

Reference number of working document: **ISO/IEC JTC1/SC22/WG20 N???**

Date: 1999-06-11

Reference number of document: **ISO/IEC draft FPDTR3 14652**

Committee identification: ISO/IEC JTC1/SC22

Secretariat: ANSI

Information technology —

Specification method for cultural conventions

Technologies de l'information —

Méthode de modélisation des conventions culturelles

1

Document type: International standard

Document subtype: if applicable

Document stage: (40) Enquiry

Document language: E

		Page
2	Contents	
3		
4	1 SCOPE	1
5	2 NORMATIVE REFERENCES	1
6	3 TERMS, DEFINITIONS AND NOTATIONS	2
7	4 FDCC-set	6
8	4.1 FDCC-set definition	6
9	4.2 LC_IDENTIFICATION	10
10	4.3 LC_CTYPE	11
11	4.4 LC_COLLATE	27
12	4.5 LC_MONETARY	42
13	4.6 LC_NUMERIC	46
14	4.7 LC_TIME	47
15	4.8 LC_MESSAGES	53
16	4.9 LC_PAPER	53
17	4.10 LC_NAME	55
18	4.11 LC_ADDRESS	57
19	4.12 LC_TELEPHONE	57
20	5 CHARMAP	58
21	6 REPERTOIREMAP	62
22	7 CONFORMANCE	90
23	Annex A (informative) DIFFERENCES FROM POSIX	91
24	Annex B (informative) RATIONALE	93
25	Annex C (informative) BNF GRAMMAR	107
26	Annex D (informative) INDEX	112
27	BIBLIOGRAPHY	115
28		

FOREWORD

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

The main task of a technical committee is to prepare International Standards but in exceptional circumstances, the publication of a Technical Report of one of the following types may be proposed:

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

Technical Reports are drafted in accordance with the rules given in the ISO/IEC Directives, Part 3.

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

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

The Annexes A, B, C and D of this Technical Report are for information only.

69 Introduction

70

71 This Technical Report defines a general mechanism to specify cultural conventions, and it
72 defines formats for a number of specific cultural conventions in the areas of character
73 classification and conversion, sorting, number formatting, monetary formatting, date
74 formatting, message display, paper formats, addressing of persons, postal address
75 formatting, telephone number handling, and a way to specify how much is covered and the
76 status of it.

77

78 There are a number of benefits coming from this standard:

79

80 Rigid specification Using this Technical Report, a user can rigidly specify a
81 number of the cultural conventions that apply to the
82 information technology environment of the user.

83

84 Cultural adaptability If an application has been designed and built in a
85 cultural neutral manner, the application may use the
86 specifications as data to its APIs, and thus the same
87 application may accommodate different users in a
88 culturally acceptable way to each of the users, without
89 change of the binary application.

90

91 Productivity This standard specifies those cultural conventions and
92 how to specify data for them. With those data an
93 application developer is relieved from getting the
94 different information to support all the cultural
95 environments for the expected customers of the product.
96 The application developer is thus ensured of culturally
97 correct behavior as specified by the customer, and
98 possibly more markets may be reached as customers may
99 have the possibility to provide the data themselves for
100 markets that were not targeted.

101

102 Uniform behaviour When a number of applications share one cultural
103 specification, which may be supplied from the user or a
104 built-in nature, their behaviour for cultural adaptation
105 become uniform.

106

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

109

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

115

116 This Technical Report defines a format compatible with the one used in the International
117 String Ordering standard, ISO/IEC 14651. This Technical Report is backwards compatible
118 with the ISO/IEC 9945-2:1993 POSIX shell and utilities standard, particularly its clauses

119 2.4 and 2.5. The major extensions from that text are listed in annex A. This Technical
120 Report has enhanced functionality in a number of areas such as ISO/IEC 10646 support,
121 more classification of characters, transliteration, dual (multi) currency support, enhanced
122 date and time formatting, paper size identification, personal name writing, postal address
123 formatting, telephone number handling, and management of categories. There is enhanced
124 support for character sets including ISO/IEC 2022 handling and an enhanced method to
125 separate the specification of cultural conventions from an actual encoding via a description
126 of the character repertoire employed. A standard set of values for all the categories has
127 been defined covering the repertoire of ISO/IEC 10646-1.
128

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*.

179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228

3 TERMS, DEFINITIONS AND NOTATIONS

3.1 Terms and definitions

For the purposes of this Technical Report, the terms and definitions given in the following apply.

3.1.1

byte:

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

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

3.1.2

character:

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

3.1.3

coded character:

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

3.1.4

text file:

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

3.1.5

cultural convention:

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

3.1.6

FDCC-set:

A Set of Formal Definitions of Cultural Conventions. The definition of the subset of a user's information technology environment that depends on language and cultural conventions. Note: the FDCC-set is a superset of the "locale" term in C and POSIX.

3.1.7

charmap:

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

3.1.8

repertoiremap:

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

229 **3.1.9**
230 **character class:**
231 A named set of characters sharing an attribute associated with the name of the class.
232

233 **3.1.10**
234 **collation:**
235 The logical ordering of strings according to defined precedence rules.
236

237 **3.1.11**
238 **collating element:**
239 The smallest entity used to determine logical ordering.
240

241 See collating sequence. A collating element shall consist of either a single character, or
242 two or more characters collating as a single entity. The LC_COLLATE category in the
243 associated FDCC-set determines the set of collating elements.
244

245 **3.1.12**
246 **multicharacter collating element:**
247 A sequence of two or more characters that collate as an entity.
248

249 For example, in some languages two characters are sorted as one letter, as in the case for
250 Danish and Norwegian "aa".
251

252 **3.1.13**
253 **collating sequence:**
254 The relative order of collating elements as determined by the setting of the LC_COLLATE
255 category in the applied FDCC-set.
256

257 **3.1.14**
258 **equivalence class:**
259 A set of collating elements with the same primary collation weight.
260

261 Elements in an equivalence class are typically elements that naturally group together, such
262 as all accented letters based on the same letter.
263

264 The collation order of elements within an equivalence class is determined by the weights
265 assigned on any subsequent levels after the primary weight.
266

267 **3.2 Notations**

268
269 The following notations and common conventions for specifications apply to this standard:
270

271 **3.2.1 Notation for defining syntax**

272
273 In this standard, the description of an individual record in a FDCC-set is done using the
274 syntax notation given in the following.
275

276 The syntax notation looks as follows:
277

278 " <format> ", [<arg1> , <arg2> , ..., <argn>]
279

280 The <format> is given in a format string enclosed in double quotes, followed by a number
 281 of parameters, separated by commas. It is similar to the format specification defined in
 282 clause 2.12 in the ISO/IEC 9945-2:1993 standard and the format specification used in C
 283 language printf() function. The format of each parameter is given by an escape sequence
 284 as follows:

285
 286 %*s* specifies a string
 287 %*d* specifies a decimal integer
 288 %*c* specifies a character
 289 %*o* specifies an octal integer
 290 %*x* specifies a hexadecimal integer

291
 292 A " " (an empty character position) in the syntax string represent one or more <blank>
 293 characters.

294
 295 All other characters in the format string except

296
 297 %% specifies a single %
 298 \n specifies an end-of-line

299
 300 represent themselves.

301
 302 The notation "... " is used to specify that repetition of the previous specification is optional,
 303 and this is done in both the format string and in the parameter list.

304
 305

306 3.2.3 Portable character set

307
 308 A set of symbolic names for characters in Table 1, which is called the portable character
 309 set, is used in character description text of this specification. The first eight entries in
 310 Table 1 are defined in ISO/IEC 6429 and others are defined in ISO/IEC 10646-1.

311

312 **Table 1: Portable character set**

313 Symbolic name	314 Glyph	315 UCS	316 Description
317 <NUL>		<U0000>	NULL (NUL)
318 <alert>		<U0007>	BELL (BEL)
319 <backspace>		<U0008>	BACKSPACE (BS)
320 <tab>		<U0009>	CHARACTER TABULATION (HT)
321 <carriage-return>		<U000D>	CARRIAGE RETURN (CR)
322 <newline>		<U000A>	LINE FEED (LF)
323 <vertical-tab>		<U000B>	LINE TABULATION (VT)
324 <form-feed>		<U000C>	FORM FEED (FF)
325 <space>		<U0020>	SPACE
326 <exclamation-mark>	!	<U0021>	EXCLAMATION MARK
327 <quotation-mark>	"	<U0022>	QUOTATION MARK
328 <number-sign>	#	<U0023>	NUMBER SIGN
329 <dollar-sign>	\$	<U0024>	DOLLAR SIGN
330 <percent-sign>	%	<U0025>	PERCENT SIGN
331 <ampersand>	&	<U0026>	AMPERSAND
332 <apostrophe>	'	<U0027>	APOSTROPHE
333 <left-parenthesis>	(<U0028>	LEFT PARENTHESIS
334 <right-parenthesis>)	<U0029>	RIGHT PARENTHESIS
335 <asterisk>	*	<U002A>	ASTERISK
336 <plus-sign>	+	<U002B>	PLUS SIGN
337 <comma>	,	<U002C>	COMMA
338 <hyphen-minus>	-	<U002D>	HYPHEN-MINUS
339 <hyphen>	-	<U002D>	HYPHEN-MINUS
<full-stop>	.	<U002E>	FULL STOP

340	<period>	.	<U002E>	FULL STOP
341	<slash>	/	<U002F>	SOLIDUS
342	<solidus>	/	<U002F>	SOLIDUS
343	<zero>	0	<U0030>	DIGIT ZERO
344	<one>	1	<U0031>	DIGIT ONE
345	<two>	2	<U0032>	DIGIT TWO
346	<three>	3	<U0033>	DIGIT THREE
347	<four>	4	<U0034>	DIGIT FOUR
348	<five>	5	<U0035>	DIGIT FIVE
349	<six>	6	<U0036>	DIGIT SIX
350	<seven>	7	<U0037>	DIGIT SEVEN
351	<eight>	8	<U0038>	DIGIT EIGHT
352	<nine>	9	<U0039>	DIGIT NINE
353	<colon>	:	<U003A>	COLON
354	<semicolon>	;	<U003B>	SEMICOLON
355	<less-than-sign>	<	<U003C>	LESS-THAN SIGN
356	<equals-sign>	=	<U003D>	EQUALS SIGN
357	<greater-than-sign>	>	<U003E>	GREATER-THAN SIGN
358	<question-mark>	?	<U003F>	QUESTION MARK
359	<commercial-at>	@	<U0040>	COMMERCIAL AT
360	<A>	A	<U0041>	LATIN CAPITAL LETTER A
361		B	<U0042>	LATIN CAPITAL LETTER B
362	<C>	C	<U0043>	LATIN CAPITAL LETTER C
363	<D>	D	<U0044>	LATIN CAPITAL LETTER D
364	<E>	E	<U0045>	LATIN CAPITAL LETTER E
365	<F>	F	<U0046>	LATIN CAPITAL LETTER F
366	<G>	G	<U0047>	LATIN CAPITAL LETTER G
367	<H>	H	<U0048>	LATIN CAPITAL LETTER H
368	<I>	I	<U0049>	LATIN CAPITAL LETTER I
369	<J>	J	<U004A>	LATIN CAPITAL LETTER J
370	<K>	K	<U004B>	LATIN CAPITAL LETTER K
371	<L>	L	<U004C>	LATIN CAPITAL LETTER L
372	<M>	M	<U004D>	LATIN CAPITAL LETTER M
373	<N>	N	<U004E>	LATIN CAPITAL LETTER N
374	<O>	O	<U004F>	LATIN CAPITAL LETTER O
375	<P>	P	<U0050>	LATIN CAPITAL LETTER P
376	<Q>	Q	<U0051>	LATIN CAPITAL LETTER Q
377	<R>	R	<U0052>	LATIN CAPITAL LETTER R
378	<S>	S	<U0053>	LATIN CAPITAL LETTER S
379	<T>	T	<U0054>	LATIN CAPITAL LETTER T
380	<U>	U	<U0055>	LATIN CAPITAL LETTER U
381	<V>	V	<U0056>	LATIN CAPITAL LETTER V
382	<W>	W	<U0057>	LATIN CAPITAL LETTER W
383	<X>	X	<U0058>	LATIN CAPITAL LETTER X
384	<Y>	Y	<U0059>	LATIN CAPITAL LETTER Y
385	<Z>	Z	<U005A>	LATIN CAPITAL LETTER Z
386	<left-square-bracket>	[<U005B>	LEFT SQUARE BRACKET
387	<backslash>	\	<U005C>	REVERSE SOLIDUS
388	<reverse-solidus>	\	<U005C>	REVERSE SOLIDUS
389	<right-square-bracket>]	<U005D>	RIGHT SQUARE BRACKET
390	<circumflex-accent>	^	<U005E>	CIRCUMFLEX ACCENT
391	<circumflex>	^	<U005E>	CIRCUMFLEX ACCENT
392	<low-line>	_	<U005F>	LOW LINE
393	<underscore>	_	<U005F>	LOW LINE
394	<grave-accent>	`	<U0060>	GRAVE ACCENT
395	<a>	a	<U0061>	LATIN SMALL LETTER A
396		b	<U0062>	LATIN SMALL LETTER B
397	<c>	c	<U0063>	LATIN SMALL LETTER C
398	<d>	d	<U0064>	LATIN SMALL LETTER D
399	<e>	e	<U0065>	LATIN SMALL LETTER E
400	<f>	f	<U0066>	LATIN SMALL LETTER F
401	<g>	g	<U0067>	LATIN SMALL LETTER G
402	<h>	h	<U0068>	LATIN SMALL LETTER H
403	<i>	i	<U0069>	LATIN SMALL LETTER I
404	<j>	j	<U006A>	LATIN SMALL LETTER J
405	<k>	k	<U006B>	LATIN SMALL LETTER K
406	<l>	l	<U006C>	LATIN SMALL LETTER L
407	<m>	m	<U006D>	LATIN SMALL LETTER M
408	<n>	n	<U006E>	LATIN SMALL LETTER N
409	<o>	o	<U006F>	LATIN SMALL LETTER O
410	<p>	p	<U0070>	LATIN SMALL LETTER P
411	<q>	q	<U0071>	LATIN SMALL LETTER Q
412	<r>	r	<U0072>	LATIN SMALL LETTER R
413	<s>	s	<U0073>	LATIN SMALL LETTER S
414	<t>	t	<U0074>	LATIN SMALL LETTER T
415	<u>	u	<U0075>	LATIN SMALL LETTER U
416	<v>	v	<U0076>	LATIN SMALL LETTER V
417	<w>	w	<U0077>	LATIN SMALL LETTER W
418	<x>	x	<U0078>	LATIN SMALL LETTER X

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

427

428 This Technical Report may use other symbolic character names than the above in
 429 examples, to illustrate the use of the range of symbols allowed by the syntax specified in
 430 4.1.1.

431

432 4 FDCC-set

433

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

439

440	LC_IDENTIFICATION	Versions and status of categories
441	LC_CTYPE	Character classification, case conversion and code 442 transformation.
443	LC_COLLATE	Collation order.
444	LC_TIME	Date and time formats.
445	LC_NUMERIC	Numeric, non-monetary formatting.
446	LC_MONETARY	Monetary formatting.
447	LC_MESSAGES	Formats of informative and diagnostic messages and 448 interactive responses.
449	LC_PAPER	Paper format
450	LC_NAME	Format of writing personal names
451	LC_ADDRESS	Format of postal addresses
452	LC_TELEPHONE	Format for telephone numbers, and other telephone 453 information

454

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

461

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

466

467 4.1 FDCC-set definition

468

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

472 The **FDCC-set definition text** shall contain one or more FDCC-set category source
473 definitions, and shall not contain more than one definition for the same FDCC-set
474 category. If the text contains source definitions for more than one category, application-
475 defined categories, if present, shall appear after the categories defined by this clause. A
476 category source definition shall contain either the definition of a category or a copy
477 directive. In the event that some of the information for a FDCC-set category, as specified
478 in this Technical Report, is missing from the FDCC-set source definition, the behaviour of
479 that category, if it is referenced, is unspecified. A FDCC-set category is the normal way of
480 specifying a single FDCC.

481
482 There are no **naming conventions** for FDCC-sets specified in this Technical Report, but
483 ISO/IEC 15897:1999 specifies naming rules for POSIX locales, charmaps and
484 repertoiremaps, that may also be applied to FDCC-sets, charmaps and repertoiremaps
485 specified according to this Technical Report.

486
487 A **category source definition** shall consist of a category header, a category body, and a
488 category trailer. A category header shall consist of the character string naming of the
489 category, beginning with the characters "LC_". The category trailer shall consist of the
490 string "END", followed by one or more "blank"s and the string used in the corresponding
491 category header.

492
493 The **category body** shall consist of one or more lines of text. Each line shall be one of the
494 following:

- 495
496 - a line containing an identifier, optionally followed by one or more operands. Identifiers
497 shall be either keywords, identifying a particular FDCC, or collating elements, or
498 section symbols,
499 - one of transliteration statements defined in 4.3.

500
501 In addition to the keywords defined in this Technical Report, the source can contain
502 application-defined keywords. Each **keyword** within a category shall have a unique name
503 (i.e., two categories can have a commonly-named keyword); no keyword shall start with
504 the characters "LC_". Identifiers shall be separated from the operands by one or more
505 "blank"s.

506
507 **Operands** shall be characters, collating elements, section symbols, or strings of characters.
508 Strings shall be enclosed in double-quotes. Literal double-quotes within strings shall be
509 preceded by the <escape character>, described below. When a keyword is followed by
510 more than one operand, the operands shall be separated by semicolons; "blank"s shall be
511 allowed before and/or after a semicolon.

512 513 514 **4.1.1 Character representation**

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

- 520
521 (0) The left angle bracket (<) is a reserved symbol, denoting the
522 start of a symbolic name; when used to represent itself

523 outside a symbolic name it shall be preceded by the escape
524 character.

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

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

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

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

555
556 , ; < > escape_char

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

558
559 Example: c ä "May"

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

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

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

574 Example: \x63;\xe7;

575

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

580

581 Example: \d99; \d231;

582

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

588

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

590

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

597

598 **4.1.2 Continuation of lines**

599

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

603

604 **4.1.3 Names for copy keyword**

605

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

608

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

614

615 **4.1.4 Pre-category statements**

616

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

619

620 **4.1.4.1 comment_char**

621

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

624

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

626

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

632

633 **4.1.4.2 escape_char**

634

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

637

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

639

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

643

644 **4.1.4.3 repertoiremap**

645

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

649

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

651

652 The name shall be one of:

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

657

658 **4.1.4.4 charmap**

659

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

662

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

664

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

669

670 The name shall be one of:

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

674

4.2 LC_IDENTIFICATION

The LC_IDENTIFICATION category defines properties of the FDCC-set, and which specification methods the FDCC-set is conforming to. All keywords are mandatory unless otherwise noted, and the operands are strings. The following keywords shall be defined:

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

If any of the above information is non-existent, it must be stated in each case; the corresponding string is then the empty string. If required information is not present in ISO 639 or ISO 3166, the relevant Maintenance Authority should be approached to get the needed item registered.

Note: Only one language can be addressed with the concepts of a FDCC-set; to address for example a bilingual culture, one need to have 2 FDCC-sets.

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

The "i18n" LC_IDENTIFICATION category is:


```

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 "1999-12-20"
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

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

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

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

The **decimal symbolic ellipsis** ("....") specification is only valid between symbolic character names. The symbolic names shall consist of zero or more nonnumeric characters from the set shown with visible glyphs in Table 1, followed by an integer formed by one or more decimal digits. The characters preceding the decimal integer shall be identical in 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 ellipsis** ("..(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 **digit** Define the characters to be classified as numeric digits. Digits
834 corresponding to the values 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9 can be specified
835 in groups of 10 digits, and in ascending order of the values they represent.
836 The digits of the portable character set are automatically included. If this
837 keyword is not specified, the digits 0 through 9 of the portable character set

838		shall automatically belong to this class, with application-defined character
839		values. The "digit" keyword is used to specify which characters are
840		accepted as digits in input to an application, such as characters typed in or
841		scanned in from an input text file, and should list digits used with all the
842		scripts supported by the FDCC-set. The keyword may be omitted.
843	outdigit	Define the characters to be classified as numeric digits for output from an
844		application, such as to a printer or a display or a output text file. Digits
845		corresponding to the values <0>, <1>, <2>, <3>, <4>, <5>, <6>, <7>, <8>,
846		and <9> can be specified, and in ascending order of the values they
847		represent. The intended use is for all places where digits are used for
848		output, including numeric and monetary formatting, and date and time
849		formatting. Only one set of 10 digits may be specified. If this keyword is
850		not specified, the digits 0 through 9 of the portable character set shall
851		automatically belong to this class, with application-defined character values.
852		The keyword may be omitted.
853	blank	Define characters to be classified as "blank" characters. If this keyword is
854		unspecified, the characters <space> and <tab>, with application-defined
855		character values, shall belong to this character class.
856	space	Define characters to be classified as white-space characters, to find
857		syntactical boundaries. No character specified for the keywords "upper",
858		"lower", "alpha", "digit", "graph", or "xdigit" shall be specified. If this
859		keyword is not specified, the characters <space>, <form-feed>, <newline>,
860		<carriage-return>, <tab>, and <vertical-tab>, shall automatically belong to
861		this class, with application-defined character values. Any characters
862		included in the class "blank" shall be automatically included. The class
863		should not include the NO-BREAK spaces characters <U00A0>, <U2007>,
864		<UFEFF>, as these characters should not be used for word boundaries. The
865		keyword may be omitted.
866	cntrl	Define characters to be classified as control characters. No character
867		specified for the keywords "upper", "lower", "alpha", "digit", "punct",
868		"graph", "print", or "xdigit" shall be specified. The keyword shall be
869		specified.
870	punct	Define characters to be classified as punctuation characters. No character
871		specified for the keywords "upper", "lower", "alpha", "digit", "cntrl",
872		"xdigit", or as the <space> character shall be specified. The keyword shall
873		be specified.
874	xdigit	Define the characters to be classified as hexadecimal digits. Only the
875		characters defined for the class "digit" shall be specified, in ascending
876		sequence by numerical value, followed by one or more sets of six characters
877		representing the hexadecimal digits 10 through 15, with each set in
878		ascending order (for example <A>, , <C>, <D>, <E>, <F>, <a>, ,
879		<c>, <d>, <e>, <f>). If this keyword is not specified, the digits <0> through
880		<9>, the uppercase letters "A" through <F>, and the lowercase letters <a>
881		through <f>, shall automatically belong to this class, with application-
882		defined character values.
883	graph	Define characters to be classified as printable characters, not including the
884		<space> character. If this keyword is not specified, characters specified for
885		the keywords "upper", "lower", "alpha", "digit", "xdigit", and "punct" shall
886		belong to this character class. No character specified for the keyword "cntrl"
887		shall be specified.

888	print	Define characters to be classified as printable characters, including the
889		<space> character. If this keyword is not provided, characters specified for
890		the keywords upper, lower, alpha, digit, xdigit, punct, graph, and the
891		<space> character shall belong to this character class. No character
892		specified for the keyword "cntrl" shall be specified.
893	toupper	Define the mapping of lowercase letters to uppercase letters. The operand
894		shall consist of character pairs, separated by semicolons. The characters in
895		each character pair shall be separated by a comma and the pair enclosed by
896		parentheses. The first character in each pair shall be the lowercase letter, the
897		second the corresponding uppercase letter. Only characters specified for the
898		keywords "lower" and "upper" shall be specified. If this keyword is not
899		specified, the lowercase letters <a> through <z>, and their corresponding
900		uppercase letters <A> through <Z>, shall automatically be included, with
901		application-defined character values.
902	tolower	Define the mapping of uppercase letters to lowercase letters. The operand
903		shall consist of character pairs, separated by semicolons. The characters in
904		each character pair are separated by a comma and the pair enclosed by
905		parentheses. The first character in each pair shall be the uppercase letter, the
906		second the corresponding lowercase letter. Only characters specified for the
907		keywords "lower" and "upper" shall be specified. If this keyword is speci-
908		fied, the uppercase letters <A> through <Z>, and their corresponding
909		lowercase letter, shall be specified. If this keyword is not specified, the
910		mapping shall be the reverse mapping of the one specified for toupper.
911	class	Define characters to be classified in the class with the name given in the
912		first operand, which is a string. This string shall only contain characters of
913		the portable character set that either has the string "LETTER" in its
914		description, or is a digit or <hyphen-minus> or <low-line>. The following
915		operands are characters. This keyword is optional. The keyword can only be
916		specified once per named class. The following two names shall be
917		recognized:
918	combining	Characters to form composite graphic symbols, such
919		as characters listed in ISO/IEC 10646:1993 annex B.1.
920	combining_level3	Characters to form composite graphic symbols, that
921		may also be represented by other characters, such as
922		characters listed in ISO/IEC 10646-1:1993 annex B.2.
923		The class names "upper", "lower", "alpha", "digit", "space", "cntrl", "punct",
924		"graph", "print", "xdigit", and "blank" are taken to mean the classes defined
925		by the respective keywords.
926	map	Define the mapping of characters. The first operand is a string, defining the
927		name of the mapping. The string shall only contain letters, digits and
928		<hyphen-minus> and <low-line> from the portable character set. The
929		following operands shall consist of character pairs, separated by semicolons.
930		The characters in each character pair shall be separated by a comma and the
931		pair enclosed by parentheses. The first character in each pair shall be the
932		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
952	upper		+	A	x	x	x	x	A	A	+	x
953	lower	+		A	x	x	x	x	A	A	+	x
954	alpha	+	+		x	x	x	x	A	A	+	x
955	digit	x	x	x		x	x	x	A	A	A	x
956	space	x	x	x	x		+	*	*	*	x	+
957	cntrl	x	x	x	x	+		x	x	x	x	+
958	punct	x	x	x	x	+	x		A	A	x	+
959	graph	+	+	+	+	+	x	+		A	+	+
960	print	+	+	+	+	+	x	+	+		+	+
961	xdigit	+	+	+	+	x	x	x	A	A		x
962	blank	x	x	x	x	A	+	*	*	*	x	

963

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 normally 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

different. The language to be transliterated to is identified with the FDCC-set, which may also be used to identify a specific language to be transliterated from. Transliteration may also be to a specific repertoire of characters, determined for example by limitations of displaying equipment, or what the user can intelligibly read. The capabilities here allows for multiple fallback, so that the specification can be valid for all target character repertoires, eliminating the need for specific data for each target repertoire. Transliteration of an incoming character string to a character string in a FDCC-set can be specified with the following keywords and transliteration statements.

translit_start The "translit_start" keyword is followed by one or more transliteration statements assigning character transliteration values to transliterating elements, and include statements copying transliteration specifications from other FDCC-sets.

translit_end The end of the transliteration statements.

include The name of the FDCC-set in text form to transliterate from, and the repertoire map for the FDCC-set to be used for the definition of the transliteration statements. Other transliteration statements may follow to replace specification of the copied FDCC-set. This keyword is optional.

default_missing defines a string of one or more characters to be used if no transliteration statement can be applied to a input <transliteration-source>.

translit_ignore defines a set of characters, separated by semicolons, that are to be ignored in the incoming character string. The characters may use the notations defined in 4.3 for lists of characters.

redefine This keyword introduces a list of transliteration statements where each of the <transliteration_source> strings have been defined previously in the specification, and the new transliteration statements then replaces the old transliteration statements for the <transliteration_source> strings specified.

4.3.2.1 Transliteration statements

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

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

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

If a transliteration statement contains more than one <transliteration_string>, the order that each <transliteration_string> occurs in the transliteration statement defines the precedence order for choosing a particular <transliteration_string> to substitute for the <transliteration_source>. When a process makes use of a transliteration statement to transliterate text, and that transliteration statement contains more than one <transliteration_string>, that process shall choose the first <transliteration_string>, in the defined precedence order, that satisfies the requirements of the transliteration.

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     translit_start
1074     include "de_DE"; "de_remap"
1075     default_missing <?>
1076     translit_ignore <U3200>..

