U2535>    BOX DRAWINGS LEFT HEAVY AND RIGHT UP LIGHT
5521    <ulR>    <U2536>    BOX DRAWINGS RIGHT HEAVY AND LEFT UP LIGHT
5522    <uH->    <U2537>    BOX DRAWINGS UP LIGHT AND HORIZONTAL HEAVY
5523    <Uh->    <U2538>    BOX DRAWINGS UP HEAVY AND HORIZONTAL LIGHT
5524    <ULr>    <U2539>    BOX DRAWINGS RIGHT LIGHT AND LEFT UP HEAVY
5525    <UlR>    <U253A>    BOX DRAWINGS LEFT LIGHT AND RIGHT UP HEAVY
5526    <UH->    <U253B>    BOX DRAWINGS HEAVY UP AND HORIZONTAL
5527    <vh>     <U253C>    BOX DRAWINGS LIGHT VERTICAL AND HORIZONTAL
5528    <vLr>    <U253D>    BOX DRAWINGS LEFT HEAVY AND RIGHT VERTICAL LIGHT
5529    <vlR>    <U253E>    BOX DRAWINGS RIGHT HEAVY AND LEFT VERTICAL LIGHT
5530    <vH->    <U253F>    BOX DRAWINGS VERTICAL LIGHT AND HORIZONTAL HEAVY
5531    <Udh>    <U2540>    BOX DRAWINGS UP HEAVY AND DOWN HORIZONTAL LIGHT
5532    <uDh>    <U2541>    BOX DRAWINGS DOWN HEAVY AND UP HORIZONTAL LIGHT
5533    <Vh->    <U2542>    BOX DRAWINGS VERTICAL HEAVY AND HORIZONTAL LIGHT
5534    <UdLr>   <U2543>    BOX DRAWINGS LEFT UP HEAVY AND RIGHT DOWN LIGHT
5535    <UdlR>   <U2544>    BOX DRAWINGS RIGHT UP HEAVY AND LEFT DOWN LIGHT
5536    <uDLr>   <U2545>    BOX DRAWINGS LEFT DOWN HEAVY AND RIGHT UP LIGHT
5537    <uDlR>   <U2546>    BOX DRAWINGS RIGHT DOWN HEAVY AND LEFT UP LIGHT
5538    <UdH>    <U2547>    BOX DRAWINGS DOWN LIGHT AND UP HORIZONTAL HEAVY
5539    <uDH>    <U2548>    BOX DRAWINGS UP LIGHT AND DOWN HORIZONTAL HEAVY
5540    <VLr>    <U2549>    BOX DRAWINGS RIGHT LIGHT AND LEFT VERTICAL HEAVY
5541    <VlR>    <U254A>    BOX DRAWINGS LEFT LIGHT AND RIGHT VERTICAL HEAVY
5542    <VH->    <U254B>    BOX DRAWINGS HEAVY VERTICAL AND HORIZONTAL
5543    <HH>     <U2550>    BOX DRAWINGS DOUBLE HORIZONTAL
5544    <VV>     <U2551>    BOX DRAWINGS DOUBLE VERTICAL
5545    <dR>     <U2552>    BOX DRAWINGS DOWN SINGLE AND RIGHT DOUBLE
5546    <Dr>     <U2553>    BOX DRAWINGS DOWN DOUBLE AND RIGHT SINGLE
5547    <DR>     <U2554>    BOX DRAWINGS DOUBLE DOWN AND RIGHT
5548    <dL>     <U2555>    BOX DRAWINGS DOWN SINGLE AND LEFT DOUBLE
5549    <Dl>     <U2556>    BOX DRAWINGS DOWN DOUBLE AND LEFT SINGLE
5550    <LD>     <U2557>    BOX DRAWINGS DOUBLE DOWN AND LEFT
5551    <uR>     <U2558>    BOX DRAWINGS UP SINGLE AND RIGHT DOUBLE
5552    <Ur>     <U2559>    BOX DRAWINGS UP DOUBLE AND RIGHT SINGLE
5553    <UR>     <U255A>    BOX DRAWINGS DOUBLE UP AND RIGHT
5554    <uL>     <U255B>    BOX DRAWINGS UP SINGLE AND LEFT DOUBLE
5555    <Ul>     <U255C>    BOX DRAWINGS UP DOUBLE AND LEFT SINGLE
5556    <UL>     <U255D>    BOX DRAWINGS DOUBLE UP AND LEFT
5557    <vR>     <U255E>    BOX DRAWINGS VERTICAL SINGLE AND RIGHT DOUBLE
5558    <Vr>     <U255F>    BOX DRAWINGS VERTICAL DOUBLE AND RIGHT SINGLE
5559    <VR>     <U2560>    BOX DRAWINGS DOUBLE VERTICAL AND RIGHT
5560    <vL>     <U2561>    BOX DRAWINGS VERTICAL SINGLE AND LEFT DOUBLE
5561    <Vl>     <U2562>    BOX DRAWINGS VERTICAL DOUBLE AND LEFT SINGLE
5562    <VL>     <U2563>    BOX DRAWINGS DOUBLE VERTICAL AND LEFT
5563    <dH>     <U2564>    BOX DRAWINGS DOWN SINGLE AND HORIZONTAL DOUBLE
5564    <Dh>     <U2565>    BOX DRAWINGS DOWN DOUBLE AND HORIZONTAL SINGLE
5565    <DH>     <U2566>    BOX DRAWINGS DOUBLE DOWN AND HORIZONTAL
5566    <uH>     <U2567>    BOX DRAWINGS UP SINGLE AND HORIZONTAL DOUBLE
5567    <Uh>     <U2568>    BOX DRAWINGS UP DOUBLE AND HORIZONTAL SINGLE
5568    <UH>     <U2569>    BOX DRAWINGS DOUBLE UP AND HORIZONTAL
5569    <vH>     <U256A>    BOX DRAWINGS VERTICAL SINGLE AND HORIZONTAL DOUBLE
5570    <Vh>     <U256B>    BOX DRAWINGS VERTICAL DOUBLE AND HORIZONTAL SINGLE
5571    <VH>     <U256C>    BOX DRAWINGS DOUBLE VERTICAL AND HORIZONTAL
5572    <FD>     <U2571>    BOX DRAWINGS LIGHT DIAGONAL UPPER RIGHT TO LOWER LEFT
5573    <BD>     <U2572>    BOX DRAWINGS LIGHT DIAGONAL UPPER LEFT TO LOWER RIGHT
5574    <TB>     <U2580>    UPPER HALF BLOCK
5575    <LB>     <U2584>    LOWER HALF BLOCK
5576    <FB>     <U2588>    FULL BLOCK
5577    <lB>     <U258C>    LEFT HALF BLOCK
5578    <RB>     <U2590>    RIGHT HALF BLOCK
5579    <.S>     <U2591>    LIGHT SHADE
5580    <:S>     <U2592>    MEDIUM SHADE
```

```
5581        <?S>        <U2593>    DARK SHADE
5582        <fS>        <U25A0>    BLACK SQUARE
5583        <OS>        <U25A1>    WHITE SQUARE
5584        <RO>        <U25A2>    WHITE SQUARE WITH ROUNDED CORNERS
5585        <Rr>        <U25A3>    WHITE SQUARE CONTAINING BLACK SMALL SQUARE
5586        <RF>        <U25A4>    SQUARE WITH HORIZONTAL FILL
5587        <RY>        <U25A5>    SQUARE WITH VERTICAL FILL
5588        <RH>        <U25A6>    SQUARE WITH ORTHOGONAL CROSSHATCH FILL
5589        <RZ>        <U25A7>    SQUARE WITH UPPER LEFT TO LOWER RIGHT FILL
5590        <RK>        <U25A8>    SQUARE WITH UPPER RIGHT TO LOWER LEFT FILL
5591        <RX>        <U25A9>    SQUARE WITH DIAGONAL CROSSHATCH FILL
5592        <sB>        <U25AA>    BLACK SMALL SQUARE
5593        <SR>        <U25AC>    BLACK RECTANGLE
5594        <Or>        <U25AD>    WHITE RECTANGLE
5595        <UT>        <U25B2>    BLACK UP-POINTING TRIANGLE
5596        <uT>        <U25B3>    WHITE UP-POINTING TRIANGLE
5597        <Tr>        <U25B7>    WHITE RIGHT-POINTING TRIANGLE
5598        <PR>        <U25BA>    BLACK RIGHT-POINTING POINTER
5599        <Dt>        <U25BC>    BLACK DOWN-POINTING TRIANGLE
5600        <dT>        <U25BD>    WHITE DOWN-POINTING TRIANGLE
5601        <Tl>        <U25C1>    WHITE LEFT-POINTING TRIANGLE
5602        <PL>        <U25C4>    BLACK LEFT-POINTING POINTER
5603        <Db>        <U25C6>    BLACK DIAMOND
5604        <Dw>        <U25C7>    WHITE DIAMOND
5605        <LZ>        <U25CA>    LOZENGE
5606        <0m>        <U25CB>    WHITE CIRCLE
5607        <0o>        <U25CE>    BULLSEYE
5608        <0M>        <U25CF>    BLACK CIRCLE
5609        <0L>        <U25D0>    CIRCLE WITH LEFT HALF BLACK
5610        <0R>        <U25D1>    CIRCLE WITH RIGHT HALF BLACK
5611        <Sn>        <U25D8>    INVERSE BULLET
5612        <Ic>        <U25D9>    INVERSE WHITE CIRCLE
5613        <Fd>        <U25E2>    BLACK LOWER RIGHT TRIANGLE
5614        <Bd>        <U25E3>    BLACK LOWER LEFT TRIANGLE
5615        <Ci>        <U25EF>    LARGE CIRCLE
5616        <*2>        <U2605>    BLACK STAR
5617        <*1>        <U2606>    WHITE STAR
5618        <TEL>       <U260E>    BLACK TELEPHONE
5619        <tel>       <U260F>    WHITE TELEPHONE
5620        <<H>        <U261C>    WHITE LEFT POINTING INDEX
5621        </>H>       <U261E>    WHITE RIGHT POINTING INDEX
5622        <0u>        <U263A>    WHITE SMILING FACE
5623        <0U>        <U263B>    BLACK SMILING FACE
5624        <SU>        <U263C>    WHITE SUN WITH RAYS
5625        <Fm>        <U2640>    FEMALE SIGN
5626        <Ml>        <U2642>    MALE SIGN
5627        <cS>        <U2660>    BLACK SPADE SUIT
5628        <cH>        <U2661>    WHITE HEART SUIT
5629        <cD>        <U2662>    WHITE DIAMOND SUIT
5630        <cC>        <U2663>    BLACK CLUB SUIT
5631        <cS->       <U2664>    WHITE SPADE SUIT
5632        <cH->       <U2665>    BLACK HEART SUIT
5633        <cD->       <U2666>    BLACK DIAMOND SUIT
5634        <cC->       <U2667>    WHITE CLUB SUIT
5635        <Md>        <U2669>    QUARTER NOTE
5636        <M8>        <U266A>    EIGHTH NOTE
5637        <M2>        <U266B>    BEAMED EIGHTH NOTES
5638        <M16>       <U266C>    BEAMED SIXTEENTH NOTES
5639        <Mb>        <U266D>    MUSIC FLAT SIGN
5640        <Mx>        <U266E>    MUSIC NATURAL SIGN
5641        <MX>        <U266F>    MUSIC SHARP SIGN
5642        <OK>        <U2713>    CHECK MARK
5643        <XX>        <U2717>    BALLOT X
5644        <-X>        <U2720>    MALTESE CROSS
5645        <IS>        <U3000>    IDEOGRAPHIC SPACE
5646        <,_>        <U3001>    IDEOGRAPHIC COMMA
5647        <._>        <U3002>    IDEOGRAPHIC FULL STOP
5648        <+">        <U3003>    DITTO MARK
5649        <JIS>       <U3004>    JAPANESE INDUSTRIAL STANDARD SYMBOL
5650        <*_>        <U3005>    IDEOGRAPHIC ITERATION MARK
5651        <;_>        <U3006>    IDEOGRAPHIC CLOSING MARK
5652        <0_>        <U3007>    IDEOGRAPHIC NUMBER ZERO
5653        <<+>        <U300A>    LEFT DOUBLE ANGLE BRACKET
5654        </>+>       <U300B>    RIGHT DOUBLE ANGLE BRACKET
5655        <<'>        <U300C>    LEFT CORNER BRACKET
5656        </>'>       <U300D>    RIGHT CORNER BRACKET
5657        <<">        <U300E>    LEFT WHITE CORNER BRACKET
5658        </>">       <U300F>    RIGHT WHITE CORNER BRACKET
5659        <(">        <U3010>    LEFT BLACK LENTICULAR BRACKET
5660        <)">        <U3011>    RIGHT BLACK LENTICULAR BRACKET
5661        <=T>        <U3012>    POSTAL MARK
5662        <=_>        <U3013>    GETA MARK
5663        <('>        <U3014>    LEFT TORTOISE SHELL BRACKET
5664        <)'>        <U3015>    RIGHT TORTOISE SHELL BRACKET
5665        <(I>        <U3016>    LEFT WHITE LENTICULAR BRACKET
5666        <)I>        <U3017>    RIGHT WHITE LENTICULAR BRACKET
5667        <-?>        <U301C>    WAVE DASH
5668        <=T:)>      <U3020>    POSTAL MARK FACE
5669        <A5>        <U3041>    HIRAGANA LETTER SMALL A
```

84