```

1081

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

1083

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

1086

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.

1090

1091 The "translit_ignore" keyword specifies that a set of Ideographic characters (the range <U3200>..
 1092 be ignored for the transliteration.

1093

1094 The next 3 lines are transliteration statements.

1095

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.

1099

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

1102
1103 The third transliteration statement transliterates the two Latin letters K and O into the Japanese Hiragana character
1104 KO.

1105
1106 The transliteration sections is terminated via the "translit_end" keyword in the above example.
1107

1108 4.3.3 "i18n" LC_CTYPE category

1109
1110 The "i18n" FDCC-set for the LC_CTYPE is defined as follows:

```

1111 LC_CTYPE
1112 % The following is the 14652 i18n fdcc-set LC_CTYPE category.
1113 % It covers ISO/IEC 10646-1 including Cor.1 and AMD 1 thru 9
1114 % The "upper" class reflects the uppercase characters of class "alpha"
1115 upper /
1116 % TABLE 1 BASIC LATIN
1117 <U0041>..<U005A>;/
1118 % TABLE 2 LATIN-1 SUPPLEMENT
1119 <U00C0>..<U00D6>;<U00D8>..<U00DE>;/
1120 % TABLE 3 LATIN EXTENDED-A
1121 <U0100>..(2)..<U0136>;/
1122 <U0139>..(2)..<U0147>;/
1123 <U014A>..(2)..<U0178>;/
1124 <U0179>..(2)..<U017D>;/
1125 % TABLE 4 LATIN EXTENDED-B
1126 <U0181>;<U0182>..(2)..<U0186>;<U0187>;/
1127 <U0189>..<U018B>;<U018E>..<U0191>;<U0193>;<U0194>;/
1128 <U0196>..<U0198>;<U019C>;<U019D>;<U019F>;/
1129 <U01A0>..(2)..<U01A4>;/
1130 <U01A7>;<U01A9>;<U01AC>;<U01AE>;<U01AF>;<U01B1>..<U01B3>;/
1131 <U01B5>;<U01B7>;<U01B8>;<U01BC>;<U01C4>;<U01C5>;<U01C7>;<U01C8>;/
1132 <U01CA>;<U01CB>;/
1133 <U01CD>..(2)..<U01DB>;/
1134 <U01DE>..(2)..<U01EE>;/
1135 <U01F1>;<U01F2>;<U01F4>;<U01FA>..(2)..<U01FE>/
1136 % TABLE 5 LATIN EXTENDED-B
1137 <U0200>..(2)..<U0216>;/
1138 % TABLE 6 IPA EXTENSIONS
1139 <U0262>;<U026A>;<U0274>;<U0276>;/
1140 <U0280>;<U0281>;<U028F>;<U0299>;<U029B>;<U029C>;<U029F>;/
1141 % TABLE 9 BASIC GREEK
1142 <U0386>;<U0388>..<U038A>;<U038C>;<U038E>;<U038F>;<U0391>..<U03A1>;/
1143 <U03A3>..<U03AB>;/
1144 % TABLE 10 GREEK SYMBOLS AND COPTIC
1145 <U03E3>..(2)..<U03EF>;/
1146 % TABLE 11 CYRILLIC
1147 <U0401>..<U040C>;<U040E>..<U042F>;<U0460>..(2)..<U047E>;/
1148 % TABLE 12 CYRILLIC
1149 <U0480>;<U0490>..(2)..<U04BE>;<U04C1>;<U04C3>;<U04C7>;<U04CB>;/
1150 <U04D0>..(2)..<U04EA>;<U04EE>..(2)..<U04F4>;<U04F8>;/
1151 % TABLE 13 ARMENIAN
1152 <U0531>..<U0556>;/
1153 % TABLE 28 GEORGIAN
1154 <U10A0>..<U10C5>;/
1155 % TABLE 31 LATIN EXTENDED ADDITIONAL
1156 <U1E00>..(2)..<U1E7E>;/
1157 % TABLE 32 LATIN EXTENDED ADDITIONAL
1158 <U1E80>..(2)..<U1E94>;/
1159 <U1EA0>..(2)..<U1EF8>;/
1160 % TABLE 33 GREEK EXTENDED
1161 <U1F08>..<U1F0F>;<U1F18>..<U1F1D>;<U1F28>..<U1F2F>;<U1F38>..<U1F3F>;/
1162 <U1F48>..<U1F4D>;<U1F59>..(2)..<U1F5F>;<U1F68>..<U1F6F>;/
1163 % TABLE 34 GREEK EXTENDED
1164 <U1F88>..<U1F8F>;<U1F98>..<U1F9F>;<U1FA8>..<U1FAF>;<U1FB8>..<U1FBC>;/
1165 <U1FC8>..<U1FCC>;<U1FD8>..<U1FDB>;<U1FE8>..<U1FEC>;<U1FF8>..<U1FFC>
1166 % TABLE 28 GEORGIAN is not addressed as the letters does not have
1167 a uppercase/lowercase relation
1168 %
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 <U1FBF>;<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 % TABLE 1 BASIC LATIN
1373     <U0021>..<U002F>;<U003A>..<U0040>;<U005B>..<U0060>;/
1374     <U007B>..<U007E>;/
1375 % TABLE 2 LATIN-1 SUPPLEMENT
1376     <U00A0>..<U00A9>;<U00AB>..<U00B9>;<U00BB>..<U00BF>;<U00D7>;<U00F7>;/
1377     <U02C7>;<U02D8>..<U02DD>;/
1378     <U037E>;<U0482>;<U055A>..<U055F>;<U0589>;<U05BE>;<U05C0>;<U05C3>;/
1379     <U05F3>;<U05F4>;<U060C>;<U061B>;<U061F>;<U0640>;<U064B>..<U0652>;/
1380     <U066A>..<U066D>;<U06D4>;<U06DD>..<U06E1>;<U06E9>..<U06EC>;<U10FB>;/
1381     <U2010>..<U2029>;<U2030>..<U2046>;<U20A0>..<U20AA>;<U2100>..<U210B>;/
1382     <U210D>..<U2110>;<U2112>..<U211B>;<U211D>..<U2127>;<U212A>..<U212C>;/
1383     <U212E>..<U2138>;<U2200>..<U22F1>;<U2300>;<U2302>..<U237A>;<U2400>..<U2424>;/
1384     <U2440>..<U244A>;<U2580>..<U2595>;<U25A0>..<U25EF>;<U2600>..<U2613>;/
1385     <U261A>..<U266F>;<U2701>..<U2704>;<U2706>..<U2709>;<U270C>..<U2727>;/
1386     <U2729>..<U274B>;<U274D>;<U274F>..<U2752>;<U2756>;<U2758>..<U275E>;/
1387     <U2761>..<U2767>;<U3000>..<U3020>;<U3030>;<U3036>;<U3037>;<U303F>;<U3164>;/
1388     <U3190>..<U319F>;<U3200>..<U321C>;<U3220>..<U3243>;<U3260>..<U327B>;/
1389     <U327F>..<U32B0>;<U32C0>..<U32CB>;<U32D0>..<U32FE>;<U3300>..<U3376>;/
1390     <U337B>..<U33DD>;<U33E0>..<U33FE>;<UFD3E>;<UFD3F>;<UFE49>..<UFE52>;/
1391     <UFE54>..<UFE66>;<UFE68>..<UFE6B>;<UFEFF>;<UFF01>..<UFF0F>;<UFF1A>..<UFF20>;/
1392     <UFF3B>..<UFF40>;<UFF5B>..<UFF5E>;<UFF61>..<UFF65>;<UFF70>;<UFF9E>..<UFFA0>;/
1393     <UFFE0>..<UFFE6>;<UFFE8>..<UFFEE>;<UFFFD>
1394 %
1395 graph /
1396     <U0021>..<U007E>;<U00A0>..<U01F5>;<U01FA>..<U0217>;/
1397     <U0250>..<U02A8>;<U02B0>..<U02DE>;<U02E0>..<U02E9>;<U0300>..<U0345>;/
1398     <U0360>;<U0361>;<U0374>;<U0375>;<U037A>;<U037E>;<U0384>..<U038A>;<U038C>;/
1399     <U038E>..<U03A1>;<U03A3>..<U03CE>;<U03D0>..<U03D6>;<U03DA>;<U03DC>;<U03DE>;/
1400     <U03E0>;<U03E2>..<U03F3>;<U0401>..<U040C>;<U040E>..<U044F>;/
1401     <U0451>..<U045C>;<U045E>..<U0486>;<U0490>..<U04C4>;<U04C7>;<U04C8>;/
1402     <U04CB>;<U04CC>;<U04D0>..<U04EB>;<U04EE>..<U04F5>;<U04F8>;<U04F9>;/
1403     <U0531>..<U0556>;<U0559>..<U055F>;<U0561>..<U0587>;<U0589>;/
1404     <U0591>..<U05A1>;<U05A3>..<U05AF>;<U05B0>..<U05B9>;/
1405     <U05BB>..<U05C4>;<U05D0>..<U05EA>;<U05F0>..<U05F4>;<U060C>;<U061B>;<U061F>;/
1406     <U0621>..<U063A>;<U0640>..<U0652>;<U0660>..<U066D>;<U0670>..<U06B7>;/
1407     <U06BA>..<U06BE>;<U06C0>..<U06CE>;<U06D0>..<U06ED>;<U06F0>..<U06F9>;/
1408     <U0901>..<U0903>;<U0905>..<U0939>;<U093C>..<U094D>;<U0950>..<U0954>;/
1409     <U0958>..<U0970>;<U0981>..<U0983>;<U0985>..<U098C>;<U098F>;<U0990>;/
1410     <U0993>..<U09A8>;<U09AA>..<U09B0>;<U09B2>;<U09B6>..<U09B9>;<U09BC>;/
1411     <U09BE>..<U09C4>;<U09C7>;<U09C8>;<U09CB>..<U09CD>;<U09D7>;<U09DC>;<U09DD>;/

```

```

1412 <U09DF>..<U09E3>;<U09E6>..<U09FA>;<U0A02>;<U0A05>..<U0A0A>;<U0A0F>;<U0A10>;/
1413 <U0A13>..<U0A28>;<U0A2A>..<U0A30>;<U0A32>;<U0A33>;<U0A35>;<U0A36>;/
1414 <U0A38>;<U0A39>;<U0A3C>;<U0A3E>..<U0A42>;<U0A47>;<U0A48>;<U0A4B>..<U0A4D>;/
1415 <U0A59>..<U0A5C>;<U0A5E>;<U0A66>..<U0A74>;<U0A81>..<U0A83>;<U0A85>..<U0A8B>;/
1416 <U0A8D>;<U0A8F>..<U0A91>;<U0A93>..<U0AA8>;<U0AAA>..<U0AB0>;/
1417 <U0AB2>;<U0AB3>;<U0AB5>..<U0AB9>..<U0ABC>..<U0AC5>;<U0AC7>..<U0AC9>;/
1418 <U0ACB>..<U0ACD>;<U0AD0>;<U0AE0>;<U0AE6>..<U0AEF>;<U0B01>..<U0B03>;/
1419 <U0B05>..<U0B0C>;<U0B0F>;<U0B10>;<U0B13>..<U0B28>;<U0B2A>..<U0B30>;/
1420 <U0B32>;<U0B33>;<U0B36>..<U0B39>;<U0B3C>..<U0B43>;<U0B47>;<U0B48>;/
1421 <U0B4B>..<U0B4D>;<U0B56>;<U0B57>;<U0B5C>;<U0B5D>;<U0B5F>..<U0B61>;/
1422 <U0B66>..<U0B70>;<U0B82>;<U0B83>;<U0B85>..<U0B8A>;<U0B8E>..<U0B90>;/
1423 <U0B92>..<U0B95>;<U0B99>;<U0B9A>;<U0B9C>;<U0B9E>;<U0B9F>;<U0BA3>;<U0BA4>;/
1424 <U0BA8>..<U0BAA>;<U0BAE>..<U0BB5>;<U0BB7>..<U0BB9>;<U0BBE>..<U0BC2>;/
1425 <U0BC6>..<U0BC8>;<U0BCA>..<U0BCD>;<U0BD7>..<U0BE7>..<U0BF2>;<U0C01>..<U0C03>;/
1426 <U0C05>..<U0C0C>;<U0C0E>..<U0C10>;<U0C12>..<U0C28>;<U0C2A>..<U0C33>;/
1427 <U0C35>..<U0C39>;<U0C3E>..<U0C44>;<U0C46>..<U0C48>;<U0C4A>..<U0C4D>;/
1428 <U0C55>;<U0C56>;<U0C60>;<U0C61>;<U0C66>..<U0C6F>;<U0C82>;<U0C83>;/
1429 <U0C85>..<U0C8C>;<U0C8E>..<U0C90>;<U0C92>..<U0CA8>;<U0CAA>..<U0CB3>;/
1430 <U0CB5>..<U0CB9>;<U0CBE>..<U0CC4>;<U0CC6>..<U0CC8>;<U0CCA>..<U0CCD>;/
1431 <U0CD5>;<U0CD6>;<U0CDE>;<U0CE0>;<U0CE1>;<U0CE6>..<U0CEF>;<U0D02>;<U0D03>;/
1432 <U0D05>..<U0D0C>;<U0D0E>..<U0D10>;<U0D12>..<U0D28>;<U0D2A>..<U0D39>;/
1433 <U0D3E>..<U0D43>;<U0D46>..<U0D48>;<U0D4A>..<U0D57>;<U0D60>;<U0D61>;/
1434 <U0D66>..<U0D6F>;<U0E01>..<U0E3A>;<U0E3F>..<U0E5B>;<U0E81>;<U0E82>;<U0E84>;/
1435 <U0E87>;<U0E88>;<U0E8A>;<U0E8D>;<U0E94>..<U0E97>;<U0E99>..<U0E9F>;/
1436 <U0EA1>..<U0EA3>;<U0EA5>;<U0EA7>;<U0EAA>;<U0EAB>;<U0EAD>..<U0EB9>;/
1437 <U0EBB>..<U0EBD>;<U0EC0>..<U0EC4>;<U0EC6>;<U0EC8>..<U0ECD>;<U0ED0>..<U0ED9>;/
1438 <U0EDC>;<U0EDD>;/
1439 <U0F00>..<U0F47>;<U0F49>..<U0F69>;<U0F71>..<U0F7F>;/
1440 <U10A0>..<U10C5>;<U10D0>..<U10F6>;<U10FB>;<U1100>..<U1159>;/
1441 <U115F>..<U11A2>;<U11A8>..<U11F9>;<U1E00>..<U1E9B>;<U1EA0>..<U1EF9>;/
1442 <U1F00>..<U1F15>;<U1F18>..<U1F1D>;<U1F20>..<U1F45>;<U1F48>..<U1F4D>;/
1443 <U1F50>..<U1F57>;<U1F59>;<U1F5B>;<U1F5D>;<U1F5F>..<U1F7D>;<U1F80>..<U1FB4>;/
1444 <U1FB6>..<U1FC4>;<U1FC6>..<U1FD3>;<U1FD6>..<U1FDB>;<U1FDD>..<U1FEF>;/
1445 <U1FF2>..<U1FF4>;<U1FF6>..<U1FF8>;<U2000>..<U202E>;<U2030>..<U2046>;/
1446 <U206A>..<U2070>;<U2074>..<U208E>;<U20A0>..<U20AB>;<U20D0>..<U20E1>;/
1447 <U2100>..<U2138>;<U2153>..<U2182>;<U2190>..<U21EA>;<U2200>..<U22F1>;<U2300>;/
1448 <U2302>..<U237A>;<U2400>..<U2424>;<U2440>..<U244A>;<U2460>..<U24EA>;/
1449 <U2500>..<U2595>;<U25A0>..<U25EF>;<U2600>..<U2613>;<U261A>..<U266F>;/
1450 <U2701>..<U2704>;<U2706>..<U2709>;<U270C>..<U2727>;<U2729>..<U274B>;<U274D>;/
1451 <U274F>..<U2752>;<U2756>;<U2758>..<U275E>;<U2761>..<U2767>;<U2776>..<U2794>;/
1452 <U2798>..<U27AF>;<U27B1>..<U27BE>;<U3000>..<U3037>;<U303F>;<U3041>..<U3094>;/
1453 <U3099>..<U309E>;<U30A1>..<U30FE>;<U3105>..<U312C>;<U3131>..<U318E>;/
1454 <U3190>..<U319F>;<U3200>..<U321C>;<U3220>..<U3243>;<U3260>..<U327B>;/
1455 <U327F>..<U32B0>;<U32C0>..<U32CB>;<U32D0>..<U32FE>;<U3300>..<U3376>;/
1456 <U337B>..<U33DD>;<U33E0>..<U33FE>;<UFB00>..<UFB06>;<UFB13>..<UFB17>;/
1457 <UFB1E>..<UFB36>;<UFB38>..<UFB3C>;<UFB3E>;<UFB40>;<UFB41>;<UFB43>;<UFB44>;/
1458 <UFB46>..<UFB81>;<UFB83>..<UFD3F>;<UFD50>..<UFD8F>;<UFD92>..<UFD97>;/
1459 <UFD9F>..<UFD9F>;<UFE20>..<UFE23>;<UFE30>..<UFE44>;<UFE49>..<UFE52>;/
1460 <UFE54>..<UFE66>;<UFE68>..<UFE6B>;<UFE70>..<UFE72>;<UFE74>;<UFE76>..<UFEFC>;/
1461 <UFEFF>;<UFF01>..<UFF5E>;<UFF61>..<UFFBE>;<UFFC2>..<UFFC7>;/
1462 <UFFCA>..<UFFCF>;<UFFD2>..<UFFD7>;<UFFDA>..<UFFDC>;<UFFE0>..<UFFE6>;/
1463 <UFFE8>..<UFFEE>;<UFFFD>
1464
1465 % "print" is by default "graph", and the <space> character
1466 %
1467 xdigit <U0030>..<U0039>;<U0041>..<U0046>;<U0061>..<U0066>
1468 %
1469 blank <U0008>;<U0020>;<U2000>..<U2006>;<U2008>..<U200B>;<U3000>
1470 %
1471 toupper /
1472 (<U0061>,<U0041>);(<U0062>,<U0042>);(<U0063>,<U0043>);(<U0064>,<U0044>);/
1473 (<U0065>,<U0045>);(<U0066>,<U0046>);(<U0067>,<U0047>);(<U0068>,<U0048>);/
1474 (<U0069>,<U0049>);(<U006A>,<U004A>);(<U006B>,<U004B>);(<U006C>,<U004C>);/
1475 (<U006D>,<U004D>);(<U006E>,<U004E>);(<U006F>,<U004F>);(<U0070>,<U0050>);/
1476 (<U0071>,<U0051>);(<U0072>,<U0052>);(<U0073>,<U0053>);(<U0074>,<U0054>);/
1477 (<U0075>,<U0055>);(<U0076>,<U0056>);(<U0077>,<U0057>);(<U0078>,<U0058>);/
1478 (<U0079>,<U0059>);(<U007A>,<U005A>);(<U00E0>,<U00C0>);(<U00E1>,<U00C1>);/
1479 (<U00E2>,<U00C2>);(<U00E3>,<U00C3>);(<U00E4>,<U00C4>);(<U00E5>,<U00C5>);/
1480 (<U00E6>,<U00C6>);(<U00E7>,<U00C7>);(<U00E8>,<U00C8>);(<U00E9>,<U00C9>);/
1481 (<U00EA>,<U00CA>);(<U00EB>,<U00CB>);(<U00EC>,<U00CC>);(<U00ED>,<U00CD>);/
1482 (<U00EE>,<U00CE>);(<U00EF>,<U00CF>);(<U00F0>,<U00D0>);(<U00F1>,<U00D1>);/
1483 (<U00F2>,<U00D2>);(<U00F3>,<U00D3>);(<U00F4>,<U00D4>);(<U00F5>,<U00D5>);/
1484 (<U00F6>,<U00D6>);(<U00F7>,<U00D7>);(<U00F8>,<U00D8>);(<U00F9>,<U00D9>);/
1485 (<U00FA>,<U00DA>);(<U00FB>,<U00DB>);(<U00FC>,<U00DC>);(<U00FD>,<U00DD>);/
1486 (<U00FE>,<U00DE>);(<U00FF>,<U00DF>);(<U0100>,<U0080>);(<U0101>,<U0081>);/
1487 (<U0102>,<U0082>);(<U0103>,<U0083>);(<U0104>,<U0084>);(<U0105>,<U0085>);/
1488 (<U0106>,<U0086>);(<U0107>,<U0087>);(<U0108>,<U0088>);(<U0109>,<U0089>);/
1489 (<U010A>,<U008A>);(<U010B>,<U008B>);(<U010C>,<U008C>);(<U010D>,<U008D>);/
1490 (<U010E>,<U008E>);(<U010F>,<U008F>);(<U0110>,<U0090>);(<U0111>,<U0091>);/
1491 (<U0112>,<U0092>);(<U0113>,<U0093>);(<U0114>,<U0094>);(<U0115>,<U0095>);/
1492 (<U0116>,<U0096>);(<U0117>,<U0097>);(<U0118>,<U0098>);(<U0119>,<U0099>);/
1493 (<U011A>,<U009A>);(<U011B>,<U009B>);(<U011C>,<U009C>);(<U011D>,<U009D>);/
1494 (<U011E>,<U009E>);(<U011F>,<U009F>);(<U0120>,<U00A0>);(<U0121>,<U00A1>);/
1495 (<U0122>,<U00A2>);(<U0123>,<U00A3>);(<U0124>,<U00A4>);(<U0125>,<U00A5>);/