```
5670        <a5>        <U3042>    HIRAGANA LETTER A
5671        <I5>        <U3043>    HIRAGANA LETTER SMALL I
5672        <i5>        <U3044>    HIRAGANA LETTER I
5673        <U5>        <U3045>    HIRAGANA LETTER SMALL U
5674        <u5>        <U3046>    HIRAGANA LETTER U
5675        <E5>        <U3047>    HIRAGANA LETTER SMALL E
5676        <e5>        <U3048>    HIRAGANA LETTER E
5677        <O5>        <U3049>    HIRAGANA LETTER SMALL O
5678        <o5>        <U304A>    HIRAGANA LETTER O
5679        <ka>        <U304B>    HIRAGANA LETTER KA
5680        <ga>        <U304C>    HIRAGANA LETTER GA
5681        <ki>        <U304D>    HIRAGANA LETTER KI
5682        <gi>        <U304E>    HIRAGANA LETTER GI
5683        <ku>        <U304F>    HIRAGANA LETTER KU
5684        <gu>        <U3050>    HIRAGANA LETTER GU
5685        <ke>        <U3051>    HIRAGANA LETTER KE
5686        <ge>        <U3052>    HIRAGANA LETTER GE
5687        <ko>        <U3053>    HIRAGANA LETTER KO
5688        <go>        <U3054>    HIRAGANA LETTER GO
5689        <sa>        <U3055>    HIRAGANA LETTER SA
5690        <za>        <U3056>    HIRAGANA LETTER ZA
5691        <si>        <U3057>    HIRAGANA LETTER SI
5692        <zi>        <U3058>    HIRAGANA LETTER ZI
5693        <su>        <U3059>    HIRAGANA LETTER SU
5694        <zu>        <U305A>    HIRAGANA LETTER ZU
5695        <se>        <U305B>    HIRAGANA LETTER SE
5696        <ze>        <U305C>    HIRAGANA LETTER ZE
5697        <so>        <U305D>    HIRAGANA LETTER SO
5698        <zo>        <U305E>    HIRAGANA LETTER ZO
5699        <ta>        <U305F>    HIRAGANA LETTER TA
5700        <da>        <U3060>    HIRAGANA LETTER DA
5701        <ti>        <U3061>    HIRAGANA LETTER TI
5702        <di>        <U3062>    HIRAGANA LETTER DI
5703        <tU>        <U3063>    HIRAGANA LETTER SMALL TU
5704        <tu>        <U3064>    HIRAGANA LETTER TU
5705        <du>        <U3065>    HIRAGANA LETTER DU
5706        <te>        <U3066>    HIRAGANA LETTER TE
5707        <de>        <U3067>    HIRAGANA LETTER DE
5708        <to>        <U3068>    HIRAGANA LETTER TO
5709        <do>        <U3069>    HIRAGANA LETTER DO
5710        <na>        <U306A>    HIRAGANA LETTER NA
5711        <ni>        <U306B>    HIRAGANA LETTER NI
5712        <nu>        <U306C>    HIRAGANA LETTER NU
5713        <ne>        <U306D>    HIRAGANA LETTER NE
5714        <no>        <U306E>    HIRAGANA LETTER NO
5715        <ha>        <U306F>    HIRAGANA LETTER HA
5716        <ba>        <U3070>    HIRAGANA LETTER BA
5717        <pa>        <U3071>    HIRAGANA LETTER PA
5718        <hi>        <U3072>    HIRAGANA LETTER HI
5719        <bi>        <U3073>    HIRAGANA LETTER BI
5720        <pi>        <U3074>    HIRAGANA LETTER PI
5721        <hu>        <U3075>    HIRAGANA LETTER HU
5722        <bu>        <U3076>    HIRAGANA LETTER BU
5723        <pu>        <U3077>    HIRAGANA LETTER PU
5724        <he>        <U3078>    HIRAGANA LETTER HE
5725        <be>        <U3079>    HIRAGANA LETTER BE
5726        <pe>        <U307A>    HIRAGANA LETTER PE
5727        <ho>        <U307B>    HIRAGANA LETTER HO
5728        <bo>        <U307C>    HIRAGANA LETTER BO
5729        <po>        <U307D>    HIRAGANA LETTER PO
5730        <ma>        <U307E>    HIRAGANA LETTER MA
5731        <mi>        <U307F>    HIRAGANA LETTER MI
5732        <mu>        <U3080>    HIRAGANA LETTER MU
5733        <me>        <U3081>    HIRAGANA LETTER ME
5734        <mo>        <U3082>    HIRAGANA LETTER MO
5735        <yA>        <U3083>    HIRAGANA LETTER SMALL YA
5736        <ya>        <U3084>    HIRAGANA LETTER YA
5737        <yU>        <U3085>    HIRAGANA LETTER SMALL YU
5738        <yu>        <U3086>    HIRAGANA LETTER YU
5739        <yO>        <U3087>    HIRAGANA LETTER SMALL YO
5740        <yo>        <U3088>    HIRAGANA LETTER YO
5741        <ra>        <U3089>    HIRAGANA LETTER RA
5742        <ri>        <U308A>    HIRAGANA LETTER RI
5743        <ru>        <U308B>    HIRAGANA LETTER RU
5744        <re>        <U308C>    HIRAGANA LETTER RE
5745        <ro>        <U308D>    HIRAGANA LETTER RO
5746        <wA>        <U308E>    HIRAGANA LETTER SMALL WA
5747        <wa>        <U308F>    HIRAGANA LETTER WA
5748        <wi>        <U3090>    HIRAGANA LETTER WI
5749        <we>        <U3091>    HIRAGANA LETTER WE
5750        <wo>        <U3092>    HIRAGANA LETTER WO
5751        <n5>        <U3093>    HIRAGANA LETTER N
5752        <vu>        <U3094>    HIRAGANA LETTER VU
5753        <"5>        <U309B>    KATAKANA-HIRAGANA VOICED SOUND MARK
5754        <O5>        <U309C>    KATAKANA-HIRAGANA SEMI-VOICED SOUND MARK
5755        <*5>        <U309D>    HIRAGANA ITERATION MARK
5756        <+5>        <U309E>    HIRAGANA VOICED ITERATION MARK
5757        <a6>        <U30A1>    KATAKANA LETTER SMALL A
```

```
5758    <A6>     <U30A2>   KATAKANA LETTER A
5759    <i6>     <U30A3>   KATAKANA LETTER SMALL I
5760    <I6>     <U30A4>   KATAKANA LETTER I
5761    <u6>     <U30A5>   KATAKANA LETTER SMALL U
5762    <U6>     <U30A6>   KATAKANA LETTER U
5763    <e6>     <U30A7>   KATAKANA LETTER SMALL E
5764    <E6>     <U30A8>   KATAKANA LETTER E
5765    <o6>     <U30A9>   KATAKANA LETTER SMALL O
5766    <O6>     <U30AA>   KATAKANA LETTER O
5767    <Ka>     <U30AB>   KATAKANA LETTER KA
5768    <Ga>     <U30AC>   KATAKANA LETTER GA
5769    <Ki>     <U30AD>   KATAKANA LETTER KI
5770    <Gi>     <U30AE>   KATAKANA LETTER GI
5771    <Ku>     <U30AF>   KATAKANA LETTER KU
5772    <Gu>     <U30B0>   KATAKANA LETTER GU
5773    <Ke>     <U30B1>   KATAKANA LETTER KE
5774    <Ge>     <U30B2>   KATAKANA LETTER GE
5775    <Ko>     <U30B3>   KATAKANA LETTER KO
5776    <Go>     <U30B4>   KATAKANA LETTER GO
5777    <Sa>     <U30B5>   KATAKANA LETTER SA
5778    <Za>     <U30B6>   KATAKANA LETTER ZA
5779    <Si>     <U30B7>   KATAKANA LETTER SI
5780    <Zi>     <U30B8>   KATAKANA LETTER ZI
5781    <Su>     <U30B9>   KATAKANA LETTER SU
5782    <Zu>     <U30BA>   KATAKANA LETTER ZU
5783    <Se>     <U30BB>   KATAKANA LETTER SE
5784    <Ze>     <U30BC>   KATAKANA LETTER ZE
5785    <So>     <U30BD>   KATAKANA LETTER SO
5786    <Zo>     <U30BE>   KATAKANA LETTER ZO
5787    <Ta>     <U30BF>   KATAKANA LETTER TA
5788    <Da>     <U30C0>   KATAKANA LETTER DA
5789    <Ti>     <U30C1>   KATAKANA LETTER TI
5790    <Di>     <U30C2>   KATAKANA LETTER DI
5791    <TU>     <U30C3>   KATAKANA LETTER SMALL TU
5792    <Tu>     <U30C4>   KATAKANA LETTER TU
5793    <Du>     <U30C5>   KATAKANA LETTER DU
5794    <Te>     <U30C6>   KATAKANA LETTER TE
5795    <De>     <U30C7>   KATAKANA LETTER DE
5796    <To>     <U30C8>   KATAKANA LETTER TO
5797    <Do>     <U30C9>   KATAKANA LETTER DO
5798    <Na>     <U30CA>   KATAKANA LETTER NA
5799    <Ni>     <U30CB>   KATAKANA LETTER NI
5800    <Nu>     <U30CC>   KATAKANA LETTER NU
5801    <Ne>     <U30CD>   KATAKANA LETTER NE
5802    <No>     <U30CE>   KATAKANA LETTER NO
5803    <Ha>     <U30CF>   KATAKANA LETTER HA
5804    <Ba>     <U30D0>   KATAKANA LETTER BA
5805    <Pa>     <U30D1>   KATAKANA LETTER PA
5806    <Hi>     <U30D2>   KATAKANA LETTER HI
5807    <Bi>     <U30D3>   KATAKANA LETTER BI
5808    <Pi>     <U30D4>   KATAKANA LETTER PI
5809    <Hu>     <U30D5>   KATAKANA LETTER HU
5810    <Bu>     <U30D6>   KATAKANA LETTER BU
5811    <Pu>     <U30D7>   KATAKANA LETTER PU
5812    <He>     <U30D8>   KATAKANA LETTER HE
5813    <Be>     <U30D9>   KATAKANA LETTER BE
5814    <Pe>     <U30DA>   KATAKANA LETTER PE
5815    <Ho>     <U30DB>   KATAKANA LETTER HO
5816    <Bo>     <U30DC>   KATAKANA LETTER BO
5817    <Po>     <U30DD>   KATAKANA LETTER PO
5818    <Ma>     <U30DE>   KATAKANA LETTER MA
5819    <Mi>     <U30DF>   KATAKANA LETTER MI
5820    <Mu>     <U30E0>   KATAKANA LETTER MU
5821    <Me>     <U30E1>   KATAKANA LETTER ME
5822    <Mo>     <U30E2>   KATAKANA LETTER MO
5823    <YA>     <U30E3>   KATAKANA LETTER SMALL YA
5824    <Ya>     <U30E4>   KATAKANA LETTER YA
5825    <YU>     <U30E5>   KATAKANA LETTER SMALL YU
5826    <Yu>     <U30E6>   KATAKANA LETTER YU
5827    <YO>     <U30E7>   KATAKANA LETTER SMALL YO
5828    <Yo>     <U30E8>   KATAKANA LETTER YO
5829    <Ra>     <U30E9>   KATAKANA LETTER RA
5830    <Ri>     <U30EA>   KATAKANA LETTER RI
5831    <Ru>     <U30EB>   KATAKANA LETTER RU
5832    <Re>     <U30EC>   KATAKANA LETTER RE
5833    <Ro>     <U30ED>   KATAKANA LETTER RO
5834    <WA>     <U30EE>   KATAKANA LETTER SMALL WA
5835    <Wa>     <U30EF>   KATAKANA LETTER WA
5836    <Wi>     <U30F0>   KATAKANA LETTER WI
5837    <We>     <U30F1>   KATAKANA LETTER WE
5838    <Wo>     <U30F2>   KATAKANA LETTER WO
5839    <N6>     <U30F3>   KATAKANA LETTER N
5840    <Vu>     <U30F4>   KATAKANA LETTER VU
5841    <KA>     <U30F5>   KATAKANA LETTER SMALL KA
5842    <KE>     <U30F6>   KATAKANA LETTER SMALL KE
5843    <Va>     <U30F7>   KATAKANA LETTER VA
5844    <Vi>     <U30F8>   KATAKANA LETTER VI
5845    <Ve>     <U30F9>   KATAKANA LETTER VE
5846    <Vo>     <U30FA>   KATAKANA LETTER VO
```

```
5847    <.6>        <U30FB>    KATAKANA MIDDLE DOT
5848    <-6>        <U30FC>    KATAKANA-HIRAGANA PROLONGED SOUND MARK
5849    <*6>        <U30FD>    KATAKANA ITERATION MARK
5850    <+6>        <U30FE>    KATAKANA VOICED ITERATION MARK
5851    <b4>        <U3105>    BOPOMOFO LETTER B
5852    <p4>        <U3106>    BOPOMOFO LETTER P
5853    <m4>        <U3107>    BOPOMOFO LETTER M
5854    <f4>        <U3108>    BOPOMOFO LETTER F
5855    <d4>        <U3109>    BOPOMOFO LETTER D
5856    <t4>        <U310A>    BOPOMOFO LETTER T
5857    <n4>        <U310B>    BOPOMOFO LETTER N
5858    <l4>        <U310C>    BOPOMOFO LETTER L
5859    <g4>        <U310D>    BOPOMOFO LETTER G
5860    <k4>        <U310E>    BOPOMOFO LETTER K
5861    <h4>        <U310F>    BOPOMOFO LETTER H
5862    <j4>        <U3110>    BOPOMOFO LETTER J
5863    <q4>        <U3111>    BOPOMOFO LETTER Q
5864    <x4>        <U3112>    BOPOMOFO LETTER X
5865    <zh>        <U3113>    BOPOMOFO LETTER ZH
5866    <ch>        <U3114>    BOPOMOFO LETTER CH
5867    <sh>        <U3115>    BOPOMOFO LETTER SH
5868    <r4>        <U3116>    BOPOMOFO LETTER R
5869    <z4>        <U3117>    BOPOMOFO LETTER Z
5870    <c4>        <U3118>    BOPOMOFO LETTER C
5871    <s4>        <U3119>    BOPOMOFO LETTER S
5872    <a4>        <U311A>    BOPOMOFO LETTER A
5873    <o4>        <U311B>    BOPOMOFO LETTER O
5874    <e4>        <U311C>    BOPOMOFO LETTER E
5875    <eh4>       <U311D>    BOPOMOFO LETTER EH
5876    <ai>        <U311E>    BOPOMOFO LETTER AI
5877    <ei>        <U311F>    BOPOMOFO LETTER EI
5878    <au>        <U3120>    BOPOMOFO LETTER AU
5879    <ou>        <U3121>    BOPOMOFO LETTER OU
5880    <an>        <U3122>    BOPOMOFO LETTER AN
5881    <en>        <U3123>    BOPOMOFO LETTER EN
5882    <aN>        <U3124>    BOPOMOFO LETTER ANG
5883    <eN>        <U3125>    BOPOMOFO LETTER ENG
5884    <er>        <U3126>    BOPOMOFO LETTER ER
5885    <i4>        <U3127>    BOPOMOFO LETTER I
5886    <u4>        <U3128>    BOPOMOFO LETTER U
5887    <iu>        <U3129>    BOPOMOFO LETTER IU
5888    <v4>        <U312A>    BOPOMOFO LETTER V
5889    <nG>        <U312B>    BOPOMOFO LETTER NG
5890    <gn>        <U312C>    BOPOMOFO LETTER GN
5891    <(JU)>      <U321C>    PARENTHESIZED HANGUL CIEUC U
5892    <1c>        <U3220>    PARENTHESIZED IDEOGRAPH ONE
5893    <2c>        <U3221>    PARENTHESIZED IDEOGRAPH TWO
5894    <3c>        <U3222>    PARENTHESIZED IDEOGRAPH THREE
5895    <4c>        <U3223>    PARENTHESIZED IDEOGRAPH FOUR
5896    <5c>        <U3224>    PARENTHESIZED IDEOGRAPH FIVE
5897    <6c>        <U3225>    PARENTHESIZED IDEOGRAPH SIX
5898    <7c>        <U3226>    PARENTHESIZED IDEOGRAPH SEVEN
5899    <8c>        <U3227>    PARENTHESIZED IDEOGRAPH EIGHT
5900    <9c>        <U3228>    PARENTHESIZED IDEOGRAPH NINE
5901    <10c>       <U3229>    PARENTHESIZED IDEOGRAPH TEN
5902    <KSC>       <U327F>    KOREAN STANDARD SYMBOL
5903    <am>        <U33C2>    SQUARE AM
5904    <pm>        <U33D8>    SQUARE PM
5905    <ff>        <UFB00>    LATIN SMALL LIGATURE FF
5906    <fi>        <UFB01>    LATIN SMALL LIGATURE FI
5907    <fl>        <UFB02>    LATIN SMALL LIGATURE FL
5908    <ffi>       <UFB03>    LATIN SMALL LIGATURE FFI
5909    <ffl>       <UFB04>    LATIN SMALL LIGATURE FFL
5910    <St>        <UFB05>    LATIN SMALL LIGATURE LONG S T
5911    <st>        <UFB06>    LATIN SMALL LIGATURE ST
5912    <3+;>       <UFE7D>    ARABIC SHADDA MEDIAL FORM
5913    <aM.>       <UFE82>    ARABIC LETTER ALEF WITH MADDA ABOVE FINAL FORM
5914    <aH.>       <UFE84>    ARABIC LETTER ALEF WITH HAMZA ABOVE FINAL FORM
5915    <ah.>       <UFE88>    ARABIC LETTER ALEF WITH HAMZA BELOW FINAL FORM
5916    <a+->       <UFE8D>    ARABIC LETTER ALEF ISOLATED FORM
5917    <a+.>       <UFE8E>    ARABIC LETTER ALEF FINAL FORM
5918    <b+->       <UFE8F>    ARABIC LETTER BEH ISOLATED FORM
5919    <b+.>       <UFE90>    ARABIC LETTER BEH FINAL FORM
5920    <b+,>       <UFE91>    ARABIC LETTER BEH INITIAL FORM
5921    <b+;>       <UFE92>    ARABIC LETTER BEH MEDIAL FORM
5922    <tm->       <UFE93>    ARABIC LETTER TEH MARBUTA ISOLATED FORM
5923    <tm.>       <UFE94>    ARABIC LETTER TEH MARBUTA FINAL FORM
5924    <t+->       <UFE95>    ARABIC LETTER TEH ISOLATED FORM
5925    <t+.>       <UFE96>    ARABIC LETTER TEH FINAL FORM
5926    <t+,>       <UFE97>    ARABIC LETTER TEH INITIAL FORM
5927    <t+;>       <UFE98>    ARABIC LETTER TEH MEDIAL FORM
5928    <tk->       <UFE99>    ARABIC LETTER THEH ISOLATED FORM
5929    <tk.>       <UFE9A>    ARABIC LETTER THEH FINAL FORM
5930    <tk,>       <UFE9B>    ARABIC LETTER THEH INITIAL FORM
5931    <tk;>       <UFE9C>    ARABIC LETTER THEH MEDIAL FORM
5932    <g+->       <UFE9D>    ARABIC LETTER JEEM ISOLATED FORM
5933    <g+.>       <UFE9E>    ARABIC LETTER JEEM FINAL FORM
5934    <g+,>       <UFE9F>    ARABIC LETTER JEEM INITIAL FORM
```

```
5935        <g+;>      <UFEA0>    ARABIC LETTER JEEM MEDIAL FORM
5936        <hk->      <UFEA1>    ARABIC LETTER HAH ISOLATED FORM
5937        <hk.>      <UFEA2>    ARABIC LETTER HAH FINAL FORM
5938        <hk,>      <UFEA3>    ARABIC LETTER HAH INITIAL FORM
5939        <hk;>      <UFEA4>    ARABIC LETTER HAH MEDIAL FORM
5940        <x+->      <UFEA5>    ARABIC LETTER KHAH ISOLATED FORM
5941        <x+.>      <UFEA6>    ARABIC LETTER KHAH FINAL FORM
5942        <x+,>      <UFEA7>    ARABIC LETTER KHAH INITIAL FORM
5943        <x+;>      <UFEA8>    ARABIC LETTER KHAH MEDIAL FORM
5944        <d+->      <UFEA9>    ARABIC LETTER DAL ISOLATED FORM
5945        <d+.>      <UFEAA>    ARABIC LETTER DAL FINAL FORM
5946        <dk->      <UFEAB>    ARABIC LETTER THAL ISOLATED FORM
5947        <dk.>      <UFEAC>    ARABIC LETTER THAL FINAL FORM
5948        <r+->      <UFEAD>    ARABIC LETTER REH ISOLATED FORM
5949        <r+.>      <UFEAE>    ARABIC LETTER REH FINAL FORM
5950        <z+->      <UFEAF>    ARABIC LETTER ZAIN ISOLATED FORM
5951        <z+.>      <UFEB0>    ARABIC LETTER ZAIN FINAL FORM
5952        <s+->      <UFEB1>    ARABIC LETTER SEEN ISOLATED FORM
5953        <s+.>      <UFEB2>    ARABIC LETTER SEEN FINAL FORM
5954        <s+,>      <UFEB3>    ARABIC LETTER SEEN INITIAL FORM
5955        <s+;>      <UFEB4>    ARABIC LETTER SEEN MEDIAL FORM
5956        <sn->      <UFEB5>    ARABIC LETTER SHEEN ISOLATED FORM
5957        <sn.>      <UFEB6>    ARABIC LETTER SHEEN FINAL FORM
5958        <sn,>      <UFEB7>    ARABIC LETTER SHEEN INITIAL FORM
5959        <sn;>      <UFEB8>    ARABIC LETTER SHEEN MEDIAL FORM
5960        <c+->      <UFEB9>    ARABIC LETTER SAD ISOLATED FORM
5961        <c+.>      <UFEBA>    ARABIC LETTER SAD FINAL FORM
5962        <c+,>      <UFEBB>    ARABIC LETTER SAD INITIAL FORM
5963        <c+;>      <UFEBC>    ARABIC LETTER SAD MEDIAL FORM
5964        <dd->      <UFEBD>    ARABIC LETTER DAD ISOLATED FORM
5965        <dd.>      <UFEBE>    ARABIC LETTER DAD FINAL FORM
5966        <dd,>      <UFEBF>    ARABIC LETTER DAD INITIAL FORM
5967        <dd;>      <UFEC0>    ARABIC LETTER DAD MEDIAL FORM
5968        <tj->      <UFEC1>    ARABIC LETTER TAH ISOLATED FORM
5969        <tj.>      <UFEC2>    ARABIC LETTER TAH FINAL FORM
5970        <tj,>      <UFEC3>    ARABIC LETTER TAH INITIAL FORM
5971        <tj;>      <UFEC4>    ARABIC LETTER TAH MEDIAL FORM
5972        <zH->      <UFEC5>    ARABIC LETTER ZAH ISOLATED FORM
5973        <zH.>      <UFEC6>    ARABIC LETTER ZAH FINAL FORM
5974        <zH,>      <UFEC7>    ARABIC LETTER ZAH INITIAL FORM
5975        <zH;>      <UFEC8>    ARABIC LETTER ZAH MEDIAL FORM
5976        <e+->      <UFEC9>    ARABIC LETTER AIN ISOLATED FORM
5977        <e+.>      <UFECA>    ARABIC LETTER AIN FINAL FORM
5978        <e+,>      <UFECB>    ARABIC LETTER AIN INITIAL FORM
5979        <e+;>      <UFECC>    ARABIC LETTER AIN MEDIAL FORM
5980        <i+->      <UFECD>    ARABIC LETTER GHAIN ISOLATED FORM
5981        <i+.>      <UFECE>    ARABIC LETTER GHAIN FINAL FORM
5982        <i+,>      <UFECF>    ARABIC LETTER GHAIN INITIAL FORM
5983        <i+;>      <UFED0>    ARABIC LETTER GHAIN MEDIAL FORM
5984        <f+->      <UFED1>    ARABIC LETTER FEH ISOLATED FORM
5985        <f+.>      <UFED2>    ARABIC LETTER FEH FINAL FORM
5986        <f+,>      <UFED3>    ARABIC LETTER FEH INITIAL FORM
5987        <f+;>      <UFED4>    ARABIC LETTER FEH MEDIAL FORM
5988        <q+->      <UFED5>    ARABIC LETTER QAF ISOLATED FORM
5989        <q+.>      <UFED6>    ARABIC LETTER QAF FINAL FORM
5990        <q+,>      <UFED7>    ARABIC LETTER QAF INITIAL FORM
5991        <q+;>      <UFED8>    ARABIC LETTER QAF MEDIAL FORM
5992        <k+->      <UFED9>    ARABIC LETTER KAF ISOLATED FORM
5993        <k+.>      <UFEDA>    ARABIC LETTER KAF FINAL FORM
5994        <k+,>      <UFEDB>    ARABIC LETTER KAF INITIAL FORM
5995        <k+;>      <UFEDC>    ARABIC LETTER KAF MEDIAL FORM
5996        <l+->      <UFEDD>    ARABIC LETTER LAM ISOLATED FORM
5997        <l+.>      <UFEDE>    ARABIC LETTER LAM FINAL FORM
5998        <l+,>      <UFEDF>    ARABIC LETTER LAM INITIAL FORM
5999        <l+;>      <UFEE0>    ARABIC LETTER LAM MEDIAL FORM
6000        <m+->      <UFEE1>    ARABIC LETTER MEEM ISOLATED FORM
6001        <m+.>      <UFEE2>    ARABIC LETTER MEEM FINAL FORM
6002        <m+,>      <UFEE3>    ARABIC LETTER MEEM INITIAL FORM
6003        <m+;>      <UFEE4>    ARABIC LETTER MEEM MEDIAL FORM
6004        <n+->      <UFEE5>    ARABIC LETTER NOON ISOLATED FORM
6005        <n+.>      <UFEE6>    ARABIC LETTER NOON FINAL FORM
6006        <n+,>      <UFEE7>    ARABIC LETTER NOON INITIAL FORM
6007        <n+;>      <UFEE8>    ARABIC LETTER NOON MEDIAL FORM
6008        <h+->      <UFEE9>    ARABIC LETTER HEH ISOLATED FORM
6009        <h+.>      <UFEEA>    ARABIC LETTER HEH FINAL FORM
6010        <h+,>      <UFEEB>    ARABIC LETTER HEH INITIAL FORM
6011        <h+;>      <UFEEC>    ARABIC LETTER HEH MEDIAL FORM
6012        <w+->      <UFEED>    ARABIC LETTER WAW ISOLATED FORM
6013        <w+.>      <UFEEE>    ARABIC LETTER WAW FINAL FORM
6014        <j+->      <UFEEF>    ARABIC LETTER ALEF MAKSURA ISOLATED FORM
6015        <j+.>      <UFEF0>    ARABIC LETTER ALEF MAKSURA FINAL FORM
6016        <y+->      <UFEF1>    ARABIC LETTER YEH ISOLATED FORM
6017        <y+.>      <UFEF2>    ARABIC LETTER YEH FINAL FORM
6018        <y+,>      <UFEF3>    ARABIC LETTER YEH INITIAL FORM
6019        <y+;>      <UFEF4>    ARABIC LETTER YEH MEDIAL FORM
6020        <lM->      <UFEF5>    ARABIC LIGATURE LAM WITH ALEF WITH MADDA ABOVE ISOLATED FORM
6021        <lM.>      <UFEF6>    ARABIC LIGATURE LAM WITH ALEF WITH MADDA ABOVE FINAL FORM
6022        <lH->      <UFEF7>    ARABIC LIGATURE LAM WITH ALEF WITH HAMZA ABOVE ISOLATED FORM
6023        <lH.>      <UFEF8>    ARABIC LIGATURE LAM WITH ALEF WITH HAMZA ABOVE FINAL FORM
```