```

1491 (<U0127>, <U0126>); (<U0129>, <U0128>); (<U012B>, <U012A>); (<U012D>, <U012C>); /
 1492 (<U012F>, <U012E>); (<U0133>, <U0132>); (<U0135>, <U0134>); (<U0137>, <U0136>); /
 1493 (<U013A>, <U0139>); (<U013C>, <U013B>); (<U013E>, <U013D>); (<U0140>, <U013F>); /
 1494 (<U0142>, <U0141>); (<U0144>, <U0143>); (<U0146>, <U0145>); (<U0148>, <U0147>); /
 1495 (<U014B>, <U014A>); (<U014D>, <U014C>); (<U014F>, <U014E>); (<U0151>, <U0150>); /
 1496 (<U0153>, <U0152>); (<U0155>, <U0154>); (<U0157>, <U0156>); (<U0159>, <U0158>); /
 1497 (<U015B>, <U015A>); (<U015D>, <U015C>); (<U015F>, <U015E>); (<U0161>, <U0160>); /
 1498 (<U0163>, <U0162>); (<U0165>, <U0164>); (<U0167>, <U0166>); (<U0169>, <U0168>); /
 1499 (<U016B>, <U016A>); (<U016D>, <U016C>); (<U016F>, <U016E>); (<U0171>, <U0170>); /
 1500 (<U0173>, <U0172>); (<U0175>, <U0174>); (<U0177>, <U0176>); (<U017A>, <U0179>); /
 1501 (<U017C>, <U017B>); (<U017E>, <U017D>); (<U017F>, <U0053>); (<U0183>, <U0182>); /
 1502 (<U0185>, <U0184>); (<U0188>, <U0187>); (<U018C>, <U018B>); (<U0192>, <U0191>); /
 1503 (<U0199>, <U0198>); (<U01A1>, <U01A0>); (<U01A3>, <U01A2>); (<U01A5>, <U01A4>); /
 1504 (<U01A8>, <U01A7>); (<U01AD>, <U01AC>); (<U01B0>, <U01AF>); (<U01B4>, <U01B3>); /
 1505 (<U01B6>, <U01B5>); (<U01B9>, <U01B8>); (<U01BD>, <U01BC>); (<U01C5>, <U01C4>); /
 1506 (<U01C6>, <U01C4>); (<U01C8>, <U01C7>); /
 1507 (<U01C9>, <U01C7>); (<U01CB>, <U01CA>); (<U01CC>, <U01CA>); /
 1508 (<U01CE>, <U01CD>); (<U01D0>, <U01CF>); (<U01D2>, <U01D1>); (<U01D4>, <U01D3>); /
 1509 (<U01D6>, <U01D5>); (<U01D8>, <U01D7>); (<U01DA>, <U01D9>); (<U01DC>, <U01DB>); /
 1510 (<U01DD>, <U018E>); (<U01DF>, <U01DE>); (<U01E1>, <U01E0>); (<U01E3>, <U01E2>); /
 1511 (<U01E5>, <U01E4>); (<U01E7>, <U01E6>); (<U01E9>, <U01E8>); (<U01EB>, <U01EA>); /
 1512 (<U01ED>, <U01EC>); (<U01EF>, <U01EE>); (<U01F2>, <U01F1>); /
 1513 (<U01F3>, <U01F1>); (<U01F5>, <U01F4>); (<U01FB>, <U01FA>); (<U01FD>, <U01FC>); /
 1514 (<U01FF>, <U01FE>); (<U0201>, <U0200>); (<U0203>, <U0202>); (<U0205>, <U0204>); /
 1515 (<U0207>, <U0206>); (<U0209>, <U0208>); (<U020B>, <U020A>); (<U020D>, <U020C>); /
 1516 (<U020F>, <U020E>); (<U0211>, <U0210>); (<U0213>, <U0212>); (<U0215>, <U0214>); /
 1517 (<U0217>, <U0216>); (<U0253>, <U0181>); (<U0254>, <U0186>); (<U0256>, <U0189>); /
 1518 (<U0257>, <U018A>); (<U0259>, <U018F>); (<U025B>, <U0190>); (<U0260>, <U0193>); /
 1519 (<U0262>, <U0194>); (<U0268>, <U0197>); (<U0269>, <U0196>); (<U026F>, <U019C>); /
 1520 (<U0272>, <U019D>); (<U0275>, <U019F>); (<U0283>, <U01A9>); (<U0288>, <U01AE>); /
 1521 (<U028A>, <U01B1>); (<U028B>, <U01B2>); (<U0292>, <U01B7>); (<U03AC>, <U0386>); /
 1522 (<U03AD>, <U0388>); (<U03AE>, <U0389>); (<U03AF>, <U038A>); (<U03B1>, <U0391>); /
 1523 (<U03B2>, <U0392>); (<U03B3>, <U0393>); (<U03B4>, <U0394>); (<U03B5>, <U0395>); /
 1524 (<U03B6>, <U0396>); (<U03B7>, <U0397>); (<U03B8>, <U0398>); (<U03B9>, <U0399>); /
 1525 (<U03BA>, <U039A>); (<U03BB>, <U039B>); (<U03BC>, <U039C>); (<U03BD>, <U039D>); /
 1526 (<U03BE>, <U039E>); (<U03BF>, <U039F>); (<U03C0>, <U03A0>); (<U03C1>, <U03A1>); /
 1527 (<U03C2>, <U03A3>); (<U03C3>, <U03A3>); (<U03C4>, <U03A4>); (<U03C5>, <U03A5>); /
 1528 (<U03C6>, <U03A6>); (<U03C7>, <U03A7>); (<U03C8>, <U03A8>); (<U03C9>, <U03A9>); /
 1529 (<U03CA>, <U03AA>); (<U03CB>, <U03AB>); (<U03CC>, <U038C>); (<U03CD>, <U038E>); /
 1530 (<U03CE>, <U038F>); /
 1531 (<U03E3>, <U03E2>); (<U03E5>, <U03E4>); (<U03E7>, <U03E6>); (<U03E9>, <U03E8>); /
 1532 (<U03EB>, <U03EA>); (<U03ED>, <U03EC>); (<U03EF>, <U03EE>); /
 1533 (<U0430>, <U0410>); (<U0431>, <U0411>); (<U0432>, <U0412>); /
 1534 (<U0433>, <U0413>); (<U0434>, <U0414>); (<U0435>, <U0415>); (<U0436>, <U0416>); /
 1535 (<U0437>, <U0417>); (<U0438>, <U0418>); (<U0439>, <U0419>); (<U043A>, <U041A>); /
 1536 (<U043B>, <U041B>); (<U043C>, <U041C>); (<U043D>, <U041D>); (<U043E>, <U041E>); /
 1537 (<U043F>, <U041F>); (<U0440>, <U0420>); (<U0441>, <U0421>); (<U0442>, <U0422>); /
 1538 (<U0443>, <U0423>); (<U0444>, <U0424>); (<U0445>, <U0425>); (<U0446>, <U0426>); /
 1539 (<U0447>, <U0427>); (<U0448>, <U0428>); (<U0449>, <U0429>); (<U044A>, <U042A>); /
 1540 (<U044B>, <U042B>); (<U044C>, <U042C>); (<U044D>, <U042D>); (<U044E>, <U042E>); /
 1541 (<U044F>, <U042F>); (<U0451>, <U0401>); (<U0452>, <U0402>); (<U0453>, <U0403>); /
 1542 (<U0454>, <U0404>); (<U0455>, <U0405>); (<U0456>, <U0406>); (<U0457>, <U0407>); /
 1543 (<U0458>, <U0408>); (<U0459>, <U0409>); (<U045A>, <U040A>); (<U045B>, <U040B>); /
 1544 (<U045C>, <U040C>); (<U045E>, <U040E>); (<U045F>, <U040F>); (<U0461>, <U0406>); /
 1545 (<U0463>, <U0462>); (<U0465>, <U0464>); (<U0467>, <U0466>); (<U0469>, <U0468>); /
 1546 (<U046B>, <U046A>); (<U046D>, <U046C>); (<U046F>, <U046E>); (<U0471>, <U0470>); /
 1547 (<U0473>, <U0472>); (<U0475>, <U0474>); (<U0477>, <U0476>); (<U0479>, <U0478>); /
 1548 (<U047B>, <U047A>); (<U047D>, <U047C>); (<U047F>, <U047E>); (<U0481>, <U0480>); /
 1549 (<U0491>, <U0490>); (<U0493>, <U0492>); (<U0495>, <U0494>); (<U0497>, <U0496>); /
 1550 (<U0499>, <U0498>); (<U049B>, <U049A>); (<U049D>, <U049C>); (<U049F>, <U049E>); /
 1551 (<U04A1>, <U04A0>); (<U04A3>, <U04A2>); (<U04A5>, <U04A4>); (<U04A7>, <U04A6>); /
 1552 (<U04A9>, <U04A8>); (<U04AB>, <U04AA>); (<U04AD>, <U04AC>); (<U04AF>, <U04AE>); /
 1553 (<U04B1>, <U04B0>); (<U04B3>, <U04B2>); (<U04B5>, <U04B4>); (<U04B7>, <U04B6>); /
 1554 (<U04B9>, <U04B8>); (<U04BB>, <U04BA>); (<U04BD>, <U04BC>); (<U04BF>, <U04BE>); /
 1555 (<U04C2>, <U04C1>); (<U04C4>, <U04C3>); (<U04C8>, <U04C7>); (<U04CC>, <U04CB>); /
 1556 (<U04D1>, <U04D0>); (<U04D3>, <U04D2>); (<U04D5>, <U04D4>); (<U04D7>, <U04D6>); /
 1557 (<U04D9>, <U04D8>); (<U04DB>, <U04DA>); (<U04DD>, <U04DC>); (<U04DF>, <U04DE>); /
 1558 (<U04E1>, <U04E0>); (<U04E3>, <U04E2>); (<U04E5>, <U04E4>); (<U04E7>, <U04E6>); /
 1559 (<U04E9>, <U04E8>); (<U04EB>, <U04EA>); (<U04EF>, <U04EE>); (<U04F1>, <U04F0>); /
 1560 (<U04F3>, <U04F2>); (<U04F5>, <U04F4>); (<U04F9>, <U04F8>); (<U0561>, <U0531>); /
 1561 (<U0562>, <U0532>); (<U0563>, <U0533>); (<U0564>, <U0534>); (<U0565>, <U0535>); /
 1562 (<U0566>, <U0536>); (<U0567>, <U0537>); (<U0568>, <U0538>); (<U0569>, <U0539>); /
 1563 (<U056A>, <U053A>); (<U056B>, <U053B>); (<U056C>, <U053C>); (<U056D>, <U053D>); /
 1564 (<U056E>, <U053E>); (<U056F>, <U053F>); (<U0570>, <U0540>); (<U0571>, <U0541>); /
 1565 (<U0572>, <U0542>); (<U0573>, <U0543>); (<U0574>, <U0544>); (<U0575>, <U0545>); /
 1566 (<U0576>, <U0546>); (<U0577>, <U0547>); (<U0578>, <U0548>); (<U0579>, <U0549>); /
 1567 (<U057A>, <U054A>); (<U057B>, <U054B>); (<U057C>, <U054C>); (<U057D>, <U054D>); /
 1568 (<U057E>, <U054E>); (<U057F>, <U054F>); (<U0580>, <U0550>); (<U0581>, <U0551>); /
 1569 (<U0582>, <U0552>); (<U0583>, <U0553>); (<U0584>, <U0554>); (<U0585>, <U0555>); /

1570 (<U0586>,<U0556>);/
 1571 (<U10D0>,<U10A0>); (<U10D1>,<U10A1>); (<U10D2>,<U10A2>); (<U10D3>,<U10A3>);/
 1572 (<U10D4>,<U10A4>); (<U10D5>,<U10A5>); (<U10D6>,<U10A6>); (<U10D7>,<U10A7>);/
 1573 (<U10D8>,<U10A8>); (<U10D9>,<U10A9>); (<U10DA>,<U10AA>); (<U10DB>,<U10AB>);/
 1574 (<U10DC>,<U10AC>); (<U10DD>,<U10AD>); (<U10DE>,<U10AE>); (<U10DF>,<U10AF>);/
 1575 (<U10E0>,<U10B0>); (<U10E1>,<U10B1>); (<U10E2>,<U10B2>); (<U10E3>,<U10B3>);/
 1576 (<U10E4>,<U10B4>); (<U10E5>,<U10B5>); (<U10E6>,<U10B6>); (<U10E7>,<U10B7>);/
 1577 (<U10E8>,<U10B8>); (<U10E9>,<U10B9>); (<U10EA>,<U10BA>); (<U10EB>,<U10BB>);/
 1578 (<U10EC>,<U10BC>); (<U10ED>,<U10BD>); (<U10EE>,<U10BE>); (<U10EF>,<U10BF>);/
 1579 (<U10F0>,<U10C0>); (<U10F1>,<U10C1>); (<U10F2>,<U10C2>); (<U10F3>,<U10C3>);/
 1580 (<U10F4>,<U10C4>); (<U10F5>,<U10C5>);/
 1581 (<U1E01>,<U1E00>); (<U1E03>,<U1E02>); (<U1E05>,<U1E04>);/
 1582 (<U1E07>,<U1E06>); (<U1E09>,<U1E08>); (<U1E0B>,<U1E0A>); (<U1E0D>,<U1E0C>);/
 1583 (<U1E0F>,<U1E0E>); (<U1E11>,<U1E10>); (<U1E13>,<U1E12>); (<U1E15>,<U1E14>);/
 1584 (<U1E17>,<U1E16>); (<U1E19>,<U1E18>); (<U1E1B>,<U1E1A>); (<U1E1D>,<U1E1C>);/
 1585 (<U1E1F>,<U1E1E>); (<U1E21>,<U1E20>); (<U1E23>,<U1E22>); (<U1E25>,<U1E24>);/
 1586 (<U1E27>,<U1E26>); (<U1E29>,<U1E28>); (<U1E2B>,<U1E2A>); (<U1E2D>,<U1E2C>);/
 1587 (<U1E2F>,<U1E2E>); (<U1E31>,<U1E30>); (<U1E33>,<U1E32>); (<U1E35>,<U1E34>);/
 1588 (<U1E37>,<U1E36>); (<U1E39>,<U1E38>); (<U1E3B>,<U1E3A>); (<U1E3D>,<U1E3C>);/
 1589 (<U1E3F>,<U1E3E>); (<U1E41>,<U1E40>); (<U1E43>,<U1E42>); (<U1E45>,<U1E44>);/
 1590 (<U1E47>,<U1E46>); (<U1E49>,<U1E48>); (<U1E4B>,<U1E4A>); (<U1E4D>,<U1E4C>);/
 1591 (<U1E4F>,<U1E4E>); (<U1E51>,<U1E50>); (<U1E53>,<U1E52>); (<U1E55>,<U1E54>);/
 1592 (<U1E57>,<U1E56>); (<U1E59>,<U1E58>); (<U1E5B>,<U1E5A>); (<U1E5D>,<U1E5C>);/
 1593 (<U1E5F>,<U1E5E>); (<U1E61>,<U1E60>); (<U1E63>,<U1E62>); (<U1E65>,<U1E64>);/
 1594 (<U1E67>,<U1E66>); (<U1E69>,<U1E68>); (<U1E6B>,<U1E6A>); (<U1E6D>,<U1E6C>);/
 1595 (<U1E6F>,<U1E6E>); (<U1E71>,<U1E70>); (<U1E73>,<U1E72>); (<U1E75>,<U1E74>);/
 1596 (<U1E77>,<U1E76>); (<U1E79>,<U1E78>); (<U1E7B>,<U1E7A>); (<U1E7D>,<U1E7C>);/
 1597 (<U1E7F>,<U1E7E>); (<U1E81>,<U1E80>); (<U1E83>,<U1E82>); (<U1E85>,<U1E84>);/
 1598 (<U1E87>,<U1E86>); (<U1E89>,<U1E88>); (<U1E8B>,<U1E8A>); (<U1E8D>,<U1E8C>);/
 1599 (<U1E8F>,<U1E8E>); (<U1E91>,<U1E90>); (<U1E93>,<U1E92>); (<U1E95>,<U1E94>);/
 1600 (<U1E9B>,<U1E9A>);/
 1601 (<U1EA1>,<U1EA0>); (<U1EA3>,<U1EA2>); (<U1EA5>,<U1EA4>); (<U1EA7>,<U1EA6>);/
 1602 (<U1EA9>,<U1EA8>); (<U1EAB>,<U1EAA>); (<U1EAD>,<U1EAC>); (<U1EAF>,<U1EAE>);/
 1603 (<U1EB1>,<U1EB0>); (<U1EB3>,<U1EB2>); (<U1EB5>,<U1EB4>); (<U1EB7>,<U1EB6>);/
 1604 (<U1EB9>,<U1EB8>); (<U1EBB>,<U1EBA>); (<U1EBD>,<U1EBC>); (<U1EBF>,<U1EBE>);/
 1605 (<U1EC1>,<U1EC0>); (<U1EC3>,<U1EC2>); (<U1EC5>,<U1EC4>); (<U1EC7>,<U1EC6>);/
 1606 (<U1EC9>,<U1EC8>); (<U1ECB>,<U1ECA>); (<U1ECD>,<U1ECC>); (<U1ECF>,<U1ECE>);/
 1607 (<U1ED1>,<U1ED0>); (<U1ED3>,<U1ED2>); (<U1ED5>,<U1ED4>); (<U1ED7>,<U1ED6>);/
 1608 (<U1ED9>,<U1ED8>); (<U1EDB>,<U1EDA>); (<U1EDD>,<U1EDC>); (<U1EDF>,<U1EDE>);/
 1609 (<U1EE1>,<U1EE0>); (<U1EE3>,<U1EE2>); (<U1EE5>,<U1EE4>); (<U1EE7>,<U1EE6>);/
 1610 (<U1EE9>,<U1EE8>); (<U1EEB>,<U1EEA>); (<U1EED>,<U1EEC>); (<U1EEF>,<U1EEE>);/
 1611 (<U1EF1>,<U1EF0>); (<U1EF3>,<U1EF2>); (<U1EF5>,<U1EF4>); (<U1EF7>,<U1EF6>);/
 1612 (<U1EF9>,<U1EF8>); (<U1F00>,<U1F08>); (<U1F01>,<U1F09>); (<U1F02>,<U1F0A>);/
 1613 (<U1F03>,<U1F0B>); (<U1F04>,<U1F0C>); (<U1F05>,<U1F0D>); (<U1F06>,<U1F0E>);/
 1614 (<U1F07>,<U1F0F>); (<U1F10>,<U1F18>); (<U1F11>,<U1F19>); (<U1F12>,<U1F1A>);/
 1615 (<U1F13>,<U1F1B>); (<U1F14>,<U1F1C>); (<U1F15>,<U1F1D>); (<U1F20>,<U1F28>);/
 1616 (<U1F21>,<U1F29>); (<U1F22>,<U1F2A>); (<U1F23>,<U1F2B>); (<U1F24>,<U1F2C>);/
 1617 (<U1F25>,<U1F2D>); (<U1F26>,<U1F2E>); (<U1F27>,<U1F2F>); (<U1F30>,<U1F38>);/
 1618 (<U1F31>,<U1F39>); (<U1F32>,<U1F3A>); (<U1F33>,<U1F3B>); (<U1F34>,<U1F3C>);/
 1619 (<U1F35>,<U1F3D>); (<U1F36>,<U1F3E>); (<U1F37>,<U1F3F>); (<U1F40>,<U1F48>);/
 1620 (<U1F41>,<U1F49>); (<U1F42>,<U1F4A>); (<U1F43>,<U1F4B>); (<U1F44>,<U1F4C>);/
 1621 (<U1F45>,<U1F4D>); (<U1F51>,<U1F59>); (<U1F53>,<U1F5B>); (<U1F55>,<U1F5D>);/
 1622 (<U1F57>,<U1F5F>); (<U1F60>,<U1F68>); (<U1F61>,<U1F69>); (<U1F62>,<U1F6A>);/
 1623 (<U1F63>,<U1F6B>); (<U1F64>,<U1F6C>); (<U1F65>,<U1F6D>); (<U1F66>,<U1F6E>);/
 1624 (<U1F67>,<U1F6F>); (<U1F70>,<U1F7A>); (<U1F71>,<U1F7B>); (<U1F72>,<U1F7C>);/
 1625 (<U1F73>,<U1F7D>); (<U1F74>,<U1F7E>); (<U1F75>,<U1F7F>); (<U1F76>,<U1F7A>);/
 1626 (<U1F77>,<U1F7B>); (<U1F78>,<U1F7C>); (<U1F79>,<U1F7D>); (<U1F7A>,<U1F7E>);/
 1627 (<U1F7B>,<U1F7F>); (<U1F7C>,<U1F7D>); (<U1F7D>,<U1F7E>); (<U1F80>,<U1F88>);/
 1628 (<U1F81>,<U1F89>); (<U1F82>,<U1F8A>); (<U1F83>,<U1F8B>); (<U1F84>,<U1F8C>);/
 1629 (<U1F85>,<U1F8D>); (<U1F86>,<U1F8E>); (<U1F87>,<U1F8F>); (<U1F90>,<U1F98>);/
 1630 (<U1F91>,<U1F99>); (<U1F92>,<U1F9A>); (<U1F93>,<U1F9B>); (<U1F94>,<U1F9C>);/
 1631 (<U1F95>,<U1F9D>); (<U1F96>,<U1F9E>); (<U1F97>,<U1F9F>); (<U1FA0>,<U1FA8>);/
 1632 (<U1FA1>,<U1FA9>); (<U1FA2>,<U1FAA>); (<U1FA3>,<U1FAB>); (<U1FA4>,<U1FAC>);/
 1633 (<U1FA5>,<U1FAD>); (<U1FA6>,<U1FAE>); (<U1FA7>,<U1FAF>); (<U1FB0>,<U1FB8>);/
 1634 (<U1FB1>,<U1FB9>); (<U1FB3>,<U1FBC>); (<U1FC3>,<U1FCC>); (<U1FD0>,<U1FD8>);/
 1635 (<U1FD1>,<U1FD9>); (<U1FE0>,<U1FE8>); (<U1FE1>,<U1FE9>); (<U1FE5>,<U1FEC>);/
 1636 (<U1FF3>,<U1FFC>);
 1637 tolower /
 1638 (<U0041>,<U0061>); (<U0042>,<U0062>); (<U0043>,<U0063>); (<U0044>,<U0064>);/
 1639 (<U0045>,<U0065>); (<U0046>,<U0066>); (<U0047>,<U0067>); (<U0048>,<U0068>);/
 1640 (<U0049>,<U0069>); (<U004A>,<U006A>); (<U004B>,<U006B>); (<U004C>,<U006C>);/
 1641 (<U004D>,<U006D>); (<U004E>,<U006E>); (<U004F>,<U006F>); (<U0050>,<U0070>);/
 1642 (<U0051>,<U0071>); (<U0052>,<U0072>); (<U0053>,<U0073>); (<U0054>,<U0074>);/
 1643 (<U0055>,<U0075>); (<U0056>,<U0076>); (<U0057>,<U0077>); (<U0058>,<U0078>);/
 1644 (<U0059>,<U0079>); (<U005A>,<U007A>); (<U00C0>,<U00E0>); (<U00C1>,<U00E1>);/
 1645 (<U00C2>,<U00E2>); (<U00C3>,<U00E3>); (<U00C4>,<U00E4>); (<U00C5>,<U00E5>);/
 1646 (<U00C6>,<U00E6>); (<U00C7>,<U00E7>); (<U00C8>,<U00E8>); (<U00C9>,<U00E9>);/
 1647 (<U00CA>,<U00EA>); (<U00CB>,<U00EB>); (<U00CC>,<U00EC>); (<U00CD>,<U00ED>);/
 1648 (<U00CE>,<U00EE>); (<U00CF>,<U00EF>); (<U00D0>,<U00F0>); (<U00D1>,<U00F1>);/