```
6024    <lh->      <UFEF9>   ARABIC LIGATURE LAM WITH ALEF WITH HAMZA BELOW ISOLATED FORM
6025    <lh.>      <UFEFA>   ARABIC LIGATURE LAM WITH ALEF WITH HAMZA BELOW FINAL FORM
6026    <la->      <UFEFB>   ARABIC LIGATURE LAM WITH ALEF ISOLATED FORM
6027    <la.>      <UFEFC>   ARABIC LIGATURE LAM WITH ALEF FINAL FORM
6028    <H->       <U0023>   NUMBER SIGN
6029    <!S>       <U0024>   DOLLAR SIGN
6030    <@>        <U0040>   COMMERCIAL AT
6031    <Oa>       <U0040>   COMMERCIAL AT
6032    <!C>       <U00A2>   CENT SIGN
6033    <L->       <U00A3>   POUND SIGN
6034    <Xo>       <U00A4>   CURRENCY SIGN
6035    <Y->       <U00A5>   YEN SIGN
6036    <!B>       <U00A6>   BROKEN BAR
6037    <So>       <U00A7>   SECTION SIGN
6038    <7!>       <U00AC>   NOT SIGN
6039    <9I>       <U00B6>   PILCROW SIGN
6040    <_->       <U2500>   BOX DRAWINGS LIGHT HORIZONTAL
6041    <_=>       <U2501>   BOX DRAWINGS HEAVY HORIZONTAL
6042    <_!>       <U2502>   BOX DRAWINGS LIGHT VERTICAL
6043    <_V/>>     <U250C>   BOX DRAWINGS LIGHT DOWN AND RIGHT
6044    <_V<w>     <U2510>   BOX DRAWINGS LIGHT DOWN AND LEFT
6045    <_A/>>     <U2514>   BOX DRAWINGS LIGHT UP AND RIGHT
6046    <_A<>      <U2518>   BOX DRAWINGS LIGHT UP AND LEFT
6047    <_!/>>     <U251C>   BOX DRAWINGS LIGHT VERTICAL AND RIGHT
6048    <_!<>      <U2524>   BOX DRAWINGS LIGHT VERTICAL AND LEFT
6049    <_V->      <U252C>   BOX DRAWINGS LIGHT DOWN AND HORIZONTAL
6050    <_-A>      <U2534>   BOX DRAWINGS LIGHT UP AND HORIZONTAL
6051    <_!->      <U253C>   BOX DRAWINGS LIGHT VERTICAL AND HORIZONTAL
6052    <_/>//>    <U2571>   BOX DRAWINGS LIGHT DIAGONAL UPPER RIGHT TO LOWER LEFT
6053    <_<\>      <U2572>   BOX DRAWINGS LIGHT DIAGONAL UPPER LEFT TO LOWER RIGHT
6054    <_./>//>   <U25E2>   BLACK LOWER RIGHT TRIANGLE
6055    <_.<\>     <U25E3>   BLACK LOWER LEFT TRIANGLE
6056    <_d!>      <U266A>   EIGHTH NOTE
6057
6058
```

7 CONFORMANCE

7.1 FDCC-set

A FDCC-set description is conforming to this Technical Report if it meets the
requirements in clause 4.

7.2 FDCC-set category

Conformance can be claimed for a category description against each of the clauses 4.3
thru 4.12, and then the requirements of clause 4.1 shall also be met, and a
LC_IDENTIFICATION category as described in clause 4.2 shall be specified.

7.3 Charmap

A charmap description is conforming to this Technical Report if it meets the requirements
in clause 5.

7.4 Repertoiremap

A repertoiremap description is conforming to this Technical Report if it meets the
requirements in clause 6.

<div align="center">

Annex A

(informative)

Differences from the ISO/IEC 9945-2 standard

</div>

This Technical Report originated from the locale and charmap specifications in the ISO/IEC 9945-2 standard, and it intends to be backwards compatible, so that what is conformant to that standard should also be conformant to this Technical Report.

A number of enhancements have been done and a number of restrictions have been lifted in comparison to the POSIX standard:

A.1   Restrictions removed

1. Dependence on specific meaning of the character NUL as termination of a string (from the C standard) has been removed, to cater for other programming languages than C.

A.2   Enhancements

1. A description of a "repertoiremap" definition was added to facilitate descriptions of FDCC-sets without charmaps, and also to provide binding from a FDCC-set using one set of character names to charmaps using another naming set.

2. The specific POSIX locale has been replaced with the "i18n" FDCC-set, defined on the repertoire on ISO/IEC 10646.

3. Transliteration support has been added in the LC_CTYPE category.

4. Terminology has been aligned with ISO/IEC TR 11017, especially the POSIX term "locale" has been changed to "FDCC-set".

5. A date escape format "%F" has been added for ISO 8601 dates, and another date escape format "%f" has been added for weekday number with Monday being the first day of the week.

6. Added to LC_MONETARY to accommodate differences between local and international formats:
> int_p_cs_precedes
> int_p_sep_by_space
> int_n_cs_precedes
> int_n_sep_by_space

7. Section symbols have been added via the "section-symbol" keyword in the LC_COLLATE category.

8. The "order_start" keyword has got an optional "section-symbol" identifier

9. The keywords "reorder-sections-after" and "reorder-sections_end" have been introduce to reorder sections.

6132   10. Symbolic elipsises (both decimal and hexadecimal) has been introduced as a notation.
6133

6134   11. The "print" CTYPE class includes automatically all "graph" characters.
6135

6136   12. The <Uxxxx> and <Uxxxxxxxx> notations have been introduced as predefined
6137   symbolic character names, together with a number of symbolic character names derived
6138   from POSIX and the Internet.
6139

6140   13. New categories LC_IDENTIFICATION, LC_PAPER, LC_NAME, LC_ADDRESS,
6141   and LC_TELEPHONE, have been introduced.
6142

6143   14. The LC_CTYPE has got support for new classes, via the new keywords class and
6144   map, which corresponds to the C standard library functions iswctype() and towctrans()
6145   respectively.
6146

6147   15. The "digit" keyword now supports digits for multiple scripts.
6148

6149   16. The LC_MONETARY category provides support for multiple currencies, such as the
6150   native currency and the Euro in some European countries.
6151

6152   17. The LC_TIME has got a number of enhancements to cater for alternate calendars, and
6153   timezone information may be given.
6154

6155   18. The charmap specification has been enhanced to support ISO 2022.

<div align="center">

Annex B

(informative)

Rationale

</div>

B.1 FDCC-set Rationale

The description of FDCC-sets is based on work performed in the UniForum Technical Committee Subcommittee on Internationalisation and on POSIX. Wherever appropriate, keywords were taken from the C Standard or the ISO/IEC 9945-2:1993 POSIX standard. The C and POSIX term "locale" has been changed into the term "FDCC-set" from ISO/IEC TR 11017 to align with that specification.

The POSIX utility "localedef" compiles locale sources into object files. The "object" definitions need not be portable, as long as "source" definitions are. Strictly speaking, "source" definitions are portable only between applications using the same character set(s). Such "source" definitions can, if they use symbolic names only, easily be ported between systems using different code sets as long as the characters in the portable character set (ISO 646) have common values between the code sets; this is frequently the case in historical applications. Of course, this requires that the symbolic names used for characters outside the portable character set are identical between character sets.

To avoid confusion between an octal constant and a backreference, the octal, hexadecimal, and decimal constants must contain at least two digits. As single-digit constants are relatively rare, this should not impose any significant hardship. Each of the constants includes "two or more" digits to account for systems in which the byte size is larger than eight bits. For example, an ISO/IEC 10646 system that has defined 16-bit bytes may require six octal, four hexadecimal, and five decimal digits, for some coded characters.

As an international (ISO/IEC) Technical Report this Technical Report should follow the ISO/IEC guidelines, including the ISO/IEC TR 10176. This TR has a rule that characters outside the invariant part of ISO/IEC 646 should not be used in portable specifications. The backslash and the number-sign character are not in the invariant part. As far as general usage of these symbols, they are covered by the "grandfather clause" specifying previous practise in international standards and in the industry such as in specifications from The Open Group, but for newly defined interfaces, ISO has requested that specifications provide alternate representations, and this Technical Report then follows POSIX for backward compatibility. Consequently, while the default escape character remains the backslash, and the default comment character is the number-sign, applications are required to recognize alternative representations, identified in the applicable source text via the "escape_char" and "comment_char" keywords.

B.1.1 LC_IDENTIFICATION Rationale.

The LC_IDENTIFICATION category gives meta-information on the FDCC-set, such as who created it, and what is the level of conformance for each of the FDCC sets.

B.1.2   LC_CTYPE Rationale

The LC_CTYPE category primarily is used to define the encoding-independent aspects of a character set, such as character classification. In addition, certain encoding-dependent characteristics are also defined for an application via the LC_CTYPE category. This Technical Report does not mandate that the encoding used in the FDCC-set is the same as the one used by the application, because an application may decide that it is advantageous to define a FDCC-set in a system-wide encoding rather than having multiple, logically identical FDCC-sets in different encodings, and to convert from the application encoding to the system-wide encoding on usage. Other applications could require encoding-dependent FDCC-sets. In either case, the LC_CTYPE attributes that are directly dependent on the encoding, such as "mb_cur_max" and the display width of characters, are not user-specifiable in a locale source, and are consequently not defined as keywords.

As the LC_CTYPE character classes are based on the C Standard character-class definition, the category does not support multicharacter elements. For instance, the German character <sharp-s> is traditionally classified as a lowercase letter. There is no corresponding uppercase letter; in proper capitalization of German text the <sharp-s> will be replaced by SS; i.e., by two characters. This kind of conversion is outside the scope of the "toupper" and "tolower" keywords.

The character classes "digit", "xdigit", "lower", "upper", and "space" have a set of automatically included characters. These only need to be specified if the character values (i.e. encoding) differs from the application default values. The definition of character class "digit" allows alternate digits (e.g., Hindi) to be specified here. The definition of character class "xdigit" requires that the characters included in character class "digit" are included here also, and allows for different symbols for the hexadecimal digits 10 through 15.

The "combining" and "combining-level3" classes are an IT-enablement of ISO/IEC 10646 definitions of combining characters. These can be used to check identifiers for consistence with the guidelines given in TR 10176 annex A.

B.1.3   LC_COLLATE Rationale.

The LC_COLLATE category governs the collation order in the FDCC-set, and may thus be useful for the processing of the ISO/IEC 14651 string ordering and comparison standard, the C Standard strxfrm() and strcoll() functions, as well as a number of ISO/IEC 9945-2:1993 POSIX utilities.

The rules governing collation depends to some extent on the use. At least five different levels of increasing complex collation rules can be distinguished:

(1)   Byte/machine code order. This is the historical collation order in the UNIX system and many proprietary operating systems. Collation is here done character by character, without any regard to context. The primary virtue is that it usually is quite fast, and also completely deterministic; it works well when the native machine collation sequence matches the user expectations.

(2)   Character order. On this level, collation is also done character by character, without regard to context. The order between characters is, however, not deter-

6256   mined by the code values, but on the user's expectations of the correct order
6257   between characters. In addition, such a (simple) collation order can specify that
6258   certain characters collate equal (e.g., upper and lowercase letters).
6259   (3) String ordering. On this level, entire strings are compared based on relatively
6260   straightforward rules. At this level, several "passes" may be required to deter-
6261   mine the order between two strings. Characters may be ignored in some passes,
6262   but not in others; the strings may be compared in different directions; and
6263   simple string substitutions may be made before strings are compared. This level
6264   is best described as "dictionary" ordering; it is based on the spelling, not the
6265   pronunciation, or meaning, of the words.
6266   (4) Text search ordering. This is a further refinement of the previous level, best de-
6267   scribed as "telephone book ordering"; some common homonyms (words spelled
6268   differently but with same pronunciation) are collated together; numbers are
6269   collated as if spelled with words, and so on.
6270   (5) Semantic level ordering. Words and strings are collated based on their meaning;
6271   entire words (such as "the") are eliminated, the ordering is not deterministic.
6272   This may requires special software, and is highly dependent on the intended
6273   use.
6274

6275   While the historical collation order formally is at level 1, for the English language it
6276   corresponds roughly to elements at level 2. The user expects to see the output from the
6277   "ls" utility sorted very much as it would be in a dictionary. While telephone book ordering
6278   would be an optimal goal for standard collation, this was ruled out as the order would be
6279   language dependent. Furthermore, a requirement was that the order must be determined
6280   solely from the text string and the collation rules; no external information (e.g., "pronu-
6281   nciation dictionaries") could be required.
6282

6283   As a result, the goal for the collation support is at level 3. This also matches the re-
6284   quirements for the Canadian collation order standard, as well as other, known collation
6285   requirements for alphabetic scripts. It specifically rules out collation based on pronun-
6286   ciation rules, or based on semantic analysis of the text. The syntax for the LC_COLLATE
6287   category source is the result of a cooperative effort between representatives for many
6288   countries and organizations working with international issues, such as UniForum, X/Open,
6289   and ISO, and it meets the requirements for level 3, and has been verified to produce the
6290   correct result with examples based on Canadian and Danish collation order.
6291

6292   The directives that can be specified in an operand to the order_start keyword are based on
6293   the requirements specified in several proposed standards and in customary use. The
6294   following is a rephrasing of rules defined for "lexical ordering in English and French" by
6295   the Canadian Standards Association (text is brackets is rephrased):
6296

6297   (1) Once special characters (punctuation) have been removed from original strings,
6298   the ordering is determined by scanning forward (left to right) [disregarding case
6299   and diacriticals].
6300   (2) In case of equivalence, special characters are once again removed from original
6301   strings and the ordering is determined scanning backward (starting from the
6302   rightmost character of the string and back), character by character, (disregarding
6303   case but considering diacriticals).
6304   (3) In case of repeated equivalence, special characters are removed again from
6305   original strings and the ordering is determined scanning forward, character by

6306   character, (considering both case and diacriticals).

6307   (4)   If there is still an ordering equivalence after rules (1) through (3) have been

6308   applied, then only special characters and the position they occupy in the string

6309   are considered to determine ordering. The string that has a special character in

6310   the lowest position comes first. If two strings have a special character in the

6311   same position, the character [with the lowest collation value] comes first. In

6312   case of equality, the other special characters are considered until there is a

6313   difference or all special characters have been exhausted.

6314

6315 It is estimated that the Technical Report covers the requirements for all European

6316 languages, and no particular problems are anticipated for Cyrillic or Middle Eastern

6317 scripts.

6318

6319 The Far East (particularly Japanese/Chinese) collations are often based on contextual

6320 information. In Japan, collations of strings containing CJK characters (ideograms) are

6321 often done considering some related information such as pronunciation, which needs a

6322 bulk dictionary (and some common sense). Such collation, in general, falls outside the

6323 desired goal of this Technical Report, and this Technical Report can support only a

6324 restricted of collations used in Japan. There are, however, several other collation rules

6325 (stroke/radical, or "most common pronunciation") which can be supported with the

6326 mechanism described here. Previous drafts contained a substitute statement, which

6327 performed a regular expression style replacement before string compares. It has been

6328 withdrawn based on balloter objections that it was not required for the types of ordering

6329 this Technical Report is aimed at.

6330

6331 The character (and collating element) order is defined by the order in which characters and

6332 elements are specified between the order_start and order_end keywords. This character

6333 order is used in range expressions in regular expressions. Weights assigned to the charac-

6334 ters and elements define the collation sequence; in the absence of weights, the character

6335 order is also the collation sequence.

6336

6337 The position keyword was introduced to provide the capability to consider, in a compare,

6338 the relative position of non-IGNOREd characters. As an example, consider the two strings

6339 "o-ring" and "or-ing". Assuming the hyphen is IGNOREd on the first pass, the two strings

6340 will compare equal, and the position of the hyphen is immaterial. On second pass, all

6341 characters except the hyphen are IGNOREd, and in the normal case the two strings would

6342 again compare equal. By taking position into account, the first collates before the second.

6343

6344 **B.1.3.1   "reorder-after" rationale**

6345

6346 Much work has been done on FDCC-sets, making them quite general. The ISO/IEC 9945-

6347 2:1993 POSIX standard introduced a "copy" command for all categories of the POSIX

6348 locale. This is useful for many purposes and it ensures that two FDCC-sets are equivalent

6349 for this category. A further step in building on previous FDCC-set work is defined in this

6350 Technical Report.

6351

6352 Collating sequences often vary a bit from country to country, and from language to

6353 language, but generally much of the collating sequence is the same. For example the

6354 Danish sequence is for the most part the same as the German or English collation, but for

6355 about a dozen letters it differs. The same can be said for Swedish or Hungarian: generally

6356   the Latin collating sequence is the same, but a few characters are different.
6357
6358   This Technical Report defines a FDCC-set defined on the character repertoire of the
6359   ISO/IEC 10646 standard, in a character set independent way. The intention is that some of
6360   the information from this FDCC-set will be acceptable in many cultures, and that it can
6361   serve as the basis for modifications in other cultures, to obtain a culturally acceptable
6362   specification. Using the "reorder-after" construct will also help improve the overview of
6363   what the changes really are for implementers and other users.
6364
6365   An example of the use of the "reorder-after" construct is the following. A default
6366   international ordering for the Latin alphabet may be adequate for Danish, with the
6367   exception of the collation rules for the letters Ü, ü, Æ, æ, Ä, ä, Ø, ø, Ö, ö, Å and å. By
6368   applying the "reorder-after" construct, the Danish specification can be made more easily
6369   by copying and reordering the existing international specification, rather than specifying
6370   collation parameters for all Latin letters (with or without diacritics). There is no obligation
6371   for Denmark to take this approach, but the "reorder-after" construct provides the
6372   mechanism for doing so if it is deemed desirable.

6373
6374

B.1.3.2   awk script for "reorder-after" construct

6376

A script has been written in the "awk" language defined in the POSIX standard ISO/IEC
9945-2 to implement the "reorder-after" construct. It functions as follows: It reads all of
the FDCC-set and if in the LC_COLLATE category, it processes the line, else it just
outputs the line. For the LC_COLLATE category it reads the lines and puts it into a
double linked list of strings identified by a line number; at the end of the LC_COLLATE
category all the lines are output. If the line is a "copy" keyword and it reads the file
referenced, extracting the LC_COLLATE section of the file in to the list of strings. If the
line is a "reorder-after" keyword, it sets a pointer to be the line number of the symbol to
of the "reorder-after" keyword. If the line is part of the "reorder-after" specification, it is
entered into the double linked list at this point, and the previous entry in the double linked
list for the <collation-element> is removed from the list. A "reorder-end" keyword
terminates the reordering.

```
BEGIN { comment = "%"; back[0]= follow[0] = 0; }
/LC_COLLATE/ { coll=1 }
/END LC_COLLATE/ { coll=0; for (lnr= 1; lnr; lnr= follow[lnr]) print cont[lnr] }

{ if (coll == 0) print $0  ;
    else { if ($1 == "copy")    {
            file = $2
            while (getline < file )
            if ( $1 == "LC_COLLATE" ) copy_lc = 1
            else if ( $1 == "END" && $2 == "LC_COLLATE" ) copy_lc =0
            else if (copy_lc) {
                    lnr++
                    follow[lnr-1] = lnr; back [ lnr ] = lnr-1
                    cont[lnr] = $0; symb[ $1 ] = lnr
            }
            close (file )
    }
    else if ($1 == "reorder-after") { ra=1 ; after = symb [ $2 ] }
    else if ($1 == "reorder-end") ra = 0
    else {
            lnr++
            if (ra) follow [ lnr ] = follow [ after ]
            if (ra) back [ follow [ after ] ] = lnr
            follow[after] = lnr; back [ lnr ] = after
            cont[lnr] = $0
            if ( ra && $1 != comment && $1 != "" )  {
                    old = symb [ $1 ];
                    follow [ back [ old ] ] = follow [ old ];
                    back [ follow [ old ] ] = back [ old ];
                    symb[ $1 ] = lnr;
            }
            after = lnr
    }
    }
}
```

6425

```
6426    B.1.3.3  Sample FDCC-set specification for Danish
6427
6428    escape_char /
6429    comment_char %
6430    repertoiremap "i18nrep"
6431    charset "ISO_8859-1:1987"
6432    % Distribution and use is free, also
6433    % for commercial purposes.
6434
6435    LC_VERSION
6436    title           "Danish language FDCC-set for Denmark"
6437    source          "Danish Standards Association"
6438    address         "Kollegievej 6, DK-2920 Charlottenlund, Danmark"
6439    contact         "Keld Simonsen"
6440    email           "Keld.Simonsen@dkuug.dk"
6441    tel             "+45 - 3996-6101"
6442    fax             "+45 - 3996-6202"
6443    language        "da"
6444    territory       "DK"
6445    revision        "4.2"
6446    date            "1997-12-22"
6447
6448    category        i18n:1998;LC_IDENTIFICATION
6449    category        i18n:1998;LC_CTYPE
6450    category        i18n:1998;LC_COLLATE
6451    category        i18n:1998;LC_TIME
6452    category        posix:1993;LC_NUMERIC
6453    category        i18n:1998;LC_MONETARY
6454    category        posix:1993;LC_MESSAGES
6455    category        i18n:1998;LC_PAPER
6456    category        i18n:1998;LC_NAME
6457    category        i18n:1998;LC_ADDRESS
6458    category        i18n:1998;LC_TELEPHONE
6459
6460    END LC_VERSION
6461
6462    LC_CTYPE
6463    copy "i18n"
6464    END LC_CTYPE
6465
6466    LC_COLLATE
6467    % The ordering algorithm is in accordance
6468    % with Danish Standard DS 377 (1980)
6469    % and the Danish Orthography Dictionary
6470    % (Retskrivningsordbogen, 2. udgave, 1996).
6471    % It is also in accordance with
6472    % Greenlandic orthography.
6473
6474    collating-element <A-A> from "<A><A>"
6475    collating-element <A-a> from "<A><a>"
6476    collating-element <a-A> from "<a><A>"
6477    collating-element <a-a> from "<a><a>"
6478    copy i18n
6479    reorder-after <CAPITAL>
6480    <CAPITAL>
6481    <CAPITAL-SMALL>
6482    <SMALL-CAPITAL>
6483    <SMALL>
6484    reorder-after <q8>
6485    <kk>      <Q>;<SPECIAL>;<SMALL>;IGNORE
6486    reorder-after <t8>
6487    <TH>      "<T><H>";"<TH><TH>";"<CAPITAL><CAPITAL>";IGNORE
6488    <th>      "<T><H>";"<TH><TH>";"<SMALL><SMALL>";IGNORE
6489    reorder-after <y8>
6490    % <U:> and <U"> are treated as <Y> in Danish
6491    <U:>      <Y>;<U:>;<CAPITAL>;IGNORE
6492    <u:>      <Y>;<U:>;<SMALL>;IGNORE
6493    <U">      <Y>;<U">;<CAPITAL>;IGNORE
6494    <u">      <Y>;<U">;<SMALL>;IGNORE
6495    reorder-after <z8>
```

98