1649 (<U00D2>, <U00F2>); (<U00D3>, <U00F3>); (<U00D4>, <U00F4>); (<U00D5>, <U00F5>); /
 1650 (<U00D6>, <U00F6>); (<U00D8>, <U00F8>); (<U00D9>, <U00F9>); (<U00DA>, <U00FA>); /
 1651 (<U00DB>, <U00FB>); (<U00DC>, <U00FC>); (<U00DD>, <U00FD>); (<U00DE>, <U00FE>); /
 1652 (<U0178>, <U00FF>); (<U0100>, <U0101>); (<U0102>, <U0103>); (<U0104>, <U0105>); /
 1653 (<U0106>, <U0107>); (<U0108>, <U0109>); (<U010A>, <U010B>); (<U010C>, <U010D>); /
 1654 (<U010E>, <U010F>); (<U0110>, <U0111>); (<U0112>, <U0113>); (<U0114>, <U0115>); /
 1655 (<U0116>, <U0117>); (<U0118>, <U0119>); (<U011A>, <U011B>); (<U011C>, <U011D>); /
 1656 (<U011E>, <U011F>); (<U0120>, <U0121>); (<U0122>, <U0123>); (<U0124>, <U0125>); /
 1657 (<U0126>, <U0127>); (<U0128>, <U0129>); (<U012A>, <U012B>); (<U012C>, <U012D>); /
 1658 (<U012E>, <U012F>); (<U0132>, <U0133>); (<U0134>, <U0135>); (<U0136>, <U0137>); /
 1659 (<U0139>, <U013A>); (<U013B>, <U013C>); (<U013D>, <U013E>); (<U013F>, <U0140>); /
 1660 (<U0141>, <U0142>); (<U0143>, <U0144>); (<U0145>, <U0146>); (<U0147>, <U0148>); /
 1661 (<U014A>, <U014B>); (<U014C>, <U014D>); (<U014E>, <U014F>); (<U0150>, <U0151>); /
 1662 (<U0152>, <U0153>); (<U0154>, <U0155>); (<U0156>, <U0157>); (<U0158>, <U0159>); /
 1663 (<U015A>, <U015B>); (<U015C>, <U015D>); (<U015E>, <U015F>); (<U0160>, <U0161>); /
 1664 (<U0162>, <U0163>); (<U0164>, <U0165>); (<U0166>, <U0167>); (<U0168>, <U0169>); /
 1665 (<U016A>, <U016B>); (<U016C>, <U016D>); (<U016E>, <U016F>); (<U0170>, <U0171>); /
 1666 (<U0172>, <U0173>); (<U0174>, <U0175>); (<U0176>, <U0177>); (<U0179>, <U017A>); /
 1667 (<U017B>, <U017C>); (<U017D>, <U017E>); (<U0182>, <U0183>); (<U0184>, <U0185>); /
 1668 (<U0187>, <U0188>); (<U0256>, <U0189>); (<U018B>, <U018C>); (<U018E>, <U01DD>); /
 1669 (<U0191>, <U0192>); (<U0198>, <U0199>); (<U01A0>, <U01A1>); (<U01A2>, <U01A3>); /
 1670 (<U01A4>, <U01A5>); (<U01A7>, <U01A8>); (<U01AC>, <U01AD>); (<U01AF>, <U01B0>); /
 1671 (<U01B3>, <U01B4>); (<U01B5>, <U01B6>); (<U01B8>, <U01B9>); (<U01BC>, <U01BD>); /
 1672 (<U01C4>, <U01C6>); (<U01C5>, <U01C6>); (<U01C7>, <U01C9>); /
 1673 (<U01C8>, <U01C9>); (<U01CA>, <U01CC>); (<U01CB>, <U01CC>); /
 1674 (<U01CD>, <U01CE>); (<U01CF>, <U01D0>); (<U01D1>, <U01D2>); /
 1675 (<U01D3>, <U01D4>); (<U01D5>, <U01D6>); (<U01D7>, <U01D8>); (<U01D9>, <U01DA>); /
 1676 (<U01DB>, <U01DC>); (<U01DE>, <U01DF>); (<U01E0>, <U01E1>); (<U01E2>, <U01E3>); /
 1677 (<U01E4>, <U01E5>); (<U01E6>, <U01E7>); (<U01E8>, <U01E9>); (<U01EA>, <U01EB>); /
 1678 (<U01EC>, <U01ED>); (<U01EE>, <U01EF>); (<U01F1>, <U01F3>); /
 1679 (<U01F2>, <U01F3>); (<U01F4>, <U01F5>); (<U01FA>, <U01FB>); (<U01FC>, <U01FD>); /
 1680 (<U01FE>, <U01FF>); (<U0200>, <U0201>); (<U0202>, <U0203>); (<U0204>, <U0205>); /
 1681 (<U0206>, <U0207>); (<U0208>, <U0209>); (<U020A>, <U020B>); (<U020C>, <U020D>); /
 1682 (<U020E>, <U020F>); (<U0210>, <U0211>); (<U0212>, <U0213>); (<U0214>, <U0215>); /
 1683 (<U0216>, <U0217>); (<U0181>, <U0253>); (<U0186>, <U0254>); (<U018A>, <U0257>); /
 1684 (<U018E>, <U0258>); (<U018F>, <U0259>); (<U0190>, <U025B>); (<U0193>, <U0260>); /
 1685 (<U0194>, <U0263>); (<U0197>, <U0268>); (<U0196>, <U0269>); (<U019C>, <U026F>); /
 1686 (<U019D>, <U0272>); (<U01A9>, <U0283>); (<U01AE>, <U0288>); (<U01B1>, <U028A>); /
 1687 (<U01B2>, <U028B>); (<U01B7>, <U0292>); (<U0386>, <U03AC>); (<U0388>, <U03AD>); /
 1688 (<U0389>, <U03AE>); (<U038A>, <U03AF>); (<U0391>, <U03B1>); (<U0392>, <U03B2>); /
 1689 (<U0393>, <U03B3>); (<U0394>, <U03B4>); (<U0395>, <U03B5>); (<U0396>, <U03B6>); /
 1690 (<U0397>, <U03B7>); (<U0398>, <U03B8>); (<U0399>, <U03B9>); (<U039A>, <U03BA>); /
 1691 (<U039B>, <U03BB>); (<U039C>, <U03BC>); (<U039D>, <U03BD>); (<U039E>, <U03BE>); /
 1692 (<U039F>, <U03BF>); (<U03A0>, <U03C0>); (<U03A1>, <U03C1>); (<U03A3>, <U03C3>); /
 1693 (<U03A4>, <U03C4>); (<U03A5>, <U03C5>); (<U03A6>, <U03C6>); (<U03A7>, <U03C7>); /
 1694 (<U03A8>, <U03C8>); (<U03A9>, <U03C9>); (<U03AA>, <U03CA>); (<U03AB>, <U03CB>); /
 1695 (<U03AC>, <U03CC>); (<U038E>, <U03CD>); (<U038F>, <U03CE>); (<U03E2>, <U03E3>); /
 1696 (<U03E4>, <U03E5>); (<U03E6>, <U03E7>); (<U03E8>, <U03E9>); (<U03EA>, <U03EB>); /
 1697 (<U03EC>, <U03ED>); (<U03EE>, <U03EF>); (<U0410>, <U0430>); /
 1698 (<U0411>, <U0431>); (<U0412>, <U0432>); (<U0413>, <U0433>); (<U0414>, <U0434>); /
 1699 (<U0415>, <U0435>); (<U0416>, <U0436>); (<U0417>, <U0437>); (<U0418>, <U0438>); /
 1700 (<U0419>, <U0439>); (<U041A>, <U043A>); (<U041B>, <U043B>); (<U041C>, <U043C>); /
 1701 (<U041D>, <U043D>); (<U041E>, <U043E>); (<U041F>, <U043F>); (<U0420>, <U0440>); /
 1702 (<U0421>, <U0441>); (<U0422>, <U0442>); (<U0423>, <U0443>); (<U0424>, <U0444>); /
 1703 (<U0425>, <U0445>); (<U0426>, <U0446>); (<U0427>, <U0447>); (<U0428>, <U0448>); /
 1704 (<U0429>, <U0449>); (<U042A>, <U044A>); (<U042B>, <U044B>); (<U042C>, <U044C>); /
 1705 (<U042D>, <U044D>); (<U042E>, <U044E>); (<U042F>, <U044F>); (<U0401>, <U0451>); /
 1706 (<U0402>, <U0452>); (<U0403>, <U0453>); (<U0404>, <U0454>); (<U0405>, <U0455>); /
 1707 (<U0406>, <U0456>); (<U0407>, <U0457>); (<U0408>, <U0458>); (<U0409>, <U0459>); /
 1708 (<U040A>, <U045A>); (<U040B>, <U045B>); (<U040C>, <U045C>); (<U040E>, <U045E>); /
 1709 (<U040F>, <U045F>); (<U0460>, <U0461>); (<U0462>, <U0463>); (<U0464>, <U0465>); /
 1710 (<U0466>, <U0467>); (<U0468>, <U0469>); (<U046A>, <U046B>); (<U046C>, <U046D>); /
 1711 (<U046E>, <U046F>); (<U0470>, <U0471>); (<U0472>, <U0473>); (<U0474>, <U0475>); /
 1712 (<U0476>, <U0477>); (<U0478>, <U0479>); (<U047A>, <U047B>); (<U047C>, <U047D>); /
 1713 (<U047E>, <U047F>); (<U0480>, <U0481>); (<U0490>, <U0491>); (<U0492>, <U0493>); /
 1714 (<U0494>, <U0495>); (<U0496>, <U0497>); (<U0498>, <U0499>); (<U049A>, <U049B>); /
 1715 (<U049C>, <U049D>); (<U049E>, <U049F>); (<U04A0>, <U04A1>); (<U04A2>, <U04A3>); /
 1716 (<U04A4>, <U04A5>); (<U04A6>, <U04A7>); (<U04A8>, <U04A9>); (<U04AA>, <U04AB>); /
 1717 (<U04AC>, <U04AD>); (<U04AE>, <U04AF>); (<U04B0>, <U04B1>); (<U04B2>, <U04B3>); /
 1718 (<U04B4>, <U04B5>); (<U04B6>, <U04B7>); (<U04B8>, <U04B9>); (<U04BA>, <U04BB>); /
 1719 (<U04BC>, <U04BD>); (<U04BE>, <U04BF>); (<U04C1>, <U04C2>); (<U04C3>, <U04C4>); /
 1720 (<U04C7>, <U04C8>); (<U04CB>, <U04CC>); (<U04D0>, <U04D1>); (<U04D2>, <U04D3>); /
 1721 (<U04D4>, <U04D5>); (<U04DE>, <U04D7>); (<U04D8>, <U04D9>); (<U04DA>, <U04DB>); /
 1722 (<U04DC>, <U04DD>); (<U04DE>, <U04DF>); (<U04E0>, <U04E1>); (<U04E2>, <U04E3>); /
 1723 (<U04E4>, <U04E5>); (<U04E6>, <U04E7>); (<U04E8>, <U04E9>); (<U04EA>, <U04EB>); /
 1724 (<U04EE>, <U04EF>); (<U04F0>, <U04F1>); (<U04F2>, <U04F3>); (<U04F4>, <U04F5>); /
 1725 (<U04F8>, <U04F9>); (<U0531>, <U0561>); (<U0532>, <U0562>); (<U0533>, <U0563>); /
 1726 (<U0534>, <U0564>); (<U0535>, <U0565>); (<U0536>, <U0566>); (<U0537>, <U0567>); /
 1727 (<U0538>, <U0568>); (<U0539>, <U0569>); (<U053A>, <U056A>); (<U053B>, <U056B>); /

```

1728 (<U053C>, <U056C>); (<U053D>, <U056D>); (<U053E>, <U056E>); (<U053F>, <U056F>); /
1729 (<U0540>, <U0570>); (<U0541>, <U0571>); (<U0542>, <U0572>); (<U0543>, <U0573>); /
1730 (<U0544>, <U0574>); (<U0545>, <U0575>); (<U0546>, <U0576>); (<U0547>, <U0577>); /
1731 (<U0548>, <U0578>); (<U0549>, <U0579>); (<U054A>, <U057A>); (<U054B>, <U057B>); /
1732 (<U054C>, <U057C>); (<U054D>, <U057D>); (<U054E>, <U057E>); (<U054F>, <U057F>); /
1733 (<U0550>, <U0580>); (<U0551>, <U0581>); (<U0552>, <U0582>); (<U0553>, <U0583>); /
1734 (<U0554>, <U0584>); (<U0555>, <U0585>); (<U0556>, <U0586>); /
1735 (<U10A0>, <U10D0>); (<U10A1>, <U10D1>); (<U10A2>, <U10D2>); (<U10A3>, <U10D3>); /
1736 (<U10A4>, <U10D4>); (<U10A5>, <U10D5>); (<U10A6>, <U10D6>); (<U10A7>, <U10D7>); /
1737 (<U10A8>, <U10D8>); (<U10A9>, <U10D9>); (<U10AA>, <U10DA>); (<U10AB>, <U10DB>); /
1738 (<U10AC>, <U10DC>); (<U10AD>, <U10DD>); (<U10AE>, <U10DE>); (<U10AF>, <U10DF>); /
1739 (<U10B0>, <U10E0>); (<U10B1>, <U10E1>); (<U10B2>, <U10E2>); (<U10B3>, <U10E3>); /
1740 (<U10B4>, <U10E4>); (<U10B5>, <U10E5>); (<U10B6>, <U10E6>); (<U10B7>, <U10E7>); /
1741 (<U10B8>, <U10E8>); (<U10B9>, <U10E9>); (<U10BA>, <U10EA>); (<U10BB>, <U10EB>); /
1742 (<U10BC>, <U10EC>); (<U10BD>, <U10ED>); (<U10BE>, <U10EE>); (<U10BF>, <U10EF>); /
1743 (<U10C0>, <U10F0>); (<U10C1>, <U10F1>); (<U10C2>, <U10F2>); (<U10C3>, <U10F3>); /
1744 (<U10C4>, <U10F4>); (<U10C5>, <U10F5>); /
1745 (<U1E00>, <U1E01>); /
1746 (<U1E02>, <U1E03>); (<U1E04>, <U1E05>); (<U1E06>, <U1E07>); (<U1E08>, <U1E09>); /
1747 (<U1E0A>, <U1E0B>); (<U1E0C>, <U1E0D>); (<U1E0E>, <U1E0F>); (<U1E10>, <U1E11>); /
1748 (<U1E12>, <U1E13>); (<U1E14>, <U1E15>); (<U1E16>, <U1E17>); (<U1E18>, <U1E19>); /
1749 (<U1E1A>, <U1E1B>); (<U1E1C>, <U1E1D>); (<U1E1E>, <U1E1F>); (<U1E20>, <U1E21>); /
1750 (<U1E22>, <U1E23>); (<U1E24>, <U1E25>); (<U1E26>, <U1E27>); (<U1E28>, <U1E29>); /
1751 (<U1E2A>, <U1E2B>); (<U1E2C>, <U1E2D>); (<U1E2E>, <U1E2F>); (<U1E30>, <U1E31>); /
1752 (<U1E32>, <U1E33>); (<U1E34>, <U1E35>); (<U1E36>, <U1E37>); (<U1E38>, <U1E39>); /
1753 (<U1E3A>, <U1E3B>); (<U1E3C>, <U1E3D>); (<U1E3E>, <U1E3F>); (<U1E40>, <U1E41>); /
1754 (<U1E42>, <U1E43>); (<U1E44>, <U1E45>); (<U1E46>, <U1E47>); (<U1E48>, <U1E49>); /
1755 (<U1E4A>, <U1E4B>); (<U1E4C>, <U1E4D>); (<U1E4E>, <U1E4F>); (<U1E50>, <U1E51>); /
1756 (<U1E52>, <U1E53>); (<U1E54>, <U1E55>); (<U1E56>, <U1E57>); (<U1E58>, <U1E59>); /
1757 (<U1E5A>, <U1E5B>); (<U1E5C>, <U1E5D>); (<U1E5E>, <U1E5F>); (<U1E60>, <U1E61>); /
1758 (<U1E62>, <U1E63>); (<U1E64>, <U1E65>); (<U1E66>, <U1E67>); (<U1E68>, <U1E69>); /
1759 (<U1E6A>, <U1E6B>); (<U1E6C>, <U1E6D>); (<U1E6E>, <U1E6F>); (<U1E70>, <U1E71>); /
1760 (<U1E72>, <U1E73>); (<U1E74>, <U1E75>); (<U1E76>, <U1E77>); (<U1E78>, <U1E79>); /
1761 (<U1E7A>, <U1E7B>); (<U1E7C>, <U1E7D>); (<U1E7E>, <U1E7F>); (<U1E80>, <U1E81>); /
1762 (<U1E82>, <U1E83>); (<U1E84>, <U1E85>); (<U1E86>, <U1E87>); (<U1E88>, <U1E89>); /
1763 (<U1E8A>, <U1E8B>); (<U1E8C>, <U1E8D>); (<U1E8E>, <U1E8F>); (<U1E90>, <U1E91>); /
1764 (<U1E92>, <U1E93>); (<U1E94>, <U1E95>); (<U1EA0>, <U1EA1>); (<U1EA2>, <U1EA3>); /
1765 (<U1EA4>, <U1EA5>); (<U1EA6>, <U1EA7>); (<U1EA8>, <U1EA9>); (<U1EAA>, <U1EAB>); /
1766 (<U1EAC>, <U1EAD>); (<U1EAE>, <U1EAF>); (<U1EB0>, <U1EB1>); (<U1EB2>, <U1EB3>); /
1767 (<U1EB4>, <U1EB5>); (<U1EB6>, <U1EB7>); (<U1EB8>, <U1EB9>); (<U1EBA>, <U1EBB>); /
1768 (<U1EBC>, <U1EBD>); (<U1EBE>, <U1EBF>); (<U1EC0>, <U1EC1>); (<U1EC2>, <U1EC3>); /
1769 (<U1EC4>, <U1EC5>); (<U1EC6>, <U1EC7>); (<U1EC8>, <U1EC9>); (<U1ECA>, <U1ECB>); /
1770 (<U1ECC>, <U1ECD>); (<U1ECE>, <U1ECF>); (<U1ED0>, <U1ED1>); (<U1ED2>, <U1ED3>); /
1771 (<U1ED4>, <U1ED5>); (<U1ED6>, <U1ED7>); (<U1ED8>, <U1ED9>); (<U1EDA>, <U1EDB>); /
1772 (<U1EDC>, <U1EDD>); (<U1EDE>, <U1EDF>); (<U1EE0>, <U1EE1>); (<U1EE2>, <U1EE3>); /
1773 (<U1EE4>, <U1EE5>); (<U1EE6>, <U1EE7>); (<U1EE8>, <U1EE9>); (<U1EEA>, <U1EEB>); /
1774 (<U1EEC>, <U1EED>); (<U1EEE>, <U1EEF>); (<U1EF0>, <U1EF1>); (<U1EF2>, <U1EF3>); /
1775 (<U1EF4>, <U1EF5>); (<U1EF6>, <U1EF7>); (<U1EF8>, <U1EF9>); (<U1F08>, <U1F00>); /
1776 (<U1F09>, <U1F01>); (<U1F0A>, <U1F02>); (<U1F0B>, <U1F03>); (<U1F0C>, <U1F04>); /
1777 (<U1F0D>, <U1F05>); (<U1F0E>, <U1F06>); (<U1F0F>, <U1F07>); (<U1F18>, <U1F10>); /
1778 (<U1F19>, <U1F11>); (<U1F1A>, <U1F12>); (<U1F1B>, <U1F13>); (<U1F1C>, <U1F14>); /
1779 (<U1F1D>, <U1F15>); (<U1F28>, <U1F20>); (<U1F29>, <U1F21>); (<U1F2A>, <U1F22>); /
1780 (<U1F2B>, <U1F23>); (<U1F2C>, <U1F24>); (<U1F2D>, <U1F25>); (<U1F2E>, <U1F26>); /
1781 (<U1F2F>, <U1F27>); (<U1F38>, <U1F30>); (<U1F39>, <U1F31>); (<U1F3A>, <U1F32>); /
1782 (<U1F3B>, <U1F33>); (<U1F3C>, <U1F34>); (<U1F3D>, <U1F35>); (<U1F3E>, <U1F36>); /
1783 (<U1F3F>, <U1F37>); (<U1F48>, <U1F40>); (<U1F49>, <U1F41>); (<U1F4A>, <U1F42>); /
1784 (<U1F4B>, <U1F43>); (<U1F4C>, <U1F44>); (<U1F4D>, <U1F45>); (<U1F4E>, <U1F46>); /
1785 (<U1F5B>, <U1F53>); (<U1F5D>, <U1F55>); (<U1F5F>, <U1F57>); (<U1F68>, <U1F60>); /
1786 (<U1F69>, <U1F61>); (<U1F6A>, <U1F62>); (<U1F6B>, <U1F63>); (<U1F6C>, <U1F64>); /
1787 (<U1F6D>, <U1F65>); (<U1F6E>, <U1F66>); (<U1F6F>, <U1F67>); (<U1FBA>, <U1F70>); /
1788 (<U1FBB>, <U1F71>); (<U1FC8>, <U1F72>); (<U1FC9>, <U1F73>); (<U1FCA>, <U1F74>); /
1789 (<U1FCB>, <U1F75>); (<U1FDA>, <U1F76>); (<U1FDB>, <U1F77>); (<U1FF8>, <U1F78>); /
1790 (<U1FF9>, <U1F79>); (<U1FEA>, <U1F7A>); (<U1FEB>, <U1F7B>); (<U1FFA>, <U1F7C>); /
1791 (<U1FFB>, <U1F7D>); (<U1F88>, <U1F80>); (<U1F89>, <U1F81>); (<U1F8A>, <U1F82>); /
1792 (<U1F8B>, <U1F83>); (<U1F8C>, <U1F84>); (<U1F8D>, <U1F85>); (<U1F8E>, <U1F86>); /
1793 (<U1F8F>, <U1F87>); (<U1F98>, <U1F90>); (<U1F99>, <U1F91>); (<U1F9A>, <U1F92>); /
1794 (<U1F9B>, <U1F93>); (<U1F9C>, <U1F94>); (<U1F9D>, <U1F95>); (<U1F9E>, <U1F96>); /
1795 (<U1F9F>, <U1F97>); (<U1FA8>, <U1FA0>); (<U1FA9>, <U1FA1>); (<U1FAA>, <U1FA2>); /
1796 (<U1FAB>, <U1FA3>); (<U1FAC>, <U1FA4>); (<U1FAD>, <U1FA5>); (<U1FAE>, <U1FA6>); /
1797 (<U1FAF>, <U1FA7>); (<U1FB8>, <U1FB0>); (<U1FB9>, <U1FB1>); (<U1FBC>, <U1FB3>); /
1798 (<U1FCC>, <U1FC3>); (<U1FD8>, <U1FD0>); (<U1FD9>, <U1FD1>); (<U1FE8>, <U1FE0>); /
1799 (<U1FE9>, <U1FE1>); (<U1FEC>, <U1FE5>); (<U1FFC>, <U1FF3>)

```

```

1800 %
1801 % The "combining" class reflects ISO/IEC 10646-1 annex B.1
1802 % That is, all combining characters (level 2+3).
1803 class "combining"; /
1804 <U0300>..<U036F>; <U20D0>..<U20FF>; <UFE20>..<UFE2F>; /
1805 <U0483>..<U0486>;<U0591>..<U05A1>;<U05A3>..<U05B9>; /
1806 <U05BB>..<U05BD>;<U05BF>;<U05C1>;<U05C2>;<U05C4>;<U064B>..<U0652>;<U0670>; /

```



```

1807 <U06D7>..

```

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.
- (5) Many-to-Many substitution. A string of one or more characters is substituted by another string (or an empty string, i.e., the character or characters shall be ignored for collation purposes).
- (6) Equivalence class definition. Two or more collating elements have the same collation value (primary weight).
- (7) Ordering by weights. When two strings are compared to determine their relative order, the two strings are first broken up into a series of collating elements, and

- 1867 each successive pair of elements are compared according to the relative primary
 1868 weights for the elements. If equal, and more than one weight has been assigned,
 1869 then the pairs of collating elements are recompared according to the relative
 1870 subsequent weights, until either a pair of collating elements compare unequal or the
 1871 weights are exhausted.
- 1872 (8) Easy reordering of characters. ISO/IEC 14651 has a template for collation
 1873 specification that with just a few modifications can be culturally correct for a
 1874 specific culture. Here the "reorder-after" keyword gives a convenient way to
 1875 modify a FDCC-set template.
- 1876 (9) Easy reordering of sections. The template in ISO/IEC 14651 gives an ordering of
 1877 the sections that may not be culturally acceptable in certain cultures. The keyword
 1878 "reorder-section-after" gives a convenient way to modify the order of sections in a
 1879 FDCC-set template.

1880
 1881 The following keywords shall be recognized in a collation sequence definition. Some of
 1882 them are described in detail in the following subclauses.

1884	copy	Specify the name of an existing FDCC-set to be used as the source for the definition of this category. If this keyword is specified, only the "reorder-after", "reorder-end", "reorder-sections-after" and "reorder-sections-end" keywords may also be specified. The FDCC-set shall be copied in source form.
1890	coll_weight_max	Define as a decimal number the number of collation levels that an interpreting system needs to support for this FDCC-set, this value is elsewhere referred as the COLL_WEIGHT_MAX limit. An interpreting system shall cater for up to 7 collating levels.
1895	section-symbol	Define a section symbol representing a set of collation order statements. The section is defined with the "order_start" keyword until the next "order_start" or "order_end" keyword. This keyword is optional.
1900	collating-element	Define a collating-element symbol representing a multicharacter collating element. This keyword is optional.
1903	collating-symbol	Define one or more collating symbols for use in collation order statements. This keyword is optional.
1905	symbol-equivalence	Define a collating-symbol to be equivalent to another defined collating-symbol.
1907	order_start	Define collation rules. This statement is followed by one or more collation order statements, assigning character collation values and collation weights to collating elements.
1911	order_end	Specify the end of the collation-order statements.
1912	reorder-after	Redefine collating rules. Specify after which collating element the redefinition of collation order shall take order. This statement is followed by one or more collation order statements, reassigning character collation values and collation weights to collating elements.

1918	reorder-end	Specify the end of the "reorder-after" collating order statements.
1919		
1920	reorder-section-after	Redefine the order of sections. This statement is followed by one or more section symbols, reassigning character collation values and collation weights to collating elements.
1921		
1922		
1923		
1924	reorder-section-end	Specify the end of the "reorder-sections" section order statements.
1925		
1926		

4.4.1 Collation statements

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

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

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

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

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

The absolute ellipsis symbol ("...") specifies that a sequence of characters shall collate according to their encoded character values. It shall be interpreted as indicating that all characters with a coded character set value higher than the value of the character in the preceding line, and lower than the coded character set value for the character in the following line, in the current coded character set, shall be placed in the character collation order between the previous and the following character in ascending order according to their coded character set values. An initial ellipsis shall be interpreted as if the preceding line specified the <NUL> character, and a trailing ellipsis as if the following line specified the highest coded character set value in the current coded character set. An ellipsis shall 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 deprecated. Symbolic ellipses may be used as the ellipses symbol, but generating symbolic character names, and thus have a better chance of portability between applications.

The symbolic ellipses (".." or "....") specifies a sequence of collating statements. It shall be interpreted as indicating that all characters with symbolic names higher than the

1968 symbolic name of the character in the preceding line, and lower than the coded character
1969 set value for the character in the following line, shall be placed in the character collation
1970 order between the previous and the following character in ascending order.

1971
1972 The symbol "UNDEFINED" shall be interpreted as including all coded character set values
1973 not specified explicitly or via the ellipsis or one of the symbolic ellipses symbols. Such
1974 characters shall be inserted in the character collation order at the point indicated by the
1975 symbol, and in ascending order according to their coded character set values. If no
1976 "UNDEFINED" symbol is specified, and the current coded character set contains
1977 characters not specified in this clause, the utility shall issue a warning message and place
1978 such characters at the end of the character collation order.

1979
1980 The optional operands for each collation-element shall be used to define the primary,
1981 secondary, or subsequent weights for the collating element. The first operand specifies the
1982 relative primary weight, the second the relative secondary weight, and so on. Two or more
1983 collation-elements can be assigned the same weight; they belong to the same equivalence
1984 class if they have the same primary weight. Collation shall behave as if, for each weight
1985 level, "IGNORE"d elements are removed. Then each successive pair of elements shall be
1986 compared according to the relative weights for the elements. If the two strings compare
1987 equal, the process shall be repeated for the next weight level, up to the limit
1988 "COLL_WEIGHTS_MAX" of the associated FDCC-set.

1989
1990 Weights shall be expressed as characters (in any of the forms specified here), <collating-
1991 symbol>s, <collating-element>s, an ellipsis, or the special symbol "IGNORE". A single
1992 character, a <collating-symbol>, or a <collating-element> shall represent the relative order
1993 in the character collating sequence of the character or symbol, rather than the character or
1994 characters themselves.

1995
1996 One-to-many mapping is indicated by specifying two or more concatenated characters or
1997 symbolic names. Thus, if the character <ss> is given the string <s><s> as a weight,
1998 comparisons shall be performed as if all occurrences of the character <ss> are replaced by
1999 <s><s>. If it is desirable to define <ss> and <s><s> as an equivalence class, then a
2000 collating-element must be defined for the string "ss", as in the example below.

2001
2002 All characters specified via an ellipsis shall by default be assigned unique weights, equal
2003 to the relative order of characters. Characters specified via an explicit or implicit
2004 "UNDEFINED" special symbol shall by default be assigned the same primary weight (i.e.,
2005 belong to the same equivalence class). An ellipsis symbol as a weight shall be interpreted
2006 to mean that each character in the sequence shall have unique weights, equal to the
2007 relative order of their character in the character collation sequence. Secondary and
2008 subsequent weights have unique values. The use of the ellipsis as a weight shall be treated
2009 as an error if the collating element is neither an ellipsis nor the special symbol
2010 "UNDEFINED".

2011
2012 The special keyword "IGNORE" as a weight shall indicate that when strings are compared
2013 using the weights at the level where "IGNORE" is specified, the collating element shall be
2014 ignored; i.e., as if the string did not contain the collating element. In regular expressions
2015 and pattern matching, all characters that are "IGNORE"d in their primary weight form an
2016 equivalence class.

2017
2018 A <comment_character> occurring where the delimiter ";" may occur, terminates the

2019 collating statement.

2020

2021 An empty operand shall be interpreted as the collating-element itself.

2022

2023 For example, the collation statement

2024

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

2025

2026

is equal to

2027

```
<a>
```

2028

2029

2030

2031 An ellipsis (absolute or symbolic) can be used as an operand if the collating-element was

2032 an ellipsis, and shall be interpreted as the value of each character defined by the ellipsis.

2033

Example:

2034

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

2051

2052 This example is interpreted as follows:

2053

2054

- 2055 (1) The UNDEFINED means that all characters not specified in this definition (explicitly or via the
- 2056 ellipsis) shall be ignored.
- 2057 (2) <LOW> defines the first collating weight, and thus the lowest weight in this example.
- 2058 (3) All characters between <space> and <a> shall have the same primary equivalence class <LOW> and
- 2059 individual secondary weights based on their ordinal encoded values. (The use of absolute ellipses is
- 2060 deprecated, but used here to illustrate generic use of ellipses. Symbolic ellipses should be used
- 2061 instead).
- 2062 (4) All characters based on the upper or lowercase character "a" belong to the same primary equivalence
- 2063 class.
- 2064 (5) The multicharacter collating element <c><h> is represented by the collating symbol <ch> and belongs
- 2065 to the same primary equivalence class as the multicharacter collating element <C><H>.
- 2066 (6) The <ss> collating element has two weights on the primary level, and it is in the same primary
- 2067 equivalence class as two consecutive <s>-es; on the secondary level the collating element has two
- 2068 weights of the equivalence class <ss>.

2069

2070 4.4.2 "copy" keyword

2071

2072 This keyword specifies the name of an existing FDCC-set to be used as the source for the

2073 definition of this category. The syntax is

2074

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

2075

2076

2077 The <FDCC-set-name> shall consist of one or more characters (in any of the forms

2078 defined in 4.1.1). If this keyword is specified, only the "reorder-after", "reorder-end",

2079 "reorder-sections-after" and "reorder-sections-end" keywords may also be specified. The

2080 FDCC-set shall be copied in source form.

2081

2082 4.4.3 "col_weight_max" keyword

2083
2084 This keyword defines as a decimal number the number of collation levels that an
2085 interpreting system needs to support, this value is elsewhere referred as the
2086 COLL_WEIGHT_MAX limit. The minimum value is 7. The syntax is

```
2087  
2088     "col_weight_max %d\n", <value>
```

2089 4.4.4 "section-symbol" keyword

2090
2091
2092 This keyword shall be used to define symbols for use in section related statements; such
2093 as the "order_start", and "reorder-sections-after" keywords and section-reordering
2094 statements. The syntax is

```
2095  
2096     "section-symbol %s\n", <section-symbol>
```

2097
2098 The <section-symbol> shall be a symbolic name, enclosed between angle brackets (< and
2099 >), and shall not duplicate any symbolic name in the current charmap (if any), or any
2100 other symbolic name defined in this collation definition. A <section-symbol> defined via
2101 this keyword is only defined with the LC_COLLATE category.

```
2102  
2103     Example:  
2104     section-symbol <LATIN>  
2105     section-symbol <ARABIC>
```

2106 4.4.5 "collating-element" keyword

2107
2108
2109 In addition to the collating elements in the character set, the collating-element keyword
2110 shall be used to define multicharacter collating elements. The syntax is

```
2111  
2112     "collating-element %s from %s\n", <collating-symbol>, <string>
```

2113
2114 The <collating-symbol> operand shall be a symbolic name, enclosed between angle
2115 brackets (< and >), and shall not duplicate any symbolic name in the current charmap or
2116 repertoiremap file (if any), or any other symbolic name defined in this collation definition.
2117 The string operand shall be a string of two or more characters that shall collate as an
2118 entity. A <collating-element> defined via this keyword is only defined within the
2119 LC_COLLATE category.

```
2120  
2121     Example with ISO/IEC 10646:  
2122     collating-element <ch> from "<c><h>"  
2123     collating-element <e-acute> from "<e><combining-acute>"  
2124     collating-element <aa> from "<a><a>"
```

2125
2126 Note: The problem of comparing a fully composed character of ISO/IEC 10646 with a
2127 decomposed representation of the same text is normally handled by the two strings
2128 comparing equal up to level 3 (the case level) of ISO/IEC 14651, but distinguishing the
2129 two at the 4th level.

2130 4.4.6 "collating-symbol" keyword

2131
2132
2133 This keyword shall be used to define symbols for use in collation sequence statements;
2134 e.g., between the order_start and the order_end keywords. The syntax is

2135 "collating-symbol %s;%s;...%s\n", <collating-symbol>, <collating-symbol> ...
 2136

2137 The <collating-symbol> shall be a symbolic name, enclosed between angle brackets (< and
 2138 >), and shall not duplicate any symbolic name in the current charmap (if any), or any
 2139 other symbolic name defined in this collation definition. A <collating-symbol> defined via
 2140 this keyword is only defined with the LC_COLLATE category. More than one <collating-
 2141 symbol> may be defined with one "collating-symbol" keyword, and symbolic ellipses may
 2142 be used.

2143
 2144 Example:
 2145 collating-symbol <CAPITAL>
 2146 collating-symbol <HIGH>

2147 **4.4.7 "symbol-equivalence" keyword**

2148
 2149 This keyword shall be used to define symbols for use in collation sequence statements;
 2150 and assign the same weight as another defined symbol. The syntax is

2151
 2152 "symbol-equivalence %s %s\n", <collating-symbol-1>, <collating-symbol-2>

2153
 2154 The <collating-symbol-1> and <collating-symbol-2> shall be symbolic names, enclosed
 2155 between angle brackets (< and >). <collating-symbol-1> shall not duplicate any symbolic
 2156 name in the current charmap (if any), or any other symbolic name defined in this collation
 2157 definition. <collating-symbol-2> is defined elsewhere in the LC_COLLATE category as a
 2158 collating-symbol. The use of <collating-symbol-2> shall be equivalent to using the
 2159 <collating-symbol-2> in the LC_COLLATE category. A <collating-symbol-1> defined via
 2160 this keyword is only defined with the LC_COLLATE category.

2161
 2162 Example
 2163 collating-symbol <CAP>
 2164 symbol-equivalence <CAPITAL> <CAP>

2165 2166 **4.4.8 "order_start" keyword**

2167
 2168 The "order_start" keyword shall precede collation order entries and also defines the
 2169 number of weights for this collation sequence definition, the collation section name and
 2170 other collation rules.

2171
 2172 The syntax of the "order_start" keyword has two forms:

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

2175
 2176 and

2177 "order_start %s;%s;...;%s\n", <section-symbol>, <sort-rules>, <sort-rules> ...

2178
 2179 The operands to the order_start keyword are optional. If present, the operands define rules
 2180 to be applied when strings are compared. The first operand may be a <section-symbol>
 2181 surrounded by "<" and ">" and the set of collating statements following the "order_start"
 2182 keyword until the "order_end" keyword are identified with this <section_symbol> or
 2183 another "order_start" keyword is encountered. The remaining number of operands define
 2184 how many weights each element is assigned; if no operands are present, one forward
 2185 operand is assumed. If present, the first operand defines rules to be applied when

2186 comparing strings using the first (primary) weight; the second when comparing strings
 2187 using the second weight, and so on. Operands shall be separated by semicolons (;). Each
 2188 operand shall consist of one or more collation directives, separated by commas (,). If the
 2189 number of operands exceeds the (COLL_WEIGHTS_MAX) limit, a utility parsing the
 2190 FDCC-set description shall issue a warning message. The following directives shall be
 2191 supported:

2192
 2193 **forward** Specifies that the direction of scanning a part of a string at a given point in a
 2194 string is done towards the logical end of the whole string for this weight level.

2195 **backward** Specifies that the direction of scanning a part of a
 2196 string at a given point in a string is done towards the
 2197 logical beginning of the whole string for this weight
 2198 level.

2199 **position** Specifies that comparison operations for the weight level will consider the
 2200 relative position of non-"IGNORE"d elements in the strings. The string
 2201 containing a non-"IGNORE"d element after the fewest IGNOREd collating
 2202 elements from the start of the compare shall collate first. If both strings contain
 2203 a non-"IGNORE"d character in the same relative position, the collating values
 2204 assigned to the elements shall determine the ordering. In case of equality,
 2205 subsequent non-IGNOREd characters shall be considered in the same manner.
 2206

2207 The directives "forward" and "backward", and "backward" and "position", are mutually
 2208 exclusive at a given level.

2209 Examples:
 2210 order_start forward;backward
 2211 order_start <CYRILLIC>;forward;forward
 2212

2213
 2214 If no operands are specified, a single forward operand shall be assumed.
 2215

2216 2217 4.4.9 "order_end" keyword

2218
 2219 The collating order entries shall be terminated with an order_end keyword.
 2220

2221 4.4.10 "reorder-after" keyword

2222
 2223 The "reorder-after" keyword shall be used to specify a modification to a copied collation
 2224 specification of an existing FDCC-set. There can be more than one "reorder-after"
 2225 statement in a collating specification. The syntax shall be:
 2226

2227 "reorder-after %s\n",<collating-symbol>
 2228

2229 The <collating-symbol> operand shall be a symbolic name, enclosed between angle
 2230 brackets, and shall be present in the source FDCC-set copied via the "copy" keyword.
 2231 The "reorder-after" statement is followed by one or more collation statements as described
 2232 in the "Collating Order" clause (4.4.5), with the exception that the ellipsis symbol (...) shall not be used.
 2233

2234
 2235 Each collation statement reassigns character collation values and collation weights to
 2236 collating elements existing in the copied collation specification, by removing the collating

2237 statement from the copied specification, and inserting the collating element in the collating
 2238 sequence with the new collation weights after the preceding collating element of the
 2239 "reorder-after" specification, the first collating element in the collation sequence being the
 2240 <collating-symbol> specified on the "reorder-after" statement.
 2241

2242 A "reorder-after" specification is terminated by another "reorder-after" specification or the
 2243 "reorder-end" statement.
 2244

2245 4.4.10.1 Example of "reorder-after"

```

2246 reorder-after <y8>
2247 <U:>      <Y;>;<U:>;<CAPITAL>
2248 <u:>      <Y;>;<U:>;<SMALL>
2249 reorder-after <z8>
2250 <AE>      <AE;>;<NONE>;<CAPITAL>
2251 <ae>      <AE;>;<NONE>;<SMALL>
2252 <A:>      <AE;>;<DIAERESIS>;<CAPITAL>
2253 <a:>      <AE;>;<DIAERESIS>;<SMALL>
2254 <O/>      <O/>;<NONE>;<CAPITAL>
2255 <o/>      <O/>;<NONE>;<SMALL>
2256 <AA>      <AA;>;<NONE>;<CAPITAL>
2257 <aa>      <AA;>;<NONE>;<SMALL>
2258 reorder-end
  
```

2259 The example is interpreted as follows (using the "i18nrep" repertoire map):
 2260

- 2261 1. The collating element <U:> is removed from the copied collating sequence and inserted after <y8> in the
 2262 collating sequence with the new weights. The collating element <u:> is removed from the copied collating
 2263 sequence and inserted in the resulting collation sequence after <U:> with the new weights. <y8> is used to
 2264 indicate the last entry of the <y> letters.
 2265
- 2266 2. The second "reorder-after" statement terminates the first list of reordering collation identifier entries, and
 2267 initiates a second list, rearranging the order and weights for the <AE>, <ae>, <A:>, <a:>, <O/>, and <o/>
 2268 collating elements after the <z8> collating symbol in the copied specification. <z8> is used to indicate the
 2269 last entry of the <z> letters.
 2270
- 2271 3. The "reorder-end" statement terminates the second list of reordering entries.
 2272

- 2273 4. Thus for the original sequence
 2274

2275 ... (U u Ü ü) V v W w X x Y y Z z
 2276

2277 this example reordering gives
 2278

2279 ... U u V v W w X x (Y y Ü ü) Z z (Æ æ Ä ä) Ø ø Å å
 2280

2281 where the parenthesis indicate ordering with the same weight on the first level for multiple upper/lowercase
 2282 pairs.
 2283
 2284
 2285

2286 4.4.11 "reorder-end" keyword

2287 The "reorder-end" keyword shall specify the end of a list of collating statements, initiated
 2288 by the "reorder-after" keyword.
 2289

2290 4.4.12 "reorder-sections-after" keyword

2291 The "reorder-sections-after" keyword shall be used to specify a modification to a copied
 2292
 2293

2294 collation specification of an existing FDCC-set. The "reorder-sections-after" statement is
 2295 followed by one or more statements consisting of section reordering statements.
 2296

2297 4.4.12.1 section reordering statements

2298
 2299 The section reordering statements rearranges the set of collating entries and changes
 2300 sorting rules for the set of collating entries identified by a section symbol in a preceding
 2301 "order_start" statement. Each section reorder statement has the syntax:
 2302

```
2303 "%s %s;...%s\n", <section-symbol>, <sort-rules>, <sort-rules> ...
```

2304
 2305 The <section-symbol> identifies the set of collating entries, and shall be defined via a
 2306 "section-symbol" keyword.
 2307

2308 The <sort-rules> are as described for the "order_start" keyword. Specified <sort-rules>
 2309 replace the specification for the ordering of the section given on the "order_start"
 2310 statement identified by the <section-symbol>. The <sort-rules> are optional and <sort-
 2311 rules> not to be changed may be given by empty specifications.
 2312

2313 The order of the section reordering statements rearranges the assignment of collation
 2314 entries for the sets of collation entries identified by the <section-symbols> to the order
 2315 that the <section-symbols> occur after the "reorder-sections-after" statement.
 2316

2317 The section reordering statements are terminated by a "reorder-sections-end" statement.
 2318

2319 4.4.12.2 Example of section reordering

```
2320 copy "i18n"  

  2321 reorder-sections-after <DIGITS>  

  2322 <ARABIC>  

  2323 <LATIN> forward;backward;forward;forward,position  

  2324 reorder-sections-end  

  2325  

  2326
```

2327 This example is interpreted as follows: The LC_COLLATE category of the "i18n" FDCC-set is copied. Then a
 2328 reordering of all collating statements for the sections <ARABIC> and <LATIN> is done, leaving the rest of the
 2329 sections as they were in the "i18n" FDCC-set. The <ARABIC> section is placed immediately after the <DIGITS>
 2330 section, and the <LATIN> section immediately following the <ARABIC> section. The ordering rules are kept as
 2331 they were in the "i18n" FDCC-set, while the <LATIN> section gets new ordering rules as indicated. The
 2332 "reorder-sections-end" keyword terminates the section reordering statements.
 2333

2334 4.4.13 "reorder-sections-end" keyword

2335
 2336 The "reorder-sections-end" keyword shall specify the end of a list of section symbols,
 2337 initiated by the "reorder-sections-after" keyword.
 2338

2339 4.4.14 "i18n" LC_COLLATE category

2340
 2341 The "i18n" LC_COLLATE category is defined as the following, which includes the
 2342 tailorable template in ISO/IEC 14651.
 2343

```
2344 LC_COLLATE  

  2345  

  2346 % Case collating symbols  

  2347 collating-symbol <RES-1>  

  2348 collating-symbol <BLK>  

  2349 collating-symbol <MIN>          % SMALL  

  2350 collating-symbol <WIDE>         % WIDE  

  2351 collating-symbol <COMPAT>
```

2352	collating-symbol		
2353	collating-symbol	<CIRCLE>	
2354	collating-symbol	<RES-2>	
2355	collating-symbol	<CAP>	% CAPITAL
2356	collating-symbol	<WIDECAP>	
2357	collating-symbol	<COMPATCAP>	
2358	collating-symbol	<FONTCAP>	
2359	collating-symbol	<CIRCLECAP>	
2360	collating-symbol	<HIRA-SMALL>	
2361	collating-symbol	<HIRA>	
2362	collating-symbol	<SMALL>	
2363	collating-symbol	<SMALL-NARROW>	
2364	collating-symbol	<KATA>	
2365	collating-symbol	<NARROW>	
2366	collating-symbol	<CIRCLE-KATA>	
2367	collating-symbol	<MNN>	
2368	collating-symbol	<MNS>	
2369	collating-symbol	<VERTICAL>	
2370	%	Arabic forms	
2371	collating-symbol	<AINI>	
2372	collating-symbol	<AMED>	
2373	collating-symbol	<AFIN>	
2374	collating-symbol	<AISO>	
2375	%		
2376	collating-symbol	<NOBREAK>	
2377	collating-symbol	<SQUARED>	
2378	collating-symbol	<SQUARED CAP>	
2379	collating-symbol	<FRACTION>	
2380	collating-symbol	<BLANK>	
2381	collating-symbol	<CAPITAL-SMALL>	
2382	collating-symbol	<SMALL-CAPITAL>	
2383	collating-symbol	<BOTH>	
2384	%	accents	
2385	collating-symbol	<LOWLINE>	% LOW LINE
2386	collating-symbol	<MACRO>	% MACRON
2387	collating-symbol	<OBLIK>	% STROKE
2388	collating-symbol	<AIGUT>	% ACUTE ACCENT
2389	collating-symbol	<GRAVE>	% GRAVE ACCENT
2390	collating-symbol	<BREVE>	% BREVE
2391	collating-symbol	<CIRCF>	% CIRCUMFLEX ACCENT
2392	collating-symbol	<CARON>	% CARON
2393	collating-symbol	<CRCL>	% RING ABOVE
2394	collating-symbol	<TREMA>	% DIAERESIS
2395	collating-symbol	<2AIGU>	% DOUBLE ACUTE ACCENT
2396	collating-symbol	<TILDE>	% TILDE
2397	collating-symbol	<POINT>	% DOT ABOVE
2398	collating-symbol	<CEDIL>	% CEDILLA
2399	collating-symbol	<OGONK>	% OGONEK
2400	collating-symbol	<OVERLINE>	% OVERLINE
2401	collating-symbol	<CROOK>	% HOOK ABOVE
2402	collating-symbol	<TONOS>	% VERTICAL LINE ABOVE
2403	collating-symbol	<D030E>	% DOUBLE VERTICAL LINE ABOVE
2404	collating-symbol	<2GRAV>	% DOUBLE GRAVE ACCENT
2405	collating-symbol	<D0310>	% CANDRABINDU
2406	collating-symbol	<BREVR>	% INVERTED BREVE
2407	collating-symbol	<D0312>	% TURNED COMMA ABOVE
2408	collating-symbol	<PSILI>	% COMMA ABOVE
2409	collating-symbol	<DASIA>	% REVERSED COMMA ABOVE
2410	collating-symbol	<D0315>	% COMMA ABOVE RIGHT
2411	collating-symbol	<D0316>	% GRAVE ACCENT BELOW
2412	collating-symbol	<D0317>	% ACUTE ACCENT BELOW
2413	collating-symbol	<D0318>	% LEFT TACK BELOW
2414	collating-symbol	<D0319>	% RIGHT TACK BELOW
2415	collating-symbol	<D031A>	% LEFT ANGLE ABOVE
2416	collating-symbol	<HORNU>	% HORN
2417	collating-symbol	<D031C>	% LEFT HALF RING BELOW
2418	collating-symbol	<D031D>	% UP TACK BELOW
2419	collating-symbol	<D031E>	% DOWN TACK BELOW
2420	collating-symbol	<D031F>	% PLUS SIGN BELOW
2421	collating-symbol	<D0320>	% MINUS SIGN BELOW
2422	collating-symbol	<PALCR>	% PALATALIZED HOOK BELOW
2423	collating-symbol	<RETCR>	% RETROFLEX HOOK BELOW
2424	collating-symbol	<POINS>	% DOT BELOW
2425	collating-symbol	<TREMS>	% DIAERESIS BELOW
2426	collating-symbol	<CRCLS>	% RING BELOW
2427	collating-symbol	<COMMS>	% COMMA BELOW
2428	collating-symbol	<D0329>	% VERTICAL LINE BELOW
2429	collating-symbol	<D032A>	% BRIDGE BELOW
2430	collating-symbol	<D032B>	% INVERTED DOUBLE ARCH BELOW

2431	collating-symbol	<D032C>	% CARON BELOW
2432	collating-symbol	<CIRCS>	% CIRCUMFLEX ACCENT BELOW
2433	collating-symbol	<BREVS>	% BREVE BELOW
2434	collating-symbol	<D032F>	% INVERTED BREVE BELOW
2435	collating-symbol	<TILDS>	% TILDE BELOW
2436	collating-symbol	<MACRS>	% MACRON BELOW
2437	collating-symbol	<D0333>	% DOUBLE LOW LINE
2438	collating-symbol	<TILDX>	% TILDE OVERLAY
2439	collating-symbol	<BARRE>	% SHORT STROKE OVERLAY
2440	collating-symbol	<D0336>	% LONG STROKE OVERLAY
2441	collating-symbol	<D0337>	% SHORT SOLIDUS OVERLAY
2442	collating-symbol	<CRCL2>	% RIGHT HALF RING BELOW
2443	collating-symbol	<D033A>	% INVERTED BRIDGE BELOW
2444	collating-symbol	<D033B>	% SQUARE BELOW
2445	collating-symbol	<D033C>	% SEAGULL BELOW
2446	collating-symbol	<D033D>	% X ABOVE
2447	collating-symbol	<D033E>	% VERTICAL TILDE
2448	collating-symbol	<D033F>	% DOUBLE OVERLINE
2449	collating-symbol	<PERIS>	% GREEK PERISPOMENI
2450	collating-symbol	<YPOGE>	% GREEK YPOGEGRAMMENI
2451	collating-symbol	<D0360>	% DOUBLE TILDE
2452	collating-symbol	<D0361>	% DOUBLE INVERTED BREVE
2453	collating-symbol	<DFE20>	% LIGATURE LEFT HALF
2454	collating-symbol	<DFE21>	% LIGATURE RIGHT HALF
2455	collating-symbol	<DFE22>	% DOUBLE TILDE LEFT HALF
2456	collating-symbol	<DFE23>	% DOUBLE TILDE RIGHT HALF
2457	collating-symbol	<D0483>	% CYRILLIC TITLO
2458	collating-symbol	<D0484>	% CYRILLIC PALATALIZATION
2459	collating-symbol	<D0485>	% CYRILLIC DASIA PNEUMATA
2460	collating-symbol	<D0486>	% CYRILLIC PSILI PNEUMATA
2461	collating-symbol	<SHEVA>	% HEBREW POINT SHEVA
2462	collating-symbol	<HTFSG>	% HEBREW POINT HATAF SEGOL
2463	collating-symbol	<HTFPT>	% HEBREW POINT HATAF PATAH
2464	collating-symbol	<HTFQM>	% HEBREW POINT HATAF QAMATS
2465	collating-symbol	<HIRIQ>	% HEBREW POINT HIRIQ
2466	collating-symbol	<TSERE>	% HEBREW POINT TSERE
2467	collating-symbol	<SEGOL>	% HEBREW POINT SEGOL
2468	collating-symbol	<PATAH>	% HEBREW POINT PATAH
2469	collating-symbol	<QAMAT>	% HEBREW POINT QAMATS
2470	collating-symbol	<HOLAM>	% HEBREW POINT HOLAM
2471	collating-symbol	<QUBUT>	% HEBREW POINT QUBUTS
2472	collating-symbol	<DAGES>	% HEBREW POINT DAGESH OR MAPIQ
2473	collating-symbol	<RAPHE>	% HEBREW POINT RAPE
2474	collating-symbol	<SHINP>	% HEBREW POINT SHIN DOT
2475	collating-symbol	<SINPT>	% HEBREW POINT SIN DOT
2476	collating-symbol	<VARIKA>	% HEBREW POINT JUDEO-SPANISH VARIKA
2477	collating-symbol	<FATHATAN>	% ARABIC FATHATAN
2478	collating-symbol	<DAMMATAN>	% ARABIC DAMMATAN
2479	collating-symbol	<KASRATAN>	% ARABIC KASRATAN
2480	collating-symbol	<FATHA>	% ARABIC FATHA
2481	collating-symbol	<DAMMA>	% ARABIC DAMMA
2482	collating-symbol	<KASRA>	% ARABIC KASRA
2483	collating-symbol	<SHADDA>	% ARABIC SHADDA
2484	collating-symbol	<SUKUN>	% ARABIC SUKUN
2485	collating-symbol	<SUPERALEF>	% ARABIC LETTER SUPERScript ALEF
2486	collating-symbol	<D06D6>	% ARABIC SMALL HIGH LIGATURE SAD WITH LAM WITH ALEF MAKSURA
2487	collating-symbol	<D06D7>	% ARABIC SMALL HIGH LIGATURE QAF WITH LAM WITH ALEF MAKSURA
2488	collating-symbol	<D06D8>	% ARABIC SMALL HIGH MEEM INITIAL FORM
2489	collating-symbol	<D06D9>	% ARABIC SMALL HIGH LAM ALEF
2490	collating-symbol	<D06DA>	% ARABIC SMALL HIGH JEEM
2491	collating-symbol	<D06DB>	% ARABIC SMALL HIGH THREE DOTS
2492	collating-symbol	<D06DC>	% ARABIC SMALL HIGH SEEN
2493	collating-symbol	<D06E1>	% ARABIC SMALL HIGH DOTLESS HEAD OF KHAH
2494	collating-symbol	<D06E2>	% ARABIC SMALL HIGH MEEM ISOLATED FORM
2495	collating-symbol	<D06E3>	% ARABIC SMALL LOW SEEN
2496	collating-symbol	<AMADD>	% ARABIC SMALL HIGH MADDA
2497	collating-symbol	<D06E7>	% ARABIC SMALL HIGH YEH
2498	collating-symbol	<D06E8>	% ARABIC SMALL HIGH NOON
2499	collating-symbol	<D06ED>	% ARABIC SMALL LOW MEEM
2500	collating-symbol	<D093C>	% DEVANAGARI SIGN NUKTA
2501	collating-symbol	<D0951>	% DEVANAGARI STRESS SIGN UDATTA
2502	collating-symbol	<D0952>	% DEVANAGARI STRESS SIGN ANUDATTA
2503	collating-symbol	<D0953>	% DEVANAGARI GRAVE ACCENT
2504	collating-symbol	<D0954>	% DEVANAGARI ACUTE ACCENT
2505	collating-symbol	<D09BC>	% BENGALI SIGN NUKTA
2506	collating-symbol	<D0A3C>	% GURMUKHI SIGN NUKTA
2507	collating-symbol	<D0ABC>	% GUJARATI SIGN NUKTA
2508	collating-symbol	<D0B3C>	% ORIYA SIGN NUKTA
2509	collating-symbol	<D0E48>	% THAI CHARACTER MAI EK

2510	collating-symbol	<D0E49>	% THAI CHARACTER MAI THO
2511	collating-symbol	<D0E4A>	% THAI CHARACTER MAI TRI
2512	collating-symbol	<D0E4B>	% THAI CHARACTER MAI CHATTAWA
2513	collating-symbol	<D0EC8>	% LAO TONE MAI EK
2514	collating-symbol	<D0EC9>	% LAO TONE MAI THO
2515	collating-symbol	<D0ECA>	% LAO TONE MAI TI
2516	collating-symbol	<D0ECB>	% LAO TONE MAI CATAWA
2517	collating-symbol	<D0F39>	% TIBETAN MARK TSA -PHRU
2518	collating-symbol	<D0F3E>	% TIBETAN SIGN YAR TSHES
2519	collating-symbol	<D0F3F>	% TIBETAN SIGN MAR TSHES
2520	collating-symbol	<D302A>	% IDEOGRAPHIC LEVEL TONE MARK
2521	collating-symbol	<D302B>	% IDEOGRAPHIC RISING TONE MARK
2522	collating-symbol	<D302C>	% IDEOGRAPHIC DEPARTING TONE MARK
2523	collating-symbol	<D302D>	% IDEOGRAPHIC ENTERING TONE MARK
2524	collating-symbol	<D302E>	% HANGUL SINGLE DOT TONE MARK
2525	collating-symbol	<D302F>	% HANGUL DOUBLE DOT TONE MARK
2526	collating-symbol	<KNVCE>	% KATAKANA-HIRAGANA VOICED SOUND MARK
2527	collating-symbol	<KNSMV>	% KATAKANA-HIRAGANA SEMI-VOICED SOUND MARK
2528	collating-symbol	<D20D0>	% LEFT HARPOON ABOVE
2529	collating-symbol	<D20D1>	% RIGHT HARPOON ABOVE
2530	collating-symbol	<D20D2>	% LONG VERTICAL LINE OVERLAY
2531	collating-symbol	<D20D3>	% SHORT VERTICAL LINE OVERLAY
2532	collating-symbol	<D20D4>	% ANTICLOCKWISE ARROW ABOVE
2533	collating-symbol	<D20D5>	% CLOCKWISE ARROW ABOVE
2534	collating-symbol	<D20D6>	% LEFT ARROW ABOVE
2535	collating-symbol	<D20D7>	% RIGHT ARROW ABOVE
2536	collating-symbol	<D20D8>	% RING OVERLAY
2537	collating-symbol	<D20D9>	% CLOCKWISE RING OVERLAY
2538	collating-symbol	<D20DA>	% ANTICLOCKWISE RING OVERLAY
2539	collating-symbol	<D20DB>	% THREE DOTS ABOVE
2540	collating-symbol	<D20DC>	% FOUR DOTS ABOVE
2541	collating-symbol	<D20DD>	% ENCLOSING CIRCLE
2542	collating-symbol	<D20DE>	% ENCLOSING SQUARE
2543	collating-symbol	<D20DF>	% ENCLOSING DIAMOND
2544	collating-symbol	<D20E0>	% ENCLOSING CIRCLE BACKSLASH
2545	collating-symbol	<D20E1>	% LEFT RIGHT ARROW ABOVE
2546	collating-symbol	<NEGATIVE>	
2547	collating-symbol	<SANSSEERIF>	
2548	collating-symbol	<NEGSANSSEERIF>	
2549	collating-symbol	<ARABIC>	
2550	collating-symbol	<EXTARABIC>	
2551	collating-symbol	<NAGAR>	
2552	collating-symbol	<BENGL>	
2553	collating-symbol	<BENGALINUMERATOR>	
2554	collating-symbol	<GURMU>	
2555	collating-symbol	<GUJAR>	
2556	collating-symbol	<ORIYA>	
2557	collating-symbol	<TAMIL>	
2558	collating-symbol	<TELGU>	
2559	collating-symbol	<KNNDA>	
2560	collating-symbol	<MALAY>	
2561	collating-symbol	<SINHALA>	
2562	collating-symbol	<THAII>	
2563	collating-symbol	<LAAOO>	
2564	collating-symbol	<BODKA>	
2565	collating-symbol	<CJKVS>	
2566	collating-symbol	<S0200>..<S1100>	% 0x0200..0x1100
2567			
2568	collating-symbol	<S4E00>..<S9FA5>	% Symbols for Han
2569			
2570	collating-symbol	<SAC00>..<SD7A3>	% Symbols for Hangul
2571			
2572	collating-symbol	<SFA0E>..<SFA29>	% Symbols for Compatibility Han
2573			
2574			% equivalences
2575	symbol-equivalence	<NONE>	<BLANK>
2576	symbol-equivalence	<CAPITAL>	<CAP>
2577	symbol-equivalence	<MACRON>	<MACRO>
2578	symbol-equivalence	<STROKE>	<OBLIK>
2579	symbol-equivalence	<ACUTE>	<AIGUT>
2580	symbol-equivalence	<CIRCUMFLEX>	<CIRCF>
2581	symbol-equivalence	<RING>	<CRCL>
2582	symbol-equivalence	<DIAERESIS>	<TREMA>
2583	symbol-equivalence	<DOT>	<POINT>
2584	symbol-equivalence	<CEDILLA>	<CEDIL>
2585	symbol-equivalence	<OGONEK>	<OGONK>
2586	symbol-equivalence	<HOOK>	<CROOK>
2587	symbol-equivalence	<HORN>	<HORNU>
2588	symbol-equivalence	<DOT-BELOW>	<POINS>

```

2589 order_start <Latin>;forward;backward;forward;forward,position
2590
2591 % Copy the template from ISO/IEC 14651
2592 copy "iso14651t1"
2593
2594 order_end
2595
2596 END LC_COLLATE
2597