```
6496    % <AE> is a separate letter in Danish
6497    <AE>     <AE>;<NONE>;<CAPITAL>;IGNORE
6498    <ae>     <AE>;<NONE>;<SMALL>;IGNORE
6499    <AE'>    <AE>;<ACUTE>;<CAPITAL>;IGNORE
6500    <ae'>    <AE>;<ACUTE>;<SMALL>;IGNORE
6501    <A3>     <AE>;<MACRON>;<CAPITAL>;IGNORE
6502    <a3>     <AE>;<MACRON>;<SMALL>;IGNORE
6503    <A:>     <AE>;<SPECIAL>;<CAPITAL>;IGNORE
6504    <a:>     <AE>;<SPECIAL>;<SMALL>;IGNORE
6505    % <O//> is a separate letter in Danish
6506    <O//>    <O//>;<NONE>;<CAPITAL>;IGNORE
6507    <o//>    <O//>;<NONE>;<SMALL>;IGNORE
6508    <O//'>   <O//>;<ACUTE>;<CAPITAL>;IGNORE
6509    <o//'>   <O//>;<ACUTE>;<SMALL>;IGNORE
6510    <O:>     <O//>;<DIAERESIS>;<CAPITAL>;IGNORE
6511    <o:>     <O//>;<DIAERESIS>;<SMALL>;IGNORE
6512    <O">     <O//>;<DOUBLE-ACUTE>;<CAPITAL>;IGNORE
6513    <o">     <O//>;<DOUBLE-ACUTE>;<SMALL>;IGNORE
6514    % <AA> is a separate letter in Danish
6515    <AA>     <AA>;<NONE>;<CAPITAL>;IGNORE
6516    <aa>     <AA>;<NONE>;<SMALL>;IGNORE
6517    <A-A>    <AA>;<A-A>;<CAPITAL>;IGNORE
6518    <A-a>    <AA>;<A-A>;<CAPITAL-SMALL>;IGNORE
6519    <a-A>    <AA>;<A-A>;<SMALL-CAPITAL>;IGNORE
6520    <a-a>    <AA>;<A-A>;<SMALL>;IGNORE
6521    <AA'>    <AA>;<AA'>;<CAPITAL>;IGNORE
6522    <aa'>    <AA>;<AA'>;<SMALL>;IGNORE
6523    reorder-end
6524    END LC_COLLATE
6525
6526    LC_MONETARY
6527    int_curr_symbol          "<D><K><K><SP>"
6528    currency_symbol          "<k><r>"
6529    mon_decimal_point        "<,>"
6530    mon_thousands_sep        "<.>"
6531    mon_grouping             3;3
6532    positive_sign            ""
6533    negative_sign            "<->"
6534    int_frac_digits          2
6535    frac_digits              2
6536    p_cs_precedes            1
6537    p_sep_by_space           2
6538    n_cs_precedes            1
6539    n_sep_by_space           2
6540    p_sign_posn              4
6541    n_sign_posn              4
6542    END LC_MONETARY
6543
6544    LC_NUMERIC
6545    decimal_point            "<,>"
6546    thousands_sep            "<.>"
6547    grouping                 3;3
6548    END LC_NUMERIC
6549
6550    LC_TIME
6551    abday       "<m><a><n>";/
6552                "<t><i><r>";"<o><n><s>";/
6553                "<t><o><r>";"<f><r><e>";/
6554                "<l><o//><r>";"<s><o/><n>
6555    day         "<m><a><n><d><a><g>";/
6556                "<t><i><r><s><d><a><g>";/
6557                "<o><n><s><d><a><g>";/
6558                "<t><o><r><s><d><a><g>";/
6559                "<f><r><e><d><a><g>";/
6560                "<l><o//><r><d><a><g>"/
6561                "<s><o//><n><d><a><g>";
6562    week     7;19971201;4
6563    abmon       "<j><a><n>";"<f><e><b>";/
6564                "<m><a><r>";"<a><p><r>";/
6565                "<m><a><j>";"<j><u><n>";/
```

```
6566                    "<j><u><l>";"<a><u><g>";/
6567                    "<s><e><p>";"<o><k><t>";/
6568                    "<n><o><v>";"<d><e><c>"
6569      mon           "<j><a><n><u><a><r>";/
6570                    "<f><e><b><r><u><a><r>";/
6571                    "<m><a><r><t><s>";/
6572                    "<a><p><r><i><l>";/
6573                    "<m><a><j>";/
6574                    "<j><u><n><i>";/
6575                    "<j><u><l><i>";/
6576                    "<a><u><g><u><s><t>";/
6577                    "<s><e><p><t><e><m><b><e><r>";/
6578                    "<o><k><t><o><b><e><r>";/
6579                    "<n><o><v><e><m><b><e><r>";/
6580                    "<d><e><c><e><m><b><e><r>"
6581      d_t_fmt       "<%><a><SP><%><F><SP><%><T><SP><%><Z>"
6582      d_fmt         "<%><O><d><.><SP><%><B><SP><%><Y>"
6583      atl_digits    "<0><.>;<1><.>;<2><.>;<3><.>;<4><.>;/
6584                    <5><.>;<6><.>;<7><.>;<8><.>;<9><.>;/
6585                    <1><0><.>;<1><1><.>;<1><2><.>;<1><3><.>;<1><4><.>;/
6586                    <1><5><.>;<1><6><.>;<1><7><.>;<1><8><.>;<1><9><.>;/
6587                    <2><0><.>;<2><1><.>;<2><2><.>;<2><3><.>;<2><4><.>;/
6588                    <2><5><.>;<2><6><.>;<2><7><.>;<2><8><.>;<2><9><.>;/
6589                    <3><0><.>;<3><1><.>"
6590      t_fmt         "<%><T>"
6591      am_pm         "";""
6592      t_fmt_ampm    ""
6593      timezone      "<C><E><T><-><1><C><E><T><SP><D><S><T><,><M><3><.><5><.><0>/
6594                    <,><M><1><0><.><5><.><0>"
6595      END LC_TIME
6596
6597      LC_MESSAGES
6598      yesexpr       "<<(><1><J><j><Y><y><)/>><.><*>"
6599      noexpr        "<<(><0><N><n><)/>><.><*>"
6600      END LC_MESSAGES
6601
6602      LC_PAPER
6603      copy "i18n"
6604      END LC_PAPER
6605
6606      LC_NAME
6607      name_fmt    "<%><p><%><t><%><g><%><t><%><m><%><t><%><f>"
6608      name_gen    ""
6609      name_mr     "<h><r>"
6610      name_mrs    "<f><r><u>"
6611      name_miss   "<f><r><o/><k><e><n>"
6612      name_ms     "<f><r>"
6613      END LC_NAME
6614
6615      LC_ADDRESS
6616      country_name      "<D><a><n><m><a><r><k>"
6617      country_post      "<D><K>"
6618      country_ab2       "<D><K>"
6619      country_ab3       "<D><N><K>"
6620      country_num       208
6621      country_car       "<D><K>"
6622      country_isbn      "<8><7>"
6623      lang_ab           "<d><a>"
6624      lang_term         "<d><a><n>"
6625      postal_fmt   "<%><a><%><N><%><f><%><N><%><d><%><N><%><b><%><N><%>/
6626                   <%><s><SP><%><h><SP><%><e><SP><%><r><%><N>/
```

6627   <%><C><-><%><z><SP><%><T><%><N><%><c><%><N>"
6628 END LC_ADDRESS
6629
6630 LC_TELEPHONE
6631 tel_int_fmt   "<+><%><c><SP><%><a><SP><%><l>"
6632 tel_dom_fmt   "<%><l>"
6633 int_select   "<0><0>"
6634 int_prefix   "<4><5>"
6635 END LC_TELEPHONE
6636
6637 **B.1.4   LC_MONETARY Rationale.**
6638
6639 The currency symbol does not appear in LC_MONETARY because it is not defined in the
6640 C Standard's C locale. The C Standard limits the size of decimal points and thousands
6641 delimiters to single-byte values. In FDCC-sets based on multibyte coded character sets this
6642 cannot be enforced, obviously; this Technical Report does not prohibit such characters, but
6643 makes the behaviour unspecified (in the text "In contexts where other standards . . . ").
6644
6645 The grouping specification is based on, but not identical to, the C Standard . The "-1"
6646 signals that no further grouping shall be performed, the equivalent of (CHAR_MAX) in
6647 the C Standard).
6648
6649 The FDCC-set definition is an extension of the C Standard localeconv() specification. In
6650 particular, rules on how currency_symbol is treated are extended to also cover int_-
6651 curr_symbol, and p_set_by_space and n_sep_by_space have been augmented with the
6652 value 2, which places a space between the sign and the symbol (if they are adjacent;
6653 otherwise it should be treated as a 0). The following table shows the result of various
6654 combinations:
6655
6656

		p_sep_by_space		
		2	1	0
$p\_cs\_precedes = 1$	$p\_sign\_posn = 0$	($ 1.25)	($ 1.25)	($1.25)
	$p\_sign\_posn = 1$	+ $1.25	+$ 1.25	+$1.25
	$p\_sign\_posn = 2$	$1.25 +	$ 1.25+	$1.25+
	$p\_sign\_posn = 3$	+ $1.25	+$ 1.25	+$1.25
	$p\_sign\_posn = 4$	$ +1.25	$+ 1.25	$+1.25
$p\_cs\_precedes = 0$	$p\_sign\_posn = 0$	(1.25 $)	(1.25 $)	(1.25$)
	$p\_sign\_posn = 1$	+1.25 $	+1.25 $	+1.25$
	$p\_sign\_posn = 2$	1.25$ +	1.25 $+	1.25$+
	$p\_sign\_posn = 3$	1.25+ $	1.25 +$	1.25+$
	$p\_sign\_posn = 4$	1.25$ +	1.25 $+	1.25$+

6657 (line numbers)
6658
6659
6660
6661
6662
6663
6664
6665
6666
6667
6668
6669
6670
6671
6672
6673 The following is an example of the interpretation of the mon_grouping keyword.
6674 Assuming that the value to be formatted is 123456789 and the mon_thousands_sep is "'",
6675 then the following table shows the result. The third column shows the equivalent C
6676 Standard string that would be used to accommodate this grouping. It is the responsibility

6677   of the utility to perform mappings of the formats in this clause to those used by language
6678   bindings such as the C Standard .
6679
6680

	Mon_grouping	Formatted Value	C String
6681 | | 3;-1 | 123456'789 | "\3\177" |
6682 | | 3 | 123'456'789 | "\3" |
6683 | | 3;2;-1 | 1234'56'789 | "\3\2\177" |
6684 | | 3;2 | 12'34'56'789 | "\3\2" |
6685 | | -1 | 123456789 | "177" |

6687
6688   In these examples, the octal value of (CHAR_MAX) is 177.
6689
6690   The multiple currency support is specified such that a FDCC-set can be used without
6691   change during the transition period in a static environment. For example in the case of the
6692   Euro currency as being employed in a number of European countries, there is no need to
6693   change the FDCC-set when shifting from one currency to two concurrent currencies; and
6694   there is no need to change FDCC-set, when changing to the Euro as the only currency.
6695   Also the same application call can be made to be valid for countries with a single
6696   currency and countries with dual currencies. The specifications can also be used without
6697   change of the FDCC-set on an installation, when converting from one national currency to
6698   another, for example when removing some zeroes to form a new currency.
6699
6700   The following example illustrates the support for multiple currencies; the example is for
6701   the Euro in Germany:
6702