```

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:

2608	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.
2611	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.
2621	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.
2626	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 application. If the currency is not the first valid currency for the period in question, the first integer is for multiplying the first valid currency, and the second for dividing this result to get the amount in the current currency. The currency to be the current currency is selected by the application from the date applicable and the currency number (first, second, third etc valid currency at that date); and whether domestic or international formatting is used is also determined by the application. Each pair of integers are separated by a <slash>. The default value is "1/100". This keyword is optional.
2641	int_curr_symbol	One or more strings separated by semicolons that shall be used as the international currency symbols. Each operand shall be a four character string, with the first three characters

2643		containing the alphabetic international currency symbol in
2644		accordance with those specified in ISO 4217, <i>Codes for the</i>
2645		<i>representation of currencies and funds</i> . The fourth character
2646		shall be the character used to separate the international
2647		currency symbol from the monetary quantity. The keyword
2648		shall be specified, unless the "copy" keyword is used.
2649	currency_symbol	One or more strings separated by semicolons that shall be
2650		used as the local currency symbol.
2651	mon_decimal_point	The operand is a string containing the symbol that shall be
2652		used as the decimal delimiter in monetary formatted
2653		quantities. In contexts where other standards limit the
2654		"mon_decimal_point" to a single byte, the result of
2655		specifying a multibyte operand is unspecified. The keyword
2656		shall be specified, unless the "copy" keyword is used.
2657	mon_thousands_sep	The operand is a string containing the symbol that shall be
2658		used as a separator for groups of digits to the left of the
2659		decimal delimiter in formatted monetary quantities. In
2660		contexts where other standards limit the
2661		"mon_thousands_sep" to a single byte, the result of speci-
2662		fyng a multibyte operand is unspecified. The keyword shall
2663		be specified, unless the "copy" keyword is used.
2664	mon_grouping	Define the size of each group of digits in formatted
2665		monetary quantities. The operand is a sequence of integers
2666		separated by semicolons. Each integer specifies the number
2667		of digits in each group, with the initial integer defining the
2668		size of the group immediately preceding the decimal
2669		delimiter, and the following integers defining the preceding
2670		groups. If the last integer is not -1, then the size of the
2671		previous group (if any) shall be repeatedly used for the
2672		remainder of the digits. If the last integer is -1, then no
2673		further grouping shall be performed. The keyword shall be
2674		specified, unless the "copy" keyword is used.
2675	positive_sign	A string that shall be used to indicate a nonnegative-valued
2676		formatted monetary quantity. The keyword shall be specified,
2677		unless the "copy" keyword is used.
2678	negative_sign	A string that shall be used to indicate a negative-valued
2679		formatted monetary quantity. The keyword shall be specified,
2680		unless the "copy" keyword is used.
2681	int_frac_digits	One or more integers separated by semicolons, representing
2682		the number of fractional digits (those to the right of the
2683		decimal delimiter) to be written in a formatted monetary
2684		quantity using int_curr_symbol. The keyword shall be
2685		specified, unless the "copy" keyword is used.
2686	frac_digits	One or more integers separated by semicolons, representing
2687		the number of fractional digits (those to the right of the
2688		decimal delimiter) to be written in a formatted monetary
2689		quantity using "currency_symbol". The keyword shall be
2690		specified, unless the "copy" keyword is used.
2691	p_cs_precedes	One or more integers separated by semicolons, set to 1 if the
2692		"currency_symbol" precedes the value for a nonnegative

2693		formatted monetary quantity, and set to 0 if the symbol
2694		succeeds the value. The keyword shall be specified, unless
2695		the "copy" keyword is used.
2696	p_sep_by_space	One or more integers separated by semicolons, set to 0 if no
2697		space separates the "currency_symbol" from the value for a
2698		nonnegative formatted monetary quantity, set to 1 if a space
2699		separates the symbol from the value, and set to 2 if a space
2700		separates the symbol and the sign string, if adjacent. The
2701		keyword shall be specified, unless the "copy" keyword is
2702		used.
2703	n_cs_precedes	One or more integers separated by semicolons, set to 1 if the
2704		"currency_symbol" precedes the value for a negative
2705		formatted monetary quantity, and set to 0 if the symbol
2706		succeeds the value. The keyword shall be specified, unless
2707		the "copy" keyword is used.
2708	n_sep_by_space	One or more integers separated by semicolons, set to 0 if no
2709		space separates the "currency_symbol" from the value for a
2710		negative formatted monetary quantity, set to 1 if a space
2711		separates the symbol from the value, and set to 2 if a space
2712		separates the symbol and the sign string, if adjacent. The
2713		keyword shall be specified, unless the "copy" keyword is
2714		used.
2715	int_p_cs_precedes	One or more integers separated by semicolons; set to 1 if the
2716		"int_curr_symbol" precedes the value for a nonnegative
2717		formatted monetary quantity, and set to 0 if the symbol
2718		succeeds the value. If not specified, the value of
2719		"p_cs_precedes" is taken.
2720	int_p_sep_by_space	One or more integers separated by semicolons; set to 0 if no
2721		space separates the "int_curr_symbol" from the value for a
2722		nonnegative formatted monetary quantity, set to 1 if a space
2723		separates the symbol from the value, and set to 2 if a space
2724		separates the symbol and the sign string, if adjacent. If not
2725		specified, the value of "p_sep_by_space" is taken.
2726	int_n_cs_precedes	One or more integers separated by semicolons; set to 1 if the
2727		"int_curr_symbol" precedes the value for a negative
2728		formatted monetary quantity, and set to 0 if the symbol
2729		succeeds the value. If not specified, the value of
2730		"n_cs_precedes" is taken.
2731	int_n_sep_by_space	One or more integers separated by semicolons; set to 0 if no
2732		space separates the "int_curr_symbol" from the value for a
2733		negative formatted monetary quantity, set to 1 if a space
2734		separates the symbol from the value, and set to 2 if a space
2735		separates the symbol and the sign string, if adjacent. If not
2736		specified, the value of "n_sep_by_space" is taken.
2737	p_sign_posn	One or more integers separated by semicolons, set to a value
2738		indicating the positioning of the "positive_sign" for a
2739		nonnegative formatted monetary quantity using the
2740		"currency_symbol". The following integer values shall be
2741		defined:
2742		
2743		0 Parentheses enclose the quantity and the

2744		"currency_symbol".
2745	1	The sign string precedes the quantity and the
2746		"currency_symbol".
2747	2	The sign string succeeds the quantity and the
2748		"currency_symbol".
2749	3	The sign string immediately precedes the
2750		"currency_symbol".
2751	4	The sign string immediately succeeds the
2752		"currency_symbol".
2753		The keyword shall be specified, unless the "copy" keyword
2754		is used.
2755		
2756	n_sign_posn	One or more integers separated by semicolons, set to a value
2757		indicating the positioning of the "negative_sign" for a
2758		negative formatted monetary quantity using the
2759		"currency_symbol". The following integer values shall be
2760		defined:
2761		
2762	0	Parentheses enclose the quantity and the
2763		"currency_symbol".
2764	1	The sign string precedes the quantity and the
2765		"currency_symbol".
2766	2	The sign string succeeds the quantity and the
2767		"currency_symbol".
2768	3	The sign string immediately precedes the
2769		"currency_symbol".
2770	4	The sign string immediately succeeds the
2771		"currency_symbol".
2772		The keyword shall be specified, unless the "copy" keyword
2773		is used.
2774		
2775	int_p_sign_posn	One or more integers separated by semicolons, set to a value
2776		indicating the positioning of the "positive_sign" for a
2777		nonnegative formatted international monetary quantity. The
2778		following integer values shall be defined:
2779		
2780	0	Parentheses enclose the quantity and the
2781		"int_curr_symbol".
2782	1	The sign string precedes the quantity and the
2783		"int_curr_symbol".
2784	2	The sign string succeeds the quantity and the
2785		"int_curr_symbol".
2786	3	The sign string immediately precedes the
2787		"int_curr_symbol".
2788	4	The sign string immediately succeeds the
2789		"int_curr_symbol".
2790		If no "int_p_sign_posn" is present the value of the
2791		"p_sign_posn" is taken.
2792		
2793	int_n_sign_posn	One or more integers separated by semicolons, set to a value

2794 indicating the positioning of the "negative_sign" for a
 2795 negative formatted international monetary quantity. The
 2796 following integer values shall be defined:

- 2797
 2798 0 Parentheses enclose the quantity and the
 2799 "int_curr_symbol".
 2800 1 The sign string precedes the quantity and the
 2801 "int_curr_symbol".
 2802 2 The sign string succeeds the quantity and the
 2803 "int_curr_symbol".
 2804 3 The sign string immediately precedes the
 2805 "int_curr_symbol".
 2806 4 The sign string immediately succeeds the
 2807 "int_curr_symbol".
 2808 If no "int_n_sign_posn" is present the value of the
 2809 "n_sign_posn" is taken.
 2810

2811 The "i18n" FDCC-set is defined as follows for the LC_MONETARY category.

```

2812 LC_MONETARY
2813 % This is the 14652 i18n fdcc-set definition for
2814 % the LC_MONETARY category.
2815 %
2816 int_curr_symbol      ""
2817 currency_symbol     ""
2818 mon_decimal_point   "<,>"
2819 mon_thousands_sep  ""
2820 mon_grouping        -1
2821 positive_sign       ""
2822 negative_sign       ""
2823 int_frac_digits     -1
2824 frac_digits         -1
2825 p_cs_precedes       -1
2826 p_sep_by_space     -1
2827 n_cs_precedes       -1
2828 n_sep_by_space     -1
2829 p_sign_posn        -1
2830 n_sign_posn        -1
2831 %
2832 END LC_MONETARY
  
```

2835 4.6 LC_NUMERIC

2837
 2838 The LC_NUMERIC category defines the rules and symbols that shall be used to format
 2839 nonmonetary numeric information. The operands are strings. For some keywords, the
 2840 strings only can contain integers. Keywords that are not provided, string values set to the
 2841 empty string (""), or integer keywords set to -1, shall be used to indicate that the value is
 2842 unspecified. The following keywords shall be defined:

- 2843
 2844 **copy** Specify the name of an existing FDCC-set to be used as the
 2845 source for the definition of this category. If this keyword is
 2846 specified, no other keyword shall be specified.
 2847 **decimal_point** The operand is a string containing the symbol that shall be used
 2848 as the decimal delimiter in numeric, nonmonetary formatted
 2849 quantities. This keyword cannot be omitted and cannot be set to
 2850 the empty string. In contexts where other standards limit the
 2851 decimal point to a single byte, the result of specifying a mul-
 2852 titype operand is unspecified.

2853 **thousands_sep** The operand is a string containing the symbol that shall be used
 2854 as a separator for groups of digits to the left of the decimal
 2855 delimiter in numeric, nonmonetary formatted monetary quan-
 2856 tities. In contexts where other standards limit the
 2857 "thousands_sep" to a single byte, the result of specifying a
 2858 multibyte operand is unspecified.

2859 **grouping** Define the size of each group of digits in formatted non-
 2860 monetary quantities. The operand is a sequence of integers
 2861 separated by semicolons. Each integer specifies the number of
 2862 digits in each group, with the initial integer defining the size of
 2863 the group immediately preceding the decimal delimiter, and the
 2864 following integers defining the preceding groups. If the last
 2865 integer is not -1, then the size of the previous group (if any)
 2866 shall be repeatedly used for the remainder of the digits. If the
 2867 last integer is -1, then no further grouping shall be performed.

2868
 2869 The "i18n" FDCC-set is for the LC_NUMERIC category:

```

2870 LC_NUMERIC
2871 % This is the 14652 i18n fdcc-set definition for
2872 % the LC_NUMERIC category.
2873 %
2874 decimal_point "<,>"
2875 thousands_sep ""
2876 grouping -1
2877 %
2878 END LC_NUMERIC
2879
2880
  
```

2881 4.7 LC_TIME

2882
 2883
 2884 The LC_TIME category defines the rules and symbols that shall be used to format date
 2885 and time information. The following keywords shall be defined:

2886
 2887 **copy** Specify the name of an existing FDCC-set to be used as the source
 2888 for the definition of this category. If this keyword is specified, no
 2889 other keyword shall be specified.

2890 **abday** Define the abbreviated weekday names for calendar systems with
 2891 weeks of constant length, to be referenced by the %a field descriptor.
 2892 The length of the week and a gregorian date for the first weekday is
 2893 defined by the "week" keyword. The operand shall consist of
 2894 semicolon-separated strings. The first string shall be the abbreviated
 2895 name of the day corresponding to the first day of the week (default
 2896 Sunday), the second the abbreviated name of the day corresponding
 2897 to the second day of the week (default Monday), and so on.

2898 **day** Define the full weekday names for calendar systems with weeks of
 2899 constant length, to be referenced by the %A field descriptor. The
 2900 length of the week and a gregorian date for the first weekday is
 2901 defined by the "week" keyword. The operand shall consist of
 2902 semicolon-separated strings. The first string shall be the full name of
 2903 the day corresponding to the first day of the week (default Sunday),
 2904 the second the full name of the day corresponding to the second day
 2905 of the week (default Monday), and so on.

2906 **week** Shall be used to define the number of days in a week, and which

2907		weekday is the first weekday (the first weekday has the value 1), and
2908		which week is to be considered the first in a year. The first operand
2909		is an integer specifying the number of days in the week. The second
2910		operand is an integer specifying the Gregorian date in the format
2911		YYYYMMDD with a leading <hyphen-minus> if before Christ. The
2912		third operand is an integer specifying the weekday number to be
2913		contained in the first week of the year. If the keyword is not
2914		specified the values are taken as 7, 19971130 (a Sunday), and 7
2915		(Saturday), respectively. ISO 8601 conforming applications should
2916		use the values 7, 19971201 (a Monday), and 4 (Thursday),
2917		respectively. This keyword is optional.
2918	abmon	Define the abbreviated month names, to be referenced by the %b
2919		field descriptor. The operand shall consist of twelve or thirteen
2920		semicolon-separated strings. The first string shall be the abbreviated
2921		name of the first month of the year (January), the second the
2922		abbreviated name of the second month, and so on.
2923	mon	Define the full month names, to be referenced by the %B field
2924		descriptor. The operand shall consist of twelve or thirteen semicolon-
2925		separated strings. The first string shall be the full name of the first
2926		month of the year (January), the second the full name of the second
2927		month, and so on.
2928	d_t_fmt	Define the appropriate date and time representation, to be referenced
2929		by the %c field descriptor. The operand shall consist of a string, and
2930		can contain any combination of characters and field descriptors. In
2931		addition, the string can contain escape sequences defined in Table 3.
2932	d_fmt	Define the appropriate date representation, to be referenced by the
2933		%x field descriptor. The operand shall consist of a string, and can
2934		contain any combination of characters and field descriptors. In
2935		addition, the string can contain escape sequences defined in Table 3.
2936	t_fmt	Define the appropriate time representation, to be referenced by the
2937		%X field descriptor. The operand shall consist of a string, and can
2938		contain any combination of characters and field descriptors. In
2939		addition, the string can contain escape sequences defined in Table 3.
2940	am_pm	Define the appropriate representation of the ante meridiem and post
2941		meridiem strings, to be referenced by the %p field descriptor. The
2942		operand shall consist of two strings, separated by a semicolon. The
2943		first string shall represent the antemeridiem designation, the last
2944		string the postmeridiem designation. The keyword is optional. If
2945		unspecified, the %p field descriptor shall refer to the empty string.
2946	t_fmt_ampm	Define the appropriate time representation in the 12-hour clock
2947		format with "am_pm", to be referenced by the %r field descriptor.
2948		The operand shall consist of a string and can contain any
2949		combination of characters and field descriptors. If the string is empty,
2950		the 12-hour format is not supported in the FDCC-set.
2951		
2952	The following keywords are all optional	
2953		
2954	era	Shall be used to define alternate Eras, corresponding to the %E field
2955		descriptor modifier. The format of the operand is unspecified, but
2956		shall support the definition of the %EC and %Ey field descriptors,
2957		and may also define the "era_year" format (%EY).

2958	era_year	Shall be used to define the format of the year in alternate Era format, corresponding to the %EY field descriptor.
2959		
2960	era_d_fmt	Shall be used to define the format of the date in alternate Era notation, corresponding to the %Ex field descriptor.
2961		
2962	alt_digits	Shall be used to define alternate symbols for digits, corresponding to the %O field descriptor modifier. The operand shall consist of semicolon-separated strings. The first string shall be the alternate symbol corresponding with zero, the second string the symbol corresponding with one, and so on. Up to 100 alternate symbol strings can be specified. The %O modifier indicates that the string corresponding to the value specified via the field descriptor shall be used instead of the value.
2963		
2964		
2965		
2966		
2967		
2968		
2969		
2970	first_weekday	Shall be used to define the first day to be displayed, for example in a calendar display utility. The operand is an integer specifying the day number (1 = first) according to the information specified with the "day" keyword. The keyword may be omitted, and then the value 1 is taken, corresponding to Sunday for a week beginning Sunday, or to Monday for a week beginning Monday.
2971		
2972		
2973		
2974		
2975		
2976	first_workday	Shall be used to define the first workday as an integer according to the day numbering specified with the "week" keyword.
2977		
2978	cal_direction	Shall be used to define the direction of the display of dates, for example in a calendar display utility. The operand is an integer, and the following values are defined:
2979		
2980		
2981		1 left-right from top
2982		2 top-down from left
2983		3 right-left from top
2984		The keyword may be omitted, and then the value 1 is taken.
2985	timezone	Shall be used to define a set of timezones, each defined by a string. In the following the characters <, >, [and] are used as metacharacters. Only characters with a visible glyph from the portable character set may be used, except in the <std> and <dst> fields. The syntax of the string is:
2986		
2987		
2988		
2989		
2990		
2991		<std><offset><dst>[<offset>][,<rule>[,<rule>...]]
2992		
2993		where
2994		
2995		<std> and <dst> Indicates no less than three, nor more than 10
2996		characters that are the designation for the
2997		standard <std> or summer <dst> time zone.
2998		only <std> is required; if <dst> is missing, then
2999		summer time does not apply in this category.
3000		Upper- and lowercase letters are explicitly
3001		allowed. Any characters except a leading colon
3002		<:> or digits, the comma <,>, the minus <->, the plus <+>, and the null character are
3003		permitted to appear in these fields, but their
3004		meaning is unspecified.
3005		
3006		<offset> Indicates the value one must add to the local
3007		time to arrive at the Coordinated Universal

3008
3009
3010
3011
3012
3013
3014
3015
3016
3017
3018
3019
3020
3021
3022
3023
3024
3025
3026
3027
3028
3029
3030
3031
3032
3033
3034
3035
3036
3037
3038
3039
3040
3041
3042
3043
3044
3045
3046
3047
3048
3049
3050
3051
3052
3053
3054
3055
3056
3057
3058

Time. The <offset> has the form:

hh[:mm[:ss]]

The minutes (mm) and seconds (ss) are optional. The hour (hh) shall be required and may be a single digit. The <offset> following <std> shall be required. If no <offset> follows <dst>, summer time is assumed to be one hour ahead of standard time. One or more digits may be used; the value is always interpreted as a decimal number. The hour shall be between zero and 24, and the minutes (and seconds) - if present - shall be between zero and 59. If preceded by a "-", the time zone shall be east of the Prime Meridian; otherwise it shall be west of (which may be indicated by an optional preceding "+").

<rule>

Indicates when to change to and back from summer time. The <rule> has the form:

<date>[/<time>/<year>],<date>[/<time>/<year>]

where the first <date> describes when the change from standard time to summer time occurs, and the second <date> describes when the change back happens. Each <time> field describes when, in current local time, the change to the other time is made. The first <year> field defines the beginning of the validity of this rule, and the second <year> field defines the end of the validity of the rule. A number of rules may be given.

The format of <date> shall be one of the following:

J<n> The Julian day <n> (1 <= n <= 365) Leap years shall not be counted. That is, in all 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.

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

M<m>.<n>.<d> the <d>th day (0 <= d <= 7)

3059 of week <n> of month <m> (1
 3060 <= n <= 5, 1 <= m <= 12,
 3061 where week 5 means "the last
 3062 <d> day in month <m>"
 3063 which may occur in either the
 3064 fourth or fifth week). Week 1
 3065 is the first week in which the
 3066 <d>th day occurs. Day zero
 3067 and day seven is Sunday.

3068
 3069 The <time> has the same format as <offset>
 3070 except that no leading sign ("- or "+) shall be
 3071 allowed. The default, if <time> is not given,
 3072 shall be "02:00:00".
 3073

3074 The <year> has the format YYYY.
 3075

3076 NOTE: This way of specifying the timezone is compatible with the
 3077 format for the environment variable TZ described in Section 8.1.1 of
 3078 POSIX.1.
 3079

3080 **4.7.1 Date Field Descriptors**

3081
 3082 The LC_TIME category defines the interpretation of a number of field descriptors. The
 3083 field descriptors are also available in the definitions with the following LC_TIME
 3084 keywords: "d_t_fmt", "d_fmt", "t_fmt", "t_fmt_ampm", "era", and "era_d_fmt". A field
 3085 descriptor may not be used with the LC_TIME keywords defining it.
 3086

3087 **Table 3: Escape sequences for the date field**

3088		
3089	%a	FDCC-set's abbreviated weekday name.
3090	%A	FDCC-set's full weekday name.
3091	%b	FDCC-set's abbreviated month name.
3092	%B	FDCC-set's full month name.
3093	%c	FDCC-set's appropriate date and time representation.
3094	%C	Century (a year divided by 100 and truncated to integer) as decimal 3095 number (00-99).
3096	%d	Day of the month as a decimal number (01-31).
3097	%D	Date in the format mm/dd/yy.
3098	%e	Day of the month as a decimal number (1-31 in at two-digit field with 3099 leading <space> fill).
3100	%F	is replaced by the date in the format YYYY-MM-DD (ISO 8601 format)
3101	%h	A synonym for %b.
3102	%H	Hour (24-hour clock) as a decimal number (00-23).
3103	%I	Hour (12-hour clock) as a decimal number (01-12).
3104	%j	Day of the year as a decimal number (001-366).
3105	%m	Month as a decimal number (01-13).
3106	%M	Minute as a decimal number (00-59).
3107	%n	A <newline> character.
3108	%p	FDCC-set's equivalent of either AM or PM.

3109	%r	12-hour clock time (01-12) using the AM/PM notation.
3110	%S	Seconds as a decimal number (00-61).
3111	%t	A <tab> character.
3112	%T	24-hour clock time in the format HH:MM:SS.
3113	%u	Weekday as a decimal number (1(Monday)-7).
3114	%U	Week number of the year (Sunday as the first day of the week) as a
3115		decimal number (00-53). All days in a new year preceding the first
3116		Sunday shall be considered to be in week 0.
3117	%v	Week number of the year as a decimal number with two digits including a
3118		possible leading zero, according to "week" keyword.
3119	%V	Week of the year (Monday as the first day of the week) as a decimal
3120		number (01-53). The method for determining the week number shall be as
3121		specified by ISO 8601.
3122	%w	Weekday as a decimal number (0(Sunday)-6).
3123	%W	Week number of the year (Monday as the first day of the week) as a
3124		decimal number (00-53).
3125	%x	FDCC-set's appropriate date representation.
3126	%X	FDCC-set's appropriate time representation.
3127	%y	Year (offset from %C) as a decimal number (00-99).
3128	%Y	Year with century as a decimal number.
3129	%Z	Time-zone name, or no characters if no time zone is determinable.
3130	%%	A <percent-sign> character.

3131

3132 **4.7.2 Modified Field Descriptors**

3133

3134 Some field descriptors can be modified by the E and O modifier characters to indicate a
 3135 different format or specification as specified in the LC_TIME FDCC-set description. If the
 3136 corresponding keyword (see "era", "era_year", "era_d_fmt", and "alt_digits") is not
 3137 specified for the current FDCC-set, the unmodified field descriptor value shall be used.