```
6703   LC_MONETARY
6704   valid_from              "";                "19990101"
6705   valid_to                "20020630";        ""
6706   conversion_rate         1;                 195/100
6707   int_curr_symbol         "<D><E><M><SP>"; "<E><U><R><SP>"
6708   currency_symbol         "<D><M>";          "<E><U><R>"
6709   mon_decimal_point       "<,>"
6710   mon_thousands_sep       "<.>"
6711   mon_grouping            3;3
6712   positive_sign           ""
6713   negative_sign           "<->"
6714   int_frac_digits         2;                 2
6715   frac_digits             2;                 2
6716   p_cs_precedes           1;                 1
6717   p_sep_by_space          2;                 2
6718   n_cs_precedes           1;                 1
6719   n_sep_by_space          2;                 2
6720   p_sign_posn             4;                 4
6721   n_sign_posn             4;                 4
6722
6723   END LC_MONETARY
```

6724

B.1.5   LC_NUMERIC Rationale.

6726

6727   See the rationale for LC_MONETARY (B1.3) for a description of the behaviour of
6728   grouping.
6729

B.1.6   LC_TIME Rationale.

6731

6732   The LC_TIME descriptions of abday, day, and abmon imply a Gregorian style calendar
6733   (7-day weeks, 12-month years, leap years, etc.). Other calendars can be supported, for
6734   example calendars with a fixed week length.

6735   In some FDCC-sets the field descriptors for weekday and month names will be given with
6736   an initial small letter. Programs using these fields may need to adjust the capitalization if
6737   the output is going to be used at the beginning of a sentence.
6738
6739   The field descriptors corresponding to the optional keywords consist of a modifier
6740   followed by a traditional field descriptor (for instance %Ex). If the optional keywords are
6741   not supported by the application or are unspecified for the current FDCC-set, these field
6742   descriptors shall be treated as the traditional field descriptor. For instance, assume the
6743   following keywords:
6744
6745   alt_digits "0th";"1st";"2nd";"3rd";"4th";"5th";"6th";"7th";"8th";"9th";"10th"
6746   d_fmt "The %Od day of %B in %Y"
6747
6748   On 7/4/1776, the %x field descriptor would result in "The 4th day of July in 1776," while
6749   7/14/1789 would come out as "The 14 day of July in 1789." It can be noted that the above
6750   example is for illustrative purposes only; the %o modifier is primarily intended to provide
6751   for Kanji or Hindi digits in date formats. While it is clear that an alternate year format is
6752   required, there is no consensus on the format or the requirements. As a result, while these
6753   keywords are reserved, the details are left unspecified. It is expected that National
6754   Standards Bodies will provide specifications.
6755
6756
6757   **B.1.7   LC_MESSAGES Rationale.**
6758
6759   The LC_MESSAGES category is described in clause 4 as affecting the language used by
6760   utilities for their output. The mechanism used by the application to accomplish this, other
6761   than the responses shown here in the FDCC-set definition, is not specified by this version
6762   of this Technical Report. The internationalization working group is developing an interface
6763   that would allow applications (and, presumably some of the standard utilities) to access
6764   messages from various message catalogs, tailored to a user's LC_MESSAGES value.
6765
6766
6767   **B.1.8   LC_PAPER Rationale.**
6768
6769   The LC_PAPER category gives information to prepare output on a printer. Only the
6770   physical measurements of the height and width is available, as this is the information most
6771   often available in various document handling applications.
6772
6773
6774   **B.1.9   LC_NAME Rationale.**
6775
6776   The LC_NAME category gives information to prepare a text for addressing a person, for
6777   example as a part of a postal address on an envelope, or as a salutating line in a letter.
6778   The information is intended to be given to an API that has the various naming information
6779   as parameters and yields a formatted string as the return value.
6780
6781
6782   **B.1.10   LC_ADDRESS Rationale.**
6783
6784   The LC_ADDRESS category gives information to prepare a text for writing an address,

6785   for example as a part of a postal address on an envelope. The information is intended to
6786   be given to an API that has the various address information as parameters and yields a
6787   formatted string as the return value.
6788
6789
6790   **B.1.11   LC_TELEPHONE Rationale.**
6791
6792   The LC_TELEPHONE category gives information to prepare a text for writing a telephone
6793   number. The information is intended to be given to an API that has the various
6794   information on a telephone number as parameters and yields a formatted string as the
6795   return value. Both an international and a domestic formatting possibility is available.
6796
6797
6798   **B.2   Character Set Rationale.**
6799
6800   This Technical Report poses no requirement that multiple character sets or code sets be
6801   supported, leaving this as a marketing differentiation for implementors. Although multiple
6802   charmaps are supported, it is the responsibility of the application to provide the file(s); if
6803   only one is provided, only that one will be accessible.
6804
6805   The character set description text provides the capability to describe character set attributes
6806   (such as collation order or character classes) independent of character set encoding, and
6807   using only the characters in the portable character set. This makes it possible to create
6808   "generic" FDCC-set source texts for all code sets that share the portable character set
6809   (such as the ISO/IEC 8859 family or IBM Extended ASCII).
6810
6811   Applications are free to describe more than one code set in a character set description text.
6812   For example, if an application defines ISO/IEC 8859-1 as the primary code set, and
6813   ISO/IEC 8859-2 as an alternate set, with each character from the alternate code set
6814   preceded in data by a shift code, a character set description text could contain a complete
6815   description of the primary set and those characters from the secondary that are not
6816   identical, the encoding of the latter including the shift code.
6817
6818   Applications are free to choose their own symbolic names, as long as the names identified
6819   by this Technical Report are also defined; this provides support for already existing
6820   "character names".
6821
6822   The charmap was introduced to resolve problems with the portability of, especially,
6823   FDCC-set sources. While the portable character set (in Table 1) is a constant across all
6824   FDCC-sets for a particular application, this is not true for the extended character set.
6825   However, the particular coded character set used for an application does not necessarily
6826   imply different characteristics or collation: on the contrary, these attributes should in many
6827   cases be identical, regardless of codeset. The charmap provides the capability to define a
6828   common FDCC-set definition for multiple codesets (the same FDCC-set source can be
6829   used for codesets with different extended characters; the ability in the charmap to define
6830   ''empty'' names allows for characters missing in certain codesets).
6831
6832   In addition, some implementors have expressed an interest in using the charmap to define
6833   certain other characteristics of codesets, such as the <mb_cur_max> value for the
6834   particular codeset. (Note that <mb_cur_max> has to be equal to or lower than the C

6835   Standard {MB_LEN_MAX}, which is the application limit). Such extensions are not
6836   described here; but may be added in a later revision of this Technical Report.
6837

6838   The <escape_char> declaration was added at the request of the international community to
6839   ease the creation of portable charmaps on terminals not implementing the default
6840   backslash escape. (This approach was adopted because this is a new interface invented by
6841   ISO/IEC 9945-2:1993 POSIX. Historical interfaces, such as the shell command language
6842   and awk, have not been modified to accommodate this type of terminal.)
6843

6844   The octal number notation was selected to match those of POSIX "awk" and "tr" utilities
6845   and is consistent with that used by the POSIX localedef utility.
6846

6847   The charmap capability implements a facility available at some X/Open compatible
6848   applications. Its prime virtue is to support "generic" collation sequence source definitions.
6849   An implementor or an applications developer can produce a template definition that can be
6850   used to produce several codeset-dependent "compiled" FDCC-set definitions. The facility
6851   also removes any dependency in many source definitions on characters outside the
6852   character set defined in this clause.
6853

6854   The charmap allows specification of more than one encoding of a character. This allows
6855   for encodings that can encode items in more than one way. For example, an item can be
6856   encoded once as a fully composed character and again as a base character plus combining
6857   character. This would allow either representation to be recognized. As only the first
6858   occurrence of the character may be output, this technique could be used to normalize a
6859   character stream.
6860

6861   The ISO 2022 support introduced gives the possibility to refer other definitions via
6862   charmaps, so the full encoding does not have to be replicated. It supports shifting with G0,
6863   G1, G2 and G3 sets, and also general shifting of coded character sets via escape
6864   sequences.
6865

6866

6867   **B.3   Repertoiremap Rationale.**
6868

6869   The repertoiremap was introduced to make FDCC-sets independent of the availability of
6870   charmaps. With the repertoiremap it is possible to use a FDCC-set encoded with one set of
6871   symbolic character names, together with charmaps with other symbolic character naming
6872   schemes, provided there are repertoiremaps available for both naming schemes.
6873

6874   Repertoiremaps are also useful to describe repertoires of characters, to be used for
6875   example for transliteration.

(informative)

BNF Grammar

C.1   BNF Syntax Rules

The syntax used here is near to ISO/IEC 14977, but "_" is allowed in identifiers, and
comma is not used as concatenator, as the items are just concatenated.

Definitions between <angle brackets> make use of terms not defined in this BNF syntax,
and assume general English usage.

Other conventions:

> * means 0 or more repetitions of a token.
>
> \+ means one or more repetitions of a token
>
> Brackets [ ] indicate optional occurrence of a token.
>
> Comments start with a % on a separate line.

There may be more specifications in the normative text that describes restrictions on the
grammar.

C.2 Grammar for FDCC-sets

```
% The following is the overall FDCC-set grammar
FDCC_set_definition         = [ global_statement* ] category+ ;
global_statement            = 'escape_char' SP char_symbol EOL
                            | 'comment_char' SP char_symbol EOL
                            | 'repertoiremap' SP quoted_string EOL
                            | 'charmap' SP quoted_string EOL ;
category                    = lc_identification | lc_ctype | lc_collate
                            | lc_monetary | lc_numeric | lc_time
                            | lc_messages | lc_paper | lc_telephone
                            | lc_name | lc_address ;

% The following is the LC_IDENTIFICATION category grammar
lc_ident                    = ident_head ident_keyword* ident_tail
                            | ident_head copy_FDCC_set ident_tail ;
ident_head                  = 'LC_IDENTIFICATION' EOL ;
ident_keyword               = ident_keyword_string SP quoted_string EOL ;
ident_keyword_string        = 'title' | 'source' | 'address' | 'contact'
                            | 'email' | 'tel' | 'fax'| 'language'
                            | 'territory' | 'audience' | 'application'
                            | 'abbreviation' | 'revision' | 'date' ;
ident_tail                  = 'END' SP 'LC_IDENTIFICATION' EOL ;


% The following is the LC_CTYPE category grammar
lc_ctype                    = ctype_head ctype_keyword* [ translit ]
                              ctype_tail
                            | ctype_head copy_FDCC_set ctype_tail ;
ctype_head                  = 'LC_CTYPE' EOL ;
ctype_keyword               = charclass_keyword SP charclass_list EOL
                            | charconv_keyword SP charconv_list EOL ;
charclass_keyword           = 'upper' | 'lower' | 'alpha' | 'digit'
                            | 'punct' | 'xdigit' | 'space' | 'print'
                            | 'graph' | 'blank' | 'cntrl' | 'outdigit'
                            | 'class' class_name semicolon ;
class_name                  = '"combining"' | '"combining_level3"'
                            | '"' identifier '"' ;
```

```
6937   charclass_list                = charclass_list semicolon char_symbol
6938                                  | charclass_list semicolon ctype_abs_ellipsis
6939                                  semicolon char_symbol
6940                                  | charclass_list semicolon charsymbol
6941                                  ctype_symbolic_ellipses charsymbol
6942                                  | char_symbol ;
6943   charconv_keyword              = 'toupper' | 'tolower'
6944                                  | 'map' '"' identifier '"' semicolon ;
6945   charconv_list                 = charconv_list semicolon charconv_entry
6946                                  | charconv_entry ;
6947   charconv_entry                = '(' char_symbol comma char_symbol ')' ;
6948   ctype_symbolic_ellipses       = '..' | '....' | '..(2)..' ;
6949   ctype_abs_ellipses            = '...' ;
6950   translit                      = translit_start [translit_include]
6951                                  [default_missing] translit_statement*
6952                                  translit_end ;
6953   translit_start                = 'translit_start' EOL ;
6954   translit_include              = 'include' SP FDCC_set_name semicolon
6955                                  quoted_nonempty_string EOL ;
6956   default_missing               = 'default_missing' SP quoted_string EOL ;
6957   translit_ignore               = 'translit_ignore' SP charclass_list EOL ;
6958   translit_statement            = char_or_string SP char_or_string [ semicolon
6959                                  char_or_string ]*  EOL ;
6960   translit_end                  = 'translit_end' EOL ;
6961   ctype_tail                    = 'END' SP 'LC_TYPE' EOL ;
6962
6963   % The following is the LC_COLLATE category grammar
6964   lc_collate                    = collate_head collate_keywords collate_tail ;
6965   collate_head                  = 'LC_COLLATE' EOL ;
6966   collate_keywords              = [ opt_statement* ] order_statements ;
6967   opt_statement                 = 'collating-symbol' SP collsymbol* EOL
6968                                  | 'collating-element' SP collelement SP
6969                                  collelem_string EOL
6970                                  | 'section-symbol' spece+ sectionsymbol EOL
6971                                  | 'copy' SP FDCC_set_name EOL
6972                                  | 'col_weight_max' SP number EOL
6973                                  | 'symbol-equivalence' SP collsymbol SP
6974                                  collsymbol ;
6975   collelem_string               = '"' char_symbol char_symbol+ '"';
6976   order_statements              = order_start collation_order order_end ;
6977   order_start                   = 'order_start' SP sectionsymbol [ semicolon
6978                                  order_opts ] EOL
6979                                  | 'order_start' SP [ order_opts ] EOL ;
6980   order_opts                    = order_opt [ semicolon order_opt ] ;
6981   order_opt                     = order_opt [ comma opt_word ] ;
6982   opt_word                      = 'forward' | 'backward' | 'position' ;
6983   collation_order               = collation_statement* ;
6984   collation_statement           = collsymbol EOL
6985                                  | collating_element [ SP weight_list ] EOL ;
6986   collation_element             = char_symbol | collelement
6987                                  | ellipses | 'UNDEFINED' ;
6988   weight_list                   = weight_symbol [ semicolon weight_symbol ]* ;
6989   weight_symbol                 = <empty>
6990                                  | char_symbol
6991                                  | collsymbol
6992                                  | '"' elem_list '"'
6993                                  | '"' symb_list '"' | 'IGNORE' ;
6994   ellipses                      = '...'  | '..' | '....' ;
6995   reorder_after                 = 'reorder-after' SP collsymbol EOL ;
6996   reorder_end                   = 'reorder-end' EOL ;
6997   reorder_section_after         = 'reorder-section-after' SP sectionsymbol SP
6998                                  sectionsymbol EOL;
6999   reorder_section_end           = 'reorder-section-end' EOL ;
7000   order_end                     = 'order_end' EOL ;
7001   collate_tail                  = 'END' SP 'LC_COLLATE' EOL ;
7002
7003   % The following is the LC_MESSAGES category grammar
7004   lc_messages                   = messages_head messages_keyword* messages_tail
7005
7006                                  | messages_head copy_FDCC_set messages_tail ;
```

```
7007  messages_head              = 'LC_MESSAGES' EOL ;
7008  messages_keyword           = 'yesexpr' SP '"' extended_reg_expr '"' EOL
7009                             | 'yesexpr' SP '"' extended_reg_expr '"' EOL ;
7010  messages_tail              = 'END' SP 'LC_MESSAGES' EOL ;
7011
7012  % The following is the LC_MONETARY category grammar
7013  lc_monetary                = monetary_head monetary_keyword* monetary_tail
7014                             | monetary_head copy_FDCC_set monetary_tail ;
7015  monetary_head              = 'LC_MONETARY' EOL ;
7016  monetary_keyword           = mon_keyword_string SP quoted_string EOL
7017                             | mon_keyword_strings SP mon_string_list EOL
7018                             | mon_keyword_char SP mon_number_list EOL
7019                             | mon_keyword_date SP mon_date_list EOL
7020                             | 'conversion_rate' SP mon_conv_list EOL
7021                             | 'mon_grouping' SP mon_group_list EOL ;
7022  mon_keyword_string         = 'mon_decimal_point' | 'mon_thousands_sep'
7023                             | 'positive_sign' | 'negative_sign' ;
7024  mon_keyword_strings        = 'int_curr_symbol' | 'currency_symbol' ;
7025  mon_keyword_char           = 'int_frac_digits' | 'frac_digits'
7026                             | 'p_cs_precedes' | 'p_sep_by_space'
7027                             | 'n_cs_precedes' | 'n_sep_by_space'
7028                             | 'int_p_cs_precedes' | 'int_p_sep_by_space'
7029                             | 'int_n_cs_precedes' | 'int_n_sep_by_space'
7030                             | 'p_sign_posn' | 'n_sign_posn'
7031                             | 'int_p_sign_posn' | 'int_n_sign_posn' ;
7032  mon_keyword_date           = 'valid_from' | 'valid_to' ;
7033  mon_date_list              = mon_date | mon_date_list semicolon mon_date ;
7034  mon_date                   = '"' 8 * digit '"' ;
7035  mon_group_list             = number  | mon_group_list semicolon number ;
7036  mon_string_list            = quoted_string [ semicolon quoted_string]* ;
7037  mon_number_list            = mon_number | mon_number_list semicolon
7038                             mon_number ;
7039  mon_number                 = number | -1 ;
7040  mon_conv_list              = mon_pair | mon_conv_list semicolon mon_pair ;
7041  mon_pair                   = number spaces* '/' spcaes* number ;
7042  monetary_tail              = 'END' SP 'LC_MONETARY' EOL ;
7043
7044  % The following is the LC_NUMERIC category grammar
7045  lc_numeric                 = numeric_head numeric_keyword* numeric_tail
7046                             | numeric_head copy_FDCC_set numeric_tail ;
7047  numeric_head               = 'LC_NUMERIC' EOL ;
7048  numeric_keyword            = num_keyword_string SP quoted_string EOL
7049                             | num_keyword_grouping SP num_group_list EOL ;
7050  num_keyword_string         = 'decimal_point' | 'thousands_sep' ;
7051  num_keyword_grouping       = 'grouping' ;
7052  num_group_list             = number
7053                             | num_group_list semicolon number ;
7054  numeric_tail               = 'END' SP 'LC_NUMERIC' EOL ;
7055
7056  % The following is the LC_TIME category grammar
7057  lc_time                    = time_head time_keyword* time_tail
7058                             | time_head copy_FDCC_set time_tail ;
7059  time_head                  = 'LC_TIME' EOL ;
7060  time_keyword               = time_keyword_name SP time_list EOL
7061                             | time_keyword_fmt SP quoted_string EOL
7062                             | time_keyword_opt SP time_list EOL
7063                             | 'week' SP number semicolon mon_date semicolon
7064                             number EOL
7065                             | time_keyword_num SP number EOL
7066                             | 'timezone' SP time_list EOL;
7067  time_keyword_name          = 'abday' | 'day' | 'abmon' | 'mon' | 'am_pm' ;
7068  time_keyword_fmt           = 'd_t_fmt' | 'd_fmt' |'t_fmt'   | 't_fmt_ampm';
7069  time_keyword_opt           = 'era' |'era_year'| 'era_d_fmt'| 'alt_digits'
7070                             ;
7071  time_keyword_week          = 'week' ;
7072  time_keyword_num           = 'first_weekday' | 'first_workday'
7073                             | 'cal_direction' ;
7074  time_list                  = time_list semicolon quoted_string
7075                             | quoted_string ;
7076  time_tail                  = 'END' SP 'LC_TIME' EOL ;
7077
```

```
7078      % The following is the LC_PAPER category grammar
7079      lc_paper                    = paper_head paper_keyword* paper_tail
7080                                  | paper_head copy_FDCC_set paper_tail ;
7081      paper_head                  = 'LC_PAPER' EOL ;
7082      paper_keyword               = paper_keyword_num SP number EOL ;
7083      paper_keyword_num           = 'height' | 'width' ;
7084      paper_tail                  = 'END' SP 'LC_PAPER' EOL ;
7085
7086      % The following is the LC_NAME category grammar
7087      lc_name                     = name_head name_keyword* name_tail
7088                                  | name_head copy_FDCC_set name_tail ;
7089      name_head                   = 'LC_NAME' EOL ;
7090      name_keyword                = name_keyword_string SP quoted_string EOL ;
7091      name_keyword_string         = 'name_fmt' | 'name_gen' | 'name_mr'
7092                                  | 'name_mrs' | 'name_ms' | 'name_miss'
7093                                  | 'name_ms' ;
7094      name_tail                   = 'END' SP 'LC_NAME' EOL ;
7095
7096      % The following is the LC_ADDRESS category grammar
7097      lc_address                  = address_head address_keyword* address_tail
7098                                  | address_head copy_FDCC_set address_tail ;
7099      address_head                = 'LC_ADDRESS' EOL ;
7100      address_keyword             = address_keyword_string SP quoted_string EOL
7101                                  | address_keyword_num SP number EOL ;
7102      address_keyword_string      = 'postal_fmt' | 'country_name' |
7103                                    'country_post' | 'country_ab2' | 'country_ab3'
7104                                  | 'country_car' | 'country_isbn'| 'lang_name' |
7105                                    'lang_ab' | 'lang_term' | 'lang_lib' ;
7106      address_keyword_num         = "country_num" ;
7107      address_tail                = 'END' SP 'LC_ADDRESS' EOL ;
7108
7109      % The following is the LC_TELEPHONE category grammar
7110      lc_tel                      = tel_head tel_keyword* tel_tail
7111                                  | tel_head copy_FDCC_set tel_tail ;
7112      tel_head                    = 'LC_TELEPHONE' EOL ;
7113      tel_keyword                 = tel_keyword_string SP quoted_string EOL ;
7114      tel_keyword_string          = 'tel_int_fmt' | 'tel_dom_fmt' | 'int_select'
7115                                  | 'int_prefix' ;
7116      tel_tail                    = 'END' SP 'LC_TELEPHONE' EOL ;
7117
7118      % The following grammar rules are common to all categories
7119      char                        = <any character except those that makes an End
7120                                    Of Line>
7121      graphic_char                = <any char except control_chars and space> ;
7122      space                       = ' ' | <TAB> ;
7123      SP                          = space+ ;
7124      EOL                         = <anything that makes an End Of Line (EOL) in
7125                                    the operating system employed> | comment EOL ;
7126      comment_char                = <defined by the 'comment_char' keyword> ;
7127      escape_char                 = <defined by the 'escape_char' keyword> ;
7128      charsymbol                  = simple_symbol | ucs_symbol ;
7129      collsymbol                  = simple_symbol ;
7130      collelement                 = simple_symbol ;
7131      sectionsymbol               = simple_symbol ;
7132      octdigit                    = '0'|'1'|'2'|'3'|'4'|'5'|'6'|'7' ;
7133      digit                       = '0'|'1'|'2'|'3'|'4'|'5'|'6'|'7'|'8'|'9' ;
7134      hex_upper                   = 'A'|'B'|'C'|'D'|'E'|'F'| digit ;
7135      hexdigit                    = hex_upper | 'a'|'b'|'c'|'d'|'e'|'f' ;
7136      letter                      = 'a'|'b'|'c'|'d'|'e'|'f'|'g'|'h'|'i'|'j'|'k'
7137                                    |'l'|'m'|'n'|'o'|'p'|'q'|'r'|'s'
7138                                    't'|'u'|'v'|'w'|'x'|'y'|'z'|'A'|'B'|'C'|'D'
7139                                    'E'|'F'|'G'|'H'|'I'|'J'|'K'|'L'|'M'|'N'|'O'
7140                                    'P'|'Q'|'R'|'S'|'T'|'U'|'V'|'W'|'X'|'Y'|'Z' ;
7141      portable_graph_gtr          = letter | digit | '!'|'"'|'#'|'$'|'%'|'&'
7142                                    | "'"|'('|')'|'*'|'+'|','|'-'|'.'|'/'|':'|';'
7143                                    | '<'|'='|'?'|'@'|'['|'\'|']'|'^'|'_'
7144                                    | '`'|'{'|'|'|'}'|'~' ;
7145      portable_graph              = portable_graph_gtr | '>' ;
7146      portable_char               = portable_graph | ' ' | <NUL> | <ALERT>
7147                                    | <BACKSPACE> | <TAB> | <CARRIAGE_RETURN>
```

```
7148                                      | <NEWLINE> | <VERTICAL_TAB> | <FORM_FEED> ;
7149    octal_char                = escape_char    octdigit octdigit octdigit* ;
7150    hex_char                  = escape_char 'x' hexdigit hexdigit hexdigit* ;
7151    decimal_char              = escape_char 'd' digit digit digit* ;
7152    number                    = digit+ ;
7153    id_part                   = letter | digit | '-' | '_' ;
7154    four_digit_hex_string     = hex_upper hex_upper hex_upper hex_upper ;
7155    identifier                = letter id_part* ;
7156    simple_symbol             = space* '<' portable_graph_gtr+ '>' ;
7157    ucs_symbol                = space* '<U' four_digit_hex_string
7158                                [ four_digit_hex_string ] '>' ;
7159    quoted_string             = '"' char_symbol* '"' ;
7160    quoted_nonempty_string    = '"' char_symbol+ '"' ;
7161    char_symbol               = char | charsymbol
7162                                | octal_char | hex_char | decimal_char ;
7163    elem_list                 = elem+ ;
7164    elem                      = char_symbol | collsymbol | collelement ;
7165    symb_list                 = collsymbol+ ;
7166    FDCC_set_name             = FDCC-name | '"' FDCC-name '"' ;
7167    copy_FDCC_set             = 'copy' FDCC_set_name EOL ;
7168    FDCC-name                 = portable_graph+ ;
7169    semicolon                 = space* ';' space* ;
7170    comma                     = space* ',' space* ;
7171    comment                   = comment_char char* ;
7172
```

Annex D

(informative)

Index

BIBLIOGRAPHY

The following specifications are considered relevant to this Technical Report, in addition to the normative references.

CEPT, CEPT-MAILCODE, *Country code for mail*.

ISO 646, *Information technology - ISO 7-bit coded character set for information interchange*.

ISO/IEC 9899, *Information technology - Programming language C*.

ISO/IEC 14977, *Information technology - Syntactic metalanguage - Extended BNF*.

The Unicode Consortium: *The Unicode Standard, Version 2.0*, Addison Wesley Developers Press, July 1996. ISBN 0-201-48345-9.

IBM: *National Language Design Guide Volume 2 - National Language Support Reference Manual*, IBM SE09-8002-03, August 1994.

STRÍ: *Nordic Cultural Requirements on Information Technology (Summary report)*, STRÍ TS3, Libris, Reykjavík, Iceland 1992. ISBN 9979-9004-3-1.