3138

3139	%Ec	FDCC-set's alternate date and time representation.
3140	%EC	The name of the base year (period) in the FDCC-set's alternate represen-
3141		tation.
3142	%Ex	FDCC-set's alternate date representation.
3143	%Ey	Offset from %EC (year only) in the FDCC-set's alternate representation.
3144	%EY	Full alternate year representation.
3145	%Od	Day of month using the FDCC-set's alternate numeric symbols.
3146	%Oe	Day of month using the FDCC-set's alternate numeric symbols.
3147	%Of	Weekday as a decimal number according to alt_day (1 is first day).
3148	%OH	Hour (24-hour clock) using the FDCC-set's alternate numeric symbols.
3149	%OI	Hour (12-hour clock) using the FDCC-set's alternate numeric symbols.
3150	%Om	Month using the FDCC-set's alternate numeric symbols.
3151	%OM	Minutes using the FDCC-set's alternate numeric symbols.
3152	%OS	Seconds using the FDCC-set's alternate numeric symbols.
3153	%Ou	Weekday as a number in the alternate representation of the FDCC-set
3154		(Monday=1).
3155	%OU	Week number of the year (Sunday as the first day of the week) using the
3156		FDCC-set's alternate numeric symbols.
3157	%OV	Week number of the year (Monday as the first day of the week, ISO 8601
3158		rules) using the alternate numeric symbols of the FDCC-set.
3159	%Ow	Weekday as number in the FDCC-set's alternate representation

3160 (Sunday=0).
 3161 %OW Week number of the year (Monday as the first day of the week) using the
 3162 FDCC-set's alternate numeric symbols.
 3163 %Oy Year (offset from %C) in alternate representation.
 3164

3165 4.7.3 "i18n" LC_TIME category

3166 The "i18n" LC_TIME category is (following ISO 8601):

```

3168 LC_TIME
3169 % This is the ISO/IEC 14652 "i18n" definition for
3170 % the LC_TIME category.
3171 %
3172 %
3173 % Weekday and week numbering according to ISO 8601
3174 abday "<1>";"<2>";"<3>";"<4>";"<5>";"<6>;<7>"
3175 day "<1>";"<2>";"<3>";"<4>";"<5>";"<6>;<7>"
3176 week 7;19971201;4
3177 abmon "<0><1>";"<0><2>";"<0><3>";"<0><4>";"<0><5>";"<0><6>";/
3178 "<0><7>";"<0><8>";"<0><9>";"<1><0>";"<1><1>";"<1><2>"
3179 mon "<0><1>";"<0><2>";"<0><3>";"<0><4>";"<0><5>";"<0><6>";/
3180 "<0><7>";"<0><8>";"<0><9>";"<1><0>";"<1><1>";"<1><2>"
3181 am_pm "";"
3182 % Date formats following ISO 8601
3183 % Appropriate date and time representation (%c)
3184 % "%F %T"
3185 d_t_fmt "<%><F><SP><%><T>"
3186 %
3187 % Appropriate date representation (%x) "%F"
3188 d_fmt "<%><F>"
3189 %
3190 % Appropriate time representation (%X) "%T"
3191 t_fmt "<%><T>"
3192 t_fmt_ampm ""
3193 %
3194 END LC_TIME

```

3196 4.8 LC_MESSAGES

3197
 3198 The LC_MESSAGES category shall define the format and values for affirmative and
 3199 negative responses. The operands shall be strings or extended regular expressions to
 3200 specify which response strings that should be considered matches; see ISO/IEC 9945-
 3201 2:1993 clause 2.8.4 for a definition of extended regular expressions. The following
 3202 keywords shall be defined:
 3203

3204
 3205 **copy** Specify the name of an existing FDCC-set to be used as the source for the
 3206 definition of this category. If this keyword is specified, no other keyword
 3207 shall be specified.
 3208 **yesexpr** The operand shall consist of an extended regular expression that describes
 3209 the acceptable affirmative response to a question expecting an affirmative
 3210 or negative response.
 3211 **noexpr** The operand shall consist of an extended regular expression that describes
 3212 the acceptable negative response to a question expecting an affirmative or
 3213 negative response.
 3214

3215 The "i18n" LC_MESSAGES category is:

```

3216 LC_MESSAGES
3217 % This is the ISO/IEC 14652 "i18n" definition for
3218 % the LC_MESSAGES category.
3219 %
3220 %
3221 yesexpr "<U005B><+><1><U005D>"
3222 noexpr "<U005B><-><0><U005D>"

```

3223 END LC_MESSAGES

3224

4.9 LC_PAPER

3225

3226

3227

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

3228

3229

3230

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.

3231

3232

3233

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.

3234

3235

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.

3236

3237

3238

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.

3239

3240

The "i18n" LC_PAPER category is:

3241

3242

3243

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

3244

3245

3246

3247

3248

3249

3250

4.10 LC_NAME

3251

3252

3253

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:

3254

3255

3256

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.

3257

3258

3259

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.

3260

3261

3262

3263

name_gen The operand is a string defining a salutation valid for all persons, example: the Japanese "-sama" salutation in a letter.

3264

3265

name_miss The operand is a string defining a salutation valid for unmarried females.

3266

3267

name_mr The operand is a string defining a salutation valid for males.

3268

name_mrs The operand is a string defining a salutation valid for married females.

3269

name_ms The operand is a string defining a salutation valid for all females.

3270

NOTE: There are a number of variations for addressing a person among the cultures.

3271

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.

3272

3273

3274

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

3275

3276

3277 keywords: "name_fmt".

3278

3279 Escape sequences for the "name_fmt" keyword:

3280

3281 %f Family names.

3282 %F Family names in uppercase.

3283 %g First given name.

3284 %G First given initial.

3285 %l First given name with latin letters.

3286 %o Other shorter name, eg. "Bill".

3287 %m Middle names.

3288 %M Middle initial.

3289 %p Profession.

3290 %s Salutation, such as "Doctor"

3291 %S Abbreviated salutation, such as "Mr." or "Dr."

3292 %d Salutation, using the FDCC-sets conventions, with 1 for the name_gen, 2
3293 for name_mr, 3 for name_mrs, 4 for name_miss, 5 for name_ms. The
3294 vaule may be stored in the database with the person information.

3295 %t If the preceding escape sequence resulted in an empty string, then the
3296 empty string, else a <space>.

3297

3298 Each escape sequence may have an <R> after the <%> to specify that the information is
3299 taken from a Romanized version string of the entity.

3300

3301 The "i18n" LC_NAME category is:

3302

```
3303 LC_NAME
3304 % This is the ISO/IEC 14652 "i18n" definition for
3305 % the LC_NAME category.
```

```
3306 %
3307 name_fmt " <%><p><%><t><%><g><%><t><%><m><%><t><%><f>"
3308 END LC_NAME
```

3309

3310 4.11 LC_ADDRESS

3311

3312 The LC_ADDRESS category defines formats to be used in specifying a location like a
3313 person's living or office, for use in a postal address or in a letter, and other items related
3314 to geography. All keywords are optional. The following keywords shall be recognized:

3315

3316 **copy** Specify the name of an existing FDCC-set to be used as the source
3317 for the definition of this category. If this keyword is specified, no
3318 other keyword shall be specified.

3319 **postal_fmt** Define the appropriate representation of a postal address such as
3320 street and city. The proper formatting of a person's name and title is
3321 done with the "name_fmt" keyword of the LC_NAME category. The
3322 operand shall consist of a string, and can contain any combination of
3323 characters and field descriptors. In addition, the string can contain
3324 escape sequences defined below.

3325 **country_name** The operand is a string with the name of the country in the language
3326 of the FDCC-set.

3327 **country_post** The operand is a string with the abbreviation of the country, used for
3328 postal addresses, for example by CEPT-MAILCODE.

3329 **country_ab2** The operand is a string with the two-letter abbreviation of the

3330		country, according to ISO 3166.
3331	country_ab3	The operand is a string with the three-letter abbreviation of the
3332		country, according to ISO 3166.
3333	country_num	The operand is an integer with the three-digit number of the country,
3334		according to ISO 3166.
3335	country_car	The operand is a string with the abbreviation of the country, used for
3336		motor vehicles and traffic, according to the Genève convention
3337		1949:68.
3338	country_isbn	The operand is a string with the abbreviation of the country, used for
3339		book numbering (ISBN), according to ISO 2108. ISBN numbers are
3340		allocated according to country.
3341	lang_name	The operand is a string with the name of the language in the
3342		language of the FDCC-set.
3343	lang_ab	The operand is a string with the two-letter abbreviation of the
3344		language, according to ISO 639.
3345	lang_term	The operand is a string with the three-letter abbreviation of the
3346		language for terminology use, according to ISO 639-2.
3347	lang_lib	The operand is a string with the three-letter abbreviation of the
3348		language for library use, according to ISO 639-2. If not specified, the
3349		value of the "lang_term" keyword is taken.

3350
 3351 The LC_ADDRESS category defines the interpretation of a number of escape sequences.
 3352 The escape sequences are also available in the definitions with the following
 3353 LC_ADDRESS keywords: "postal_fmt".
 3354

3355 Escape sequences for the "postal_fmt" keyword:

3356		
3357	%a	C/O address.
3358	%f	Firm name.
3359	%d	department name.
3360	%b	Building name.
3361	%s	street or block (eg. Japanese) name.
3362	%h	house number or designation.
3363	%N	if any graphical characters have been specified then an end of line is
3364		made.
3365	%t	if the preceding escape sequence resulted in an empty string, then the
3366		empty string, else a <space>.
3367	%r	room number, door designation.
3368	%e	floor number.
3369	%C	country designation, from the <country_post> keyword.
3370	%z	zip number, postal code.
3371	%T	town, city.
3372	%S	state, province, or prefecture.
3373	%c	country.

3374
 3375 Each escape sequence may have an <R> after the <%> to specify that the information is
 3376 taken from a Romanized version string of the entity.
 3377

3378 NOTE: There are a number of variations for specifying a location among the cultures.
 3379 Some of the information, like the middle names, or even the family name, is not used
 3380 in some cultures. The specification here should be regarded as a start point for this

3381 problem.

3382

3383 The "i18n" LC_ADDRESS category is:

3384

3385 LC_ADDRESS

3386 % This is the ISO/IEC 14652 "i18n" definition for

3387 % the LC_ADDRESS category.

3388 %

3389 postal_fmt " <%><a><%><N><%><f><%><N><%><d><%><N><%><%><N>/

3390 <%><s><SP><%><h><SP><%><e><SP><%><r><%><N>/

3391 <%><C><-><%><z><SP><%><T><%><N><%><c><%><N>"

3392 END LC_ADDRESS

3393

3394

3395

4.12 LC_TELEPHONE

3396

3397 The LC_TELEPHONE category defines formats to be used with telephone services. All

3398 keywords are optional. The following keywords shall be defined:

3399

3400 **copy** Specify the name of an existing FDCC-set to be used as the source
3401 for the definition of this category. If this keyword is specified, no
3402 other keyword shall be specified.

3403 **tel_int_fmt** Define the appropriate representation of a telephone number for
3404 international use. The operand shall consist of a string, and can
3405 contain any combination of characters and field descriptors. In
3406 addition, the string can contain escape sequences defined below.

3407 **tel_dom_fmt** Define the appropriate representation of a telephone number for
3408 domestic use. The operand shall consist of a string, and can contain
3409 any combination of characters and field descriptors. In addition, the
3410 string can contain escape sequences defined below.

3411 **int_select** The operand is a string with the digits used to call international
3412 telephone numbers.

3413 **int_prefix** The operand is a string with the prefix used from other countries to
3414 call the area

3415

3416 The LC_TELEPHONE category defines the interpretation of a number of escape
3417 sequences. The escape sequences are also available in the definitions with the following
3418 LC_TELEPHONE keywords: "tel_int_fmt" and "tel_dom_fmt".

3419

3420 %a area code without prefix (prefix is often <0>).

3421 %A area code including prefix (prefix is often <0>).

3422 %l local number.

3423 %c country code

3424 %C alternative carrier service code used for dialling abroad

3425

3426 The "i18n" LC_TELEPHONE category is:

3427

3428 LC_TELEPHONE

3429 % This is the ISO/IEC 14652 "i18n" definition for

3430 % the LC_TELEPHONE category.

3431 %

3432 tel_int_fmt "<+><%><c><SP><%><a><SP><%><l>"

3433 END LC_TELEPHONE

3434

3435

5. CHARMAP

3436

3437

3438 A character set description may exist for each coded character set supported by an
3439 application. This text is referred elsewhere in this Technical Report as a charmap.

3440
3441 A conforming charmap to be used with a FDCC-set shall support the portable character set
3442 specified in Table 1.

3443
3444 Conforming charmaps shall specify certain character and character set attributes, as
3445 defined in 5.1.

3446 5.1 Character Set Description Text

3447
3448
3449 The character set description text (charmap) describes the mapping between symbolic
3450 character names and actual encoding of a coded character set. It is used to bind the
3451 symbolic character names in a FDCC-set to an actual encoding, so an application can
3452 process data in this encoding.

3453
3454 The following declarations can precede the character definitions. Each shall consist of the
3455 symbol shown in the following list, starting in column 1, including the surrounding
3456 brackets, followed by one or more "blank"s, followed by the value to be assigned to the
3457 symbol. If any of the declarations are included, they shall be specified in the order shown
3458 in the following list:

3459	<code_set_name>	The name of the coded character set for which the character set description text is defined. The characters of the name shall be taken from the set of characters with visible glyphs defined in Table 1.
3460		
3461	<mb_cur_max>	The maximum number of bytes in a multibyte character. This shall default to 1.
3462		
3463	<mb_cur_min>	An unsigned positive integer value that shall define the minimum number of bytes in a character for the encoded character set. The value shall be less or equal to "mb_cur_max". If not specified, the minimum number shall be equal to "mb_cur_max".
3464		
3465	<escape_char>	The escape character used to indicate that the characters following shall be interpreted in a special way, as defined later in this subclause. This shall default to backslash (\). The character slash (/) is used in all the following text and examples, unless otherwise noted.
3466		
3467	<comment_char>	The character that when placed in column 1 of a charmap line, is used to indicate that the line shall be ignored. The default character shall be the number sign (#). The character percent-sign (%) is used in all the following text and examples, unless otherwise noted.
3468		
3469	<repertoiremap>	The name of the repertoiremap used to define the symbolic character names in the charmap. The characters of the name shall be taken from the set of characters with visible glyphs
3470		
3471		
3472		
3473		
3474		
3475		
3476		
3477		
3478		
3479		
3480		
3481		
3482		
3483		
3484		
3485		
3486		
3487		
3488		

3489 defined in Table 1.

3490		
3491	<escseq>	defines the escape sequences for ISO 2022 shifting for the coded
3492		character set defined by the charmap. The semicolon-separated
3493		operands are all strings with characters taken from the set of
3494		characters with visible glyphs defined in table 1. The first
3495		operand defines the g-set or c-set to be defined, and the
3496		following values are defined: c0, c1, g0, g1, g2, g3. The second
3497		operand defines what range of characters in the charmap is
3498		affected, and the values defined are: c0, c1, g0, g1. The third
3499		operand is the escape sequence that is defined.
3500		
3501	<addset>	the name of the charmap to be added the current coded character
3502		set and to be selected by the escape sequences defined by
3503		<escseq> of the added charmap.
3504		
3505	<include>	include the encoding of another charmap in the current charmap.
3506		The semicolon-separated operands are all strings with characters
3507		taken from the set of characters with visible glyphs defined in
3508		table 1. The first operand defines the g-set or c-set to be defined
3509		in the current charmap, and the following values are defined: c0,
3510		c1, g0, g1, g2, g3. The second operand defines a range of
3511		characters in the referenced charmap, and the values defined are:
3512		c0, c1, g0, g1. The third operand is the name of the charmap to
3513		be included. The coded character sets are defined initially for the
3514		encoding, and therefore do not need escape sequences for
3515		identification. If two g0 sets are defined, the second is switched
3516		to using the SHIFT OUT control character, while the first is
3517		shifted to using the SHIFT IN control character.
3518		

3519 The character set mapping definitions shall be all the lines immediately following an
 3520 identifier line containing the string "CHARMAP" starting in column 1, and preceding a
 3521 trailer line containing the string "END CHARMAP" starting in column 1. Empty lines
 3522 and lines containing a <comment_char> in the first column shall be ignored. Each
 3523 noncomment line of the character set mapping definition (i.e., between the "CHARMAP"
 3524 and "END CHARMAP" lines of the text) shall be in one of the following syntaxes.

3525
 3526
 3527 "%s %s %s\n", <symbolic-name>,<encoding>,<comments>

3528
 3529 "%s...%s %s %s\n", <symbolic-name>,<symbolic-name>,<encoding>,<comments>

3530
 3531 "%s....%s %s %s\n", <symbolic-name>,<symbolic-name>,<encoding>,<comments>

3532
 3533 "%s..%s %s %s\n", <symbolic-name>,<symbolic-name>,<encoding>,<comments>

3534
 3535 In the first syntax, the line of the character set mapping definition shall start with the
 3536 symbolic name, immediately preceded by a <less-than> character and immediately
 3537 followed by a <greater-than> character. Symbolic names shall only contain characters
 3538 from the set shown with a visible glyph in Table 1.

3539 The same symbolic name may occur several times, with different values. The first value is
3540 the one used when generating an encoding, while the other values are accepted in
3541 decoding. Symbolic names may be included to identify values that can overlap with each
3542 other or with the values of the symbolic names shown in Table 1. It is possible to specify
3543 symbolic names for which no encoding exists in the encoded character set, by not
3544 specifying a value.

3545
3546 In the second and third syntax (symbolic decimal ellipsis), the line in the character set
3547 mapping defines a range of one or more symbolic names. The difference between the
3548 second and the third syntax is the number of dots in the ellipsis: the second has 3 dots, the
3549 third has 4 dots. In these forms the symbolic names shall consist of zero or more
3550 nonnumeric characters from the set shown with visible glyphs in Table 1, followed by an
3551 integer formed by one or more decimal digits. The characters preceding the integer shall
3552 be identical in the two symbolic names, and the integer formed by the digits in the second
3553 symbolic name shall be identical to or greater than the integer formed by the digits in the
3554 first name. This shall be interpreted as a series of symbolic names formed from the
3555 common part and each of the integers in decimal format between the first and the second
3556 integer, inclusive, and with a length of the symbolic names generated that is equal to the
3557 length of the first (and also the second) symbolic name. As an example,
3558 <j0101>....<j0104> is interpreted as the symbolic names <j0101>, <j0102>, <j0103>, and
3559 <j0104>, in that order.

3560
3561 Note: The rationale to allow both a 3-dot and a 4-dot symbol for symbolic decimal
3562 ellipses is that in the POSIX standard the decimal symbolic ellipses was defined by a 3-
3563 dot symbol for charmaps, while the 3-dot symbol was an absolute ellipses for POSIX
3564 locales, and this International standard specifies a 4-dot symbol for the decimal
3565 symbolic ellipses. The 3-dot symbolic decimal ellipses in charmaps is deprecated.

3566
3567 In the fourth syntax (symbolic hexadecimal ellipsis, with two dots), the line in the
3568 character set mapping defines a range of one or more symbolic names. In this form the
3569 symbolic names shall consist of zero or more nonnumeric characters from the set shown
3570 with visible glyphs in Table 1, followed by an integer formed by one or more hexadecimal
3571 digits, using uppercase letters only for the range "A" to "F". The characters preceding the
3572 hexadecimal integer shall be identical in the two symbolic names, and the integer formed
3573 by the hexadecimal digits in the second symbolic name shall be identical to or greater than
3574 the integer formed by the hexadecimal digits in the first name. This shall be interpreted as
3575 a series of symbolic names formed from the common part and each of the integers in
3576 hexadecimal format using uppercase letters only between the first and the second integer,
3577 inclusive, and with a length of the symbolic names generated that is equal to the length of
3578 the first (and also the second) symbolic name. As an example, <U010E>..<U0111> is
3579 interpreted as the symbolic names <U010E>, <U010F>, <U0110>, and <U0111>, in that
3580 order.

3581
3582 The encoding part shall be expressed as one (for single-byte values) or more concatenated
3583 decimal, octal or hexadecimal constants. Decimal constants shall be represented by two or
3584 three decimal digits, preceded by the escape character and the lowercase letter "d"; for
3585 example /d05, /d97, or /d143. Hexadecimal constants shall be represented by two
3586 hexadecimal digits, preceded by the escape character and the lowercase letter "x"; for
3587 example /x05, /x61, or /x8f. Octal constants shall be represented by two or three octal
3588 digits, preceded by the escape character; for example /05, /141, or /217. In a charmap,
3589 each constant should represent an 8 bit byte for portability reasons. Applications

3590 supporting other byte sizes may allow constants to represent values larger than those that
 3591 can be represented in 8 bit bytes, and to allow additional digits in constants. When
 3592 constants are concatenated for multibyte character values, they may be of different types,
 3593 and interpreted in byte order from the first to the last with the least significant byte of the
 3594 multibyte character specified by the last byte. The manner in which these constants are
 3595 represented in the character stored in the system is application defined. Omitting bytes
 3596 from a multibyte character produces undefined results.
 3597

3598 In lines defining ranges of symbolic names, the encoded value is the value for the first
 3599 symbolic name in the range (the symbolic name preceding the ellipsis). Subsequent
 3600 symbolic names defined by the range shall have encoding values in increasing order. For
 3601 example the line

```
3602
3603     <j0101>....<j0104>      /d129/d254
3604
```

3605 shall be interpreted as

```
3606
3607     <j0101>    /d129/d254
3608     <j0102>    /d129/d255
3609     <j0103>    /d130/d000
3610     <j0104>    /d130/d001
3611
```

3612 The comments parameter is optional.

3613
 3614

3615 Example of using ISO 2022 techniques:

3616

3617 The following example defines two coded character sets, a 7-bit and a 14-bit. They are then merged into one
 3618 encoding. It is an example on how encodings used in Eastern Asia could be specified.

3619

3620 The 7-bit charmap

3621

```
3622     <escape_char> /
3623     <comment_char> %
3624     % The 7bit charmap defines both control and graphic characters
3625     <code_set_name> "eastern7bit"
3626     <escseq>      "c0";"c0","/x21/x40"
3627     <escseq>      "g0";"g0","/x28/x48"
3628     <escseq>      "g1";"g0","/x29/x48"
3629     <escseq>      "g2";"g0","/x2A/x48"
3630     <escseq>      "g3";"g0","/x2B/x48"
3631
```

3632 CHARMAP

```
3633     <tab>          /x08
3634     <newline>     /x0D
3635     <a>           /x61
3636     % more character encodings to be defined here
3637     END CHARMAP
3638
```

3639

3640

3640 The 14-bit charmap

3641

```
3642     <escape_char> /
3643     <comment_char> %
3644     <code_set_name> "eastern14bit"
3645     <mb_cur_max>   2
```

```

3646 <esqseq> "g0";"g0";"/x24/x40"
3647 <esqseq> "g1";"g0";"/x24/x29/x40"
3648 <esqseq> "g2";"g0";"/x24/x2A/x40"
3649 <esqseq> "g3";"g0";"/x24/x2B/x40"
3650 CHARMAP
3651 <U0365> /d036/d055 % the character codes are only examples
3652 <U0744> /d036/d056
3653 % more character encodings to be defined here
3654 END CHARMAP

```

The merged encoding

```

3659 <escape_char> /
3660 <comment_char> %
3661 <code_set_name> "shift-eastern"
3662 <mb_cur_max> 2
3663 <mb_cur_min> 1
3664 <include> "c0";"c0";"eastern7bit"
3665 <include> "g0";"g0";"eastern7bit"
3666 <include> "g1";"g0";"eastern14bit"
3667 % This defines the g0 values of "eastern14bit" (without the 8th
3668 % bit set) to be the g1 in this encoding (with the 8th bit set).
3669 %
3670 % So the bytes without the 8th bit set is from the "shift7bit"
3671 % coded character set, while bytes with the 8th bit set are from
3672 % the 14-bit set.

```

Another merged encoding using the same charmaps:

```

3676 <escape_char> /
3677 <comment_char> %
3678 <code_set_name> "EUC-eastern"
3679 <mb_cur_max> 2
3680 <mb_cur_min> 1
3681 <include> "c0";"c0";"eastern7bit"
3682 <include> "g0";"g0";"eastern7bit"
3683 <include> "g0";"g0";"eastern14bit"
3684 % As there are two "g0" sets defined, the first referenced is the
3685 % initial g0 set, while the second can be shifted to via the SHIFT OUT
3686 % control character. The first can then be shifted to by the SHIFT IN
3687 % control character.

```

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>
```

3704 The symbolic character name and the ISO/IEC 10646 short identifier are each surrounded
 3705 by angle brackets <>, and the fields shall be separated by one or more spaces or tabs on a
 3706 line. If a right angle bracket or an escape character is used within a symbolic name, it
 3707 shall be preceded by the escape character. Characters not in ISO/IEC 10646 may be
 3708 referenced by the symbolic character names <P00000000>..<<PF8FFFFFFF>.

3709
 3710 The escape character can be redefined from the default reverse solidus (\) with the first
 3711 line of the Repertoiremap containing the string "escape_char" followed by one or more
 3712 spaces or tabs and then the escape character.

3713
 3714 Several symbolic character names can refer to the same abstract character, and are then
 3715 used as synonyms in FDCC-sets and charmaps. The set of <U0000>..<<UFFFF> and
 3716 <U00000000>..<<U7FFFFFFF> symbolic names (no lowercase letters) are predefined and
 3717 refers to the corresponding code points of ISO/IEC 10646 with the same short identifier.

3718
 3719 The "i18nrep" repertoiremap is defined to accommodate prior art, such as defined in the
 3720 ISO/IEC 9945-2:1993 standard annex G, and used by ISO and IEC member bodies in their
 3721 national POSIX locale specifications, and as used in POSIX locales distributed by the
 3722 ISO/IEC POSIX working group and X/Open. Many POSIX charmaps registered with
 3723 ISO/IEC 15897 use these symbolic names. It also reflects use on the Internet, and many of
 3724 the Internet registered charsets are specified using these symbolic names. The "i18nrep"
 3725 repertoiremap thus facilitates reuse of both POSIX locale data and POSIX charmaps with
 3726 data from this Technical Report. The contents of the "i18nrep" repertoiremap is as follows:

```

3727 escape_char /
3728 <NUL>          <U0000>  NULL (NUL)
3729 <SOH>         <U0001>  START OF HEADING (SOH)
3730 <STX>         <U0002>  START OF TEXT (STX)
3731 <ETX>         <U0003>  END OF TEXT (ETX)
3732 <EOT>         <U0004>  END OF TRANSMISSION (EOT)
3733 <ENQ>         <U0005>  ENQUIRY (ENQ)
3734 <ACK>         <U0006>  ACKNOWLEDGE (ACK)
3735 <alert>       <U0007>  BELL (BEL)
3736 <BEL>         <U0007>  BELL (BEL)
3737 <backspace>   <U0008>  BACKSPACE (BS)
3738 <tab>         <U0009>  CHARACTER TABULATION (HT)
3739 <newline>     <U000A>  LINE FEED (LF)
3740 <vertical-tab> <U000B>  LINE TABULATION (VT)
3741 <form-feed>   <U000C>  FORM FEED (FF)
3742 <carriage-return> <U000D>  CARRIAGE RETURN (CR)
3743 <DLE>         <U0010>  DATALINK ESCAPE (DLE)
3744 <DC1>         <U0011>  DEVICE CONTROL ONE (DC1)
3745 <DC2>         <U0012>  DEVICE CONTROL TWO (DC2)
3746 <DC3>         <U0013>  DEVICE CONTROL THREE (DC3)
3747 <DC4>         <U0014>  DEVICE CONTROL FOUR (DC4)
3748 <NAK>         <U0015>  NEGATIVE ACKNOWLEDGE (NAK)
3749 <SYN>         <U0016>  SYNCHRONOUS IDLE (SYN)
3750 <ETB>         <U0017>  END OF TRANSMISSION BLOCK (ETB)
3751 <CAN>         <U0018>  CANCEL (CAN)
3752 <SUB>         <U001A>  SUBSTITUTE (SUB)
3753 <ESC>         <U001B>  ESCAPE (ESC)
3754 <IS4>         <U001C>  FILE SEPARATOR (IS4)
3755 <IS3>         <U001D>  GROUP SEPARATOR (IS3)
3756 <intro>       <U001D>  GROUP SEPARATOR (IS3)
3757 <IS2>         <U001E>  RECORD SEPARATOR (IS2)
3758 <IS1>         <U001F>  UNIT SEPARATOR (IS1)
3759 <DEL>         <U007F>  DELETE (DEL)
3760 <space>       <U0020>  SPACE
3761 <exclamation-mark> <U0021>  EXCLAMATION MARK
3762 <quotation-mark> <U0022>  QUOTATION MARK
3763 <number-sign>   <U0023>  NUMBER SIGN
3764 <dollar-sign>   <U0024>  DOLLAR SIGN
3765 <percent-sign>  <U0025>  PERCENT SIGN
3766 <ampersand>       <U0026>  AMPERSAND
3767 <apostrophe>   <U0027>  APOSTROPHE
3768 <left-parenthesis> <U0028>  LEFT PARENTHESIS
3769 <right-parenthesis> <U0029>  RIGHT PARENTHESIS
3770 <asterisk>      <U002A>  ASTERISK
3771 <plus-sign>     <U002B>  PLUS SIGN
3772 <comma>        <U002C>  COMMA
3773 <hyphen>       <U002D>  HYPHEN-MINUS
3774 <hyphen-minus> <U002D>  HYPHEN-MINUS
3775

```

3776	<period>	<U002E>	FULL STOP
3777	<full-stop>	<U002E>	FULL STOP
3778	<slash>	<U002F>	SOLIDUS
3779	<solidus>	<U002F>	SOLIDUS
3780	<zero>	<U0030>	DIGIT ZERO
3781	<one>	<U0031>	DIGIT ONE
3782	<two>	<U0032>	DIGIT TWO
3783	<three>	<U0033>	DIGIT THREE
3784	<four>	<U0034>	DIGIT FOUR
3785	<five>	<U0035>	DIGIT FIVE
3786	<six>	<U0036>	DIGIT SIX
3787	<seven>	<U0037>	DIGIT SEVEN
3788	<eight>	<U0038>	DIGIT EIGHT
3789	<nine>	<U0039>	DIGIT NINE
3790	<colon>	<U003A>	COLON
3791	<semicolon>	<U003B>	SEMICOLON
3792	<less-than-sign>	<U003C>	LESS-THAN SIGN
3793	<equals-sign>	<U003D>	EQUALS SIGN
3794	<greater-than-sign>	<U003E>	GREATER-THAN SIGN
3795	<question-mark>	<U003F>	QUESTION MARK
3796	<commercial-at>	<U0040>	COMMERCIAL AT
3797	<left-square-bracket>	<U005B>	LEFT SQUARE BRACKET
3798	<backslash>	<U005C>	REVERSE SOLIDUS
3799	<reverse-solidus>	<U005C>	REVERSE SOLIDUS
3800	<right-square-bracket>	<U005D>	RIGHT SQUARE BRACKET
3801	<circumflex>	<U005E>	CIRCUMFLEX ACCENT
3802	<circumflex-accent>	<U005E>	CIRCUMFLEX ACCENT
3803	<underscore>	<U005F>	LOW LINE
3804	<low-line>	<U005F>	LOW LINE
3805	<grave-accent>	<U0060>	GRAVE ACCENT
3806	<left-brace>	<U007B>	LEFT CURLY BRACKET
3807	<left-curly-bracket>	<U007B>	LEFT CURLY BRACKET
3808	<vertical-line>	<U007C>	VERTICAL LINE
3809	<right-brace>	<U007D>	RIGHT CURLY BRACKET
3810	<right-curly-bracket>	<U007D>	RIGHT CURLY BRACKET
3811	<tilde>	<U007E>	TILDE
3812			
3813	<a8>	<U0252>	Weight indicating the position of the last a
3814	<b8>	<U0182>	Weight indicating the position of the last b
3815	<c8>	<U0255>	Weight indicating the position of the last c
3816	<d8>	<U018D>	Weight indicating the position of the last d
3817	<e8>	<U0264>	Weight indicating the position of the last e
3818	<f8>	<U0191>	Weight indicating the position of the last f
3819	<g8>	<U01A2>	Weight indicating the position of the last g
3820	<h8>	<U02BD>	Weight indicating the position of the last h
3821	<i8>	<U0196>	Weight indicating the position of the last i
3822	<j8>	<U0284>	Weight indicating the position of the last j
3823	<k8>	<U029E>	Weight indicating the position of the last k
3824	<l8>	<U028E>	Weight indicating the position of the last l
3825	<m8>	<U0271>	Weight indicating the position of the last m
3826	<n8>	<U014A>	Weight indicating the position of the last n
3827	<o8>	<U0277>	Weight indicating the position of the last o
3828	<p8>	<U0278>	Weight indicating the position of the last p
3829	<q8>	<U0138>	Weight indicating the position of the last q
3830	<r8>	<U02B6>	Weight indicating the position of the last r
3831	<s8>	<U0286>	Weight indicating the position of the last s
3832	<t8>	<U0287>	Weight indicating the position of the last t
3833	<u8>	<U01B1>	Weight indicating the position of the last u
3834	<v8>	<U028C>	Weight indicating the position of the last v
3835	<w8>	<U028D>	Weight indicating the position of the last w
3836	<x8>	<U216B>	Weight indicating the position of the last x
3837	<y8>	<U01B3>	Weight indicating the position of the last y
3838	<z8>	<U0293>	Weight indicating the position of the last z
3839			
3840	<NU>	<U0000>	NULL (NUL)
3841	<SH>	<U0001>	START OF HEADING (SOH)
3842	<SX>	<U0002>	START OF TEXT (STX)
3843	<EX>	<U0003>	END OF TEXT (ETX)
3844	<ET>	<U0004>	END OF TRANSMISSION (EOT)
3845	<EQ>	<U0005>	ENQUIRY (ENQ)
3846	<AK>	<U0006>	ACKNOWLEDGE (ACK)
3847	<BL>	<U0007>	BELL (BEL)
3848	<BS>	<U0008>	BACKSPACE (BS)
3849	<HT>	<U0009>	CHARACTER TABULATION (HT)
3850	<LF>	<U000A>	LINE FEED (LF)
3851	<VT>	<U000B>	LINE TABULATION (VT)
3852	<FF>	<U000C>	FORM FEED (FF)
3853	<CR>	<U000D>	CARRIAGE RETURN (CR)
3854	<SO>	<U000E>	SHIFT OUT (SO)
3855	<SI>	<U000F>	SHIFT IN (SI)
3856	<DL>	<U0010>	DATALINK ESCAPE (DLE)
3857	<D1>	<U0011>	DEVICE CONTROL ONE (DC1)
3858	<D2>	<U0012>	DEVICE CONTROL TWO (DC2)
3859	<D3>	<U0013>	DEVICE CONTROL THREE (DC3)
3860	<D4>	<U0014>	DEVICE CONTROL FOUR (DC4)
3861	<NK>	<U0015>	NEGATIVE ACKNOWLEDGE (NAK)
3862	<SY>	<U0016>	SYNCHRONOUS IDLE (SYN)
3863	<EB>	<U0017>	END OF TRANSMISSION BLOCK (ETB)
3864	<CN>	<U0018>	CANCEL (CAN)

3865		<U0019>	END OF MEDIUM (EM)
3866	<SB>	<U001A>	SUBSTITUTE (SUB)
3867	<EC>	<U001B>	ESCAPE (ESC)
3868	<FS>	<U001C>	FILE SEPARATOR (IS4)
3869	<GS>	<U001D>	GROUP SEPARATOR (IS3)
3870	<RS>	<U001E>	RECORD SEPARATOR (IS2)
3871	<US>	<U001F>	UNIT SEPARATOR (IS1)
3872	<DT>	<U007F>	DELETE (DEL)
3873	<PA>	<U0080>	PADDING CHARACTER (PAD)
3874	<HO>	<U0081>	HIGH OCTET PRESET (HOP)
3875	<BH>	<U0082>	BREAK PERMITTED HERE (BPH)
3876	<NH>	<U0083>	NO BREAK HERE (NBH)
3877	<IN>	<U0084>	INDEX (IND)
3878	<NL>	<U0085>	NEXT LINE (NEL)
3879	<SA>	<U0086>	START OF SELECTED AREA (SSA)
3880	<ES>	<U0087>	END OF SELECTED AREA (ESA)
3881	<HS>	<U0088>	CHARACTER TABULATION SET (HTS)
3882	<HJ>	<U0089>	CHARACTER TABULATION WITH JUSTIFICATION (HTJ)
3883	<VS>	<U008A>	LINE TABULATION SET (VTS)
3884	<PD>	<U008B>	PARTIAL LINE FORWARD (PLD)
3885	<PU>	<U008C>	PARTIAL LINE BACKWARD (PLU)
3886	<RI>	<U008D>	REVERSE LINE FEED (RI)
3887	<S2>	<U008E>	SINGLE-SHIFT TWO (SS2)
3888	<S3>	<U008F>	SINGLE-SHIFT THREE (SS3)
3889	<DC>	<U0090>	DEVICE CONTROL STRING (DCS)
3890	<P1>	<U0091>	PRIVATE USE ONE (PU1)
3891	<P2>	<U0092>	PRIVATE USE TWO (PU2)
3892	<TS>	<U0093>	SET TRANSMIT STATE (STS)
3893	<CC>	<U0094>	CANCEL CHARACTER (CCH)
3894	<MW>	<U0095>	MESSAGE WAITING (MW)
3895	<SG>	<U0096>	START OF GUARDED AREA (SPA)
3896	<EG>	<U0097>	END OF GUARDED AREA (EPA)
3897	<SS>	<U0098>	START OF STRING (SOS)
3898	<GC>	<U0099>	SINGLE GRAPHIC CHARACTER INTRODUCER (SGCI)
3899	<SC>	<U009A>	SINGLE CHARACTER INTRODUCER (SCI)
3900	<CI>	<U009B>	CONTROL SEQUENCE INTRODUCER (CSI)
3901	<ST>	<U009C>	STRING TERMINATOR (ST)
3902	<OC>	<U009D>	OPERATING SYSTEM COMMAND (OSC)
3903	<PM>	<U009E>	PRIVACY MESSAGE (PM)
3904	<AC>	<U009F>	APPLICATION PROGRAM COMMAND (APC)
3905	<SP>	<U0020>	SPACE
3906	<!>	<U0021>	EXCLAMATION MARK
3907	<">	<U0022>	QUOTATION MARK
3908	<#>	<U0023>	NUMBER SIGN
3909	<\$>	<U0024>	DOLLAR SIGN
3910	<%>	<U0025>	PERCENT SIGN
3911	<&>	<U0026>	AMPERSAND
3912	<'>	<U0027>	APOSTROPHE
3913	<(>	<U0028>	LEFT PARENTHESIS
3914	<)>	<U0029>	RIGHT PARENTHESIS
3915	<*>	<U002A>	ASTERISK
3916	<+>	<U002B>	PLUS SIGN
3917	<,>	<U002C>	COMMA
3918	<->	<U002D>	HYPHEN-MINUS
3919	<.>	<U002E>	FULL STOP
3920	<//>	<U002F>	SOLIDUS
3921	<0>	<U0030>	DIGIT ZERO
3922	<1>	<U0031>	DIGIT ONE
3923	<2>	<U0032>	DIGIT TWO
3924	<3>	<U0033>	DIGIT THREE
3925	<4>	<U0034>	DIGIT FOUR
3926	<5>	<U0035>	DIGIT FIVE
3927	<6>	<U0036>	DIGIT SIX
3928	<7>	<U0037>	DIGIT SEVEN
3929	<8>	<U0038>	DIGIT EIGHT
3930	<9>	<U0039>	DIGIT NINE
3931	<:>	<U003A>	COLON
3932	<:>	<U003B>	SEMICOLON
3933	<<>	<U003C>	LESS-THAN SIGN
3934	<=>	<U003D>	EQUALS SIGN
3935	</>>	<U003E>	GREATER-THAN SIGN
3936	<?>	<U003F>	QUESTION MARK
3937	<@>	<U0040>	COMMERCIAL AT
3938	<A>	<U0041>	LATIN CAPITAL LETTER A
3939		<U0042>	LATIN CAPITAL LETTER B
3940	<C>	<U0043>	LATIN CAPITAL LETTER C
3941	<D>	<U0044>	LATIN CAPITAL LETTER D
3942	<E>	<U0045>	LATIN CAPITAL LETTER E
3943	<F>	<U0046>	LATIN CAPITAL LETTER F
3944	<G>	<U0047>	LATIN CAPITAL LETTER G
3945	<H>	<U0048>	LATIN CAPITAL LETTER H
3946	<I>	<U0049>	LATIN CAPITAL LETTER I
3947	<J>	<U004A>	LATIN CAPITAL LETTER J
3948	<K>	<U004B>	LATIN CAPITAL LETTER K
3949	<L>	<U004C>	LATIN CAPITAL LETTER L
3950	<M>	<U004D>	LATIN CAPITAL LETTER M
3951	<N>	<U004E>	LATIN CAPITAL LETTER N
3952	<O>	<U004F>	LATIN CAPITAL LETTER O
3953	<P>	<U0050>	LATIN CAPITAL LETTER P

3954	<Q>	<U0051>	LATIN CAPITAL LETTER Q
3955	<R>	<U0052>	LATIN CAPITAL LETTER R
3956	<S>	<U0053>	LATIN CAPITAL LETTER S
3957	<T>	<U0054>	LATIN CAPITAL LETTER T
3958	<U>	<U0055>	LATIN CAPITAL LETTER U
3959	<V>	<U0056>	LATIN CAPITAL LETTER V
3960	<W>	<U0057>	LATIN CAPITAL LETTER W
3961	<X>	<U0058>	LATIN CAPITAL LETTER X
3962	<Y>	<U0059>	LATIN CAPITAL LETTER Y
3963	<Z>	<U005A>	LATIN CAPITAL LETTER Z
3964	<<(>	<U005B>	LEFT SQUARE BRACKET
3965	</////>	<U005C>	REVERSE SOLIDUS
3966	<)/>>	<U005D>	RIGHT SQUARE BRACKET
3967	<' />>	<U005E>	CIRCUMFLEX ACCENT
3968	<_>	<U005F>	LOW LINE
3969	<' !>	<U0060>	GRAVE ACCENT
3970	<a>	<U0061>	LATIN SMALL LETTER A
3971		<U0062>	LATIN SMALL LETTER B
3972	<c>	<U0063>	LATIN SMALL LETTER C
3973	<d>	<U0064>	LATIN SMALL LETTER D
3974	<e>	<U0065>	LATIN SMALL LETTER E
3975	<f>	<U0066>	LATIN SMALL LETTER F
3976	<g>	<U0067>	LATIN SMALL LETTER G
3977	<h>	<U0068>	LATIN SMALL LETTER H
3978	<i>	<U0069>	LATIN SMALL LETTER I
3979	<j>	<U006A>	LATIN SMALL LETTER J
3980	<k>	<U006B>	LATIN SMALL LETTER K
3981	<l>	<U006C>	LATIN SMALL LETTER L
3982	<m>	<U006D>	LATIN SMALL LETTER M
3983	<n>	<U006E>	LATIN SMALL LETTER N
3984	<o>	<U006F>	LATIN SMALL LETTER O
3985	<p>	<U0070>	LATIN SMALL LETTER P
3986	<q>	<U0071>	LATIN SMALL LETTER Q
3987	<r>	<U0072>	LATIN SMALL LETTER R
3988	<s>	<U0073>	LATIN SMALL LETTER S
3989	<t>	<U0074>	LATIN SMALL LETTER T
3990	<u>	<U0075>	LATIN SMALL LETTER U
3991	<v>	<U0076>	LATIN SMALL LETTER V
3992	<w>	<U0077>	LATIN SMALL LETTER W
3993	<x>	<U0078>	LATIN SMALL LETTER X
3994	<y>	<U0079>	LATIN SMALL LETTER Y
3995	<z>	<U007A>	LATIN SMALL LETTER Z
3996	<(!>	<U007B>	LEFT CURLY BRACKET
3997	<!!>	<U007C>	VERTICAL LINE
3998	<!)>	<U007D>	RIGHT CURLY BRACKET
3999	<' ?>	<U007E>	TILDE
4000	<NS>	<U00A0>	NO-BREAK SPACE
4001	<!I>	<U00A1>	INVERTED EXCLAMATION MARK
4002	<Ct>	<U00A2>	CENT SIGN
4003	<Pd>	<U00A3>	POUND SIGN
4004	<Cu>	<U00A4>	CURRENCY SIGN
4005	<Ye>	<U00A5>	YEN SIGN
4006	<BB>	<U00A6>	BROKEN BAR
4007	<SE>	<U00A7>	SECTION SIGN
4008	<' :>	<U00A8>	DIAERESIS
4009	<Co>	<U00A9>	COPYRIGHT SIGN
4010	<-a>	<U00AA>	FEMININE ORDINAL INDICATOR
4011	<<<>	<U00AB>	LEFT-POINTING DOUBLE ANGLE QUOTATION MARK
4012	<NO>	<U00AC>	NOT SIGN
4013	<-->	<U00AD>	SOFT HYPHEN
4014	<Rg>	<U00AE>	REGISTERED SIGN
4015	<' m>	<U00AF>	MACRON
4016	<DG>	<U00B0>	DEGREE SIGN
4017	<+>	<U00B1>	PLUS-MINUS SIGN
4018	<2S>	<U00B2>	SUPERSCRIP TWO
4019	<3S>	<U00B3>	SUPERSCRIP THREE
4020	<' ' >	<U00B4>	ACUTE ACCENT
4021	<My>	<U00B5>	MICRO SIGN
4022	<PI>	<U00B6>	PILCROW SIGN
4023	<.M>	<U00B7>	MIDDLE DOT
4024	<' , >	<U00B8>	CEDILLA
4025	<1S>	<U00B9>	SUPERSCRIP ONE
4026	<-o>	<U00BA>	MASCULINE ORDINAL INDICATOR
4027	</>/>>	<U00BB>	RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK
4028	<14>	<U00BC>	VULGAR FRACTION ONE QUARTER
4029	<12>	<U00BD>	VULGAR FRACTION ONE HALF
4030	<34>	<U00BE>	VULGAR FRACTION THREE QUARTERS
4031	<?I>	<U00BF>	INVERTED QUESTION MARK
4032	<A!>	<U00C0>	LATIN CAPITAL LETTER A WITH GRAVE
4033	<A'>	<U00C1>	LATIN CAPITAL LETTER A WITH ACUTE
4034	<A/ >>	<U00C2>	LATIN CAPITAL LETTER A WITH CIRCUMFLEX
4035	<A?>	<U00C3>	LATIN CAPITAL LETTER A WITH TILDE
4036	<A :>	<U00C4>	LATIN CAPITAL LETTER A WITH DIAERESIS
4037	<AA>	<U00C5>	LATIN CAPITAL LETTER A WITH RING ABOVE
4038	<AE>	<U00C6>	LATIN CAPITAL LETTER AE (ash)
4039	<C , >	<U00C7>	LATIN CAPITAL LETTER C WITH CEDILLA
4040	<E!>	<U00C8>	LATIN CAPITAL LETTER E WITH GRAVE
4041	<E'>	<U00C9>	LATIN CAPITAL LETTER E WITH ACUTE
4042	<E/ >>	<U00CA>	LATIN CAPITAL LETTER E WITH CIRCUMFLEX

4043	<E:>	<U00CB>	LATIN CAPITAL LETTER E WITH DIAERESIS
4044	<I!>	<U00CC>	LATIN CAPITAL LETTER I WITH GRAVE
4045	<I'>	<U00CD>	LATIN CAPITAL LETTER I WITH ACUTE
4046	<I/>>	<U00CE>	LATIN CAPITAL LETTER I WITH CIRCUMFLEX
4047	<I:>	<U00CF>	LATIN CAPITAL LETTER I WITH DIAERESIS
4048	<D->	<U00D0>	LATIN CAPITAL LETTER ETH (Icelandic)
4049	<N?>	<U00D1>	LATIN CAPITAL LETTER N WITH TILDE
4050	<O!>	<U00D2>	LATIN CAPITAL LETTER O WITH GRAVE
4051	<O'>	<U00D3>	LATIN CAPITAL LETTER O WITH ACUTE
4052	<O/>>	<U00D4>	LATIN CAPITAL LETTER O WITH CIRCUMFLEX
4053	<O?>	<U00D5>	LATIN CAPITAL LETTER O WITH TILDE
4054	<O:>	<U00D6>	LATIN CAPITAL LETTER O WITH DIAERESIS
4055	<*X>	<U00D7>	MULTIPLICATION SIGN
4056	<O//>	<U00D8>	LATIN CAPITAL LETTER O WITH STROKE
4057	<U!>	<U00D9>	LATIN CAPITAL LETTER U WITH GRAVE
4058	<U'>	<U00DA>	LATIN CAPITAL LETTER U WITH ACUTE
4059	<U/>>	<U00DB>	LATIN CAPITAL LETTER U WITH CIRCUMFLEX
4060	<U:>	<U00DC>	LATIN CAPITAL LETTER U WITH DIAERESIS
4061	<Y'>	<U00DD>	LATIN CAPITAL LETTER Y WITH ACUTE
4062	<TH>	<U00DE>	LATIN CAPITAL LETTER THORN (Icelandic)
4063	<ss>	<U00DF>	LATIN SMALL LETTER SHARP S (German)
4064	<a!>	<U00E0>	LATIN SMALL LETTER A WITH GRAVE
4065	<a'>	<U00E1>	LATIN SMALL LETTER A WITH ACUTE
4066	<a/>>	<U00E2>	LATIN SMALL LETTER A WITH CIRCUMFLEX
4067	<a?>	<U00E3>	LATIN SMALL LETTER A WITH TILDE
4068	<a:>	<U00E4>	LATIN SMALL LETTER A WITH DIAERESIS
4069	<aa>	<U00E5>	LATIN SMALL LETTER A WITH RING ABOVE
4070	<ae>	<U00E6>	LATIN SMALL LETTER AE (ash)
4071	<c.>	<U00E7>	LATIN SMALL LETTER C WITH CEDILLA
4072	<e!>	<U00E8>	LATIN SMALL LETTER E WITH GRAVE
4073	<e'>	<U00E9>	LATIN SMALL LETTER E WITH ACUTE
4074	<e/>>	<U00EA>	LATIN SMALL LETTER E WITH CIRCUMFLEX
4075	<e:>	<U00EB>	LATIN SMALL LETTER E WITH DIAERESIS
4076	<i!>	<U00EC>	LATIN SMALL LETTER I WITH GRAVE
4077	<i'>	<U00ED>	LATIN SMALL LETTER I WITH ACUTE
4078	<i/>>	<U00EE>	LATIN SMALL LETTER I WITH CIRCUMFLEX
4079	<i:>	<U00EF>	LATIN SMALL LETTER I WITH DIAERESIS
4080	<d->	<U00F0>	LATIN SMALL LETTER ETH (Icelandic)
4081	<n?>	<U00F1>	LATIN SMALL LETTER N WITH TILDE
4082	<o!>	<U00F2>	LATIN SMALL LETTER O WITH GRAVE
4083	<o'>	<U00F3>	LATIN SMALL LETTER O WITH ACUTE
4084	<o/>>	<U00F4>	LATIN SMALL LETTER O WITH CIRCUMFLEX
4085	<o?>	<U00F5>	LATIN SMALL LETTER O WITH TILDE
4086	<o:>	<U00F6>	LATIN SMALL LETTER O WITH DIAERESIS
4087	<-:>	<U00F7>	DIVISION SIGN
4088	<o//>	<U00F8>	LATIN SMALL LETTER O WITH STROKE
4089	<u!>	<U00F9>	LATIN SMALL LETTER U WITH GRAVE
4090	<u'>	<U00FA>	LATIN SMALL LETTER U WITH ACUTE
4091	<u/>>	<U00FB>	LATIN SMALL LETTER U WITH CIRCUMFLEX
4092	<u:>	<U00FC>	LATIN SMALL LETTER U WITH DIAERESIS
4093	<y'>	<U00FD>	LATIN SMALL LETTER Y WITH ACUTE
4094	<th>	<U00FE>	LATIN SMALL LETTER THORN (Icelandic)
4095	<y:>	<U00FF>	LATIN SMALL LETTER Y WITH DIAERESIS
4096	<A->	<U0100>	LATIN CAPITAL LETTER A WITH MACRON
4097	<a->	<U0101>	LATIN SMALL LETTER A WITH MACRON
4098	<A>	<U0102>	LATIN CAPITAL LETTER A WITH BREVE
4099	<a>	<U0103>	LATIN SMALL LETTER A WITH BREVE
4100	<A; >	<U0104>	LATIN CAPITAL LETTER A WITH OGONEK
4101	<a; >	<U0105>	LATIN SMALL LETTER A WITH OGONEK
4102	<C'>	<U0106>	LATIN CAPITAL LETTER C WITH ACUTE
4103	<c'>	<U0107>	LATIN SMALL LETTER C WITH ACUTE
4104	<C/>>	<U0108>	LATIN CAPITAL LETTER C WITH CIRCUMFLEX
4105	<c/>>	<U0109>	LATIN SMALL LETTER C WITH CIRCUMFLEX
4106	<C.>	<U010A>	LATIN CAPITAL LETTER C WITH DOT ABOVE
4107	<c.>	<U010B>	LATIN SMALL LETTER C WITH DOT ABOVE
4108	<C<>	<U010C>	LATIN CAPITAL LETTER C WITH CARON
4109	<c<>	<U010D>	LATIN SMALL LETTER C WITH CARON
4110	<D<>	<U010E>	LATIN CAPITAL LETTER D WITH CARON
4111	<d<>	<U010F>	LATIN SMALL LETTER D WITH CARON
4112	<D//>	<U0110>	LATIN CAPITAL LETTER D WITH STROKE
4113	<d//>	<U0111>	LATIN SMALL LETTER D WITH STROKE
4114	<E->	<U0112>	LATIN CAPITAL LETTER E WITH MACRON
4115	<e->	<U0113>	LATIN SMALL LETTER E WITH MACRON
4116	<E>	<U0114>	LATIN CAPITAL LETTER E WITH BREVE
4117	<e>	<U0115>	LATIN SMALL LETTER E WITH BREVE
4118	<E.>	<U0116>	LATIN CAPITAL LETTER E WITH DOT ABOVE
4119	<e.>	<U0117>	LATIN SMALL LETTER E WITH DOT ABOVE
4120	<E; >	<U0118>	LATIN CAPITAL LETTER E WITH OGONEK
4121	<e; >	<U0119>	LATIN SMALL LETTER E WITH OGONEK
4122	<E<>	<U011A>	LATIN CAPITAL LETTER E WITH CARON
4123	<e<>	<U011B>	LATIN SMALL LETTER E WITH CARON
4124	<G/>>	<U011C>	LATIN CAPITAL LETTER G WITH CIRCUMFLEX
4125	<g/>>	<U011D>	LATIN SMALL LETTER G WITH CIRCUMFLEX
4126	<G>	<U011E>	LATIN CAPITAL LETTER G WITH BREVE
4127	<g>	<U011F>	LATIN SMALL LETTER G WITH BREVE
4128	<G.>	<U0120>	LATIN CAPITAL LETTER G WITH DOT ABOVE
4129	<g.>	<U0121>	LATIN SMALL LETTER G WITH DOT ABOVE
4130	<G.>	<U0122>	LATIN CAPITAL LETTER G WITH CEDILLA
4131	<g.>	<U0123>	LATIN SMALL LETTER G WITH CEDILLA

4132	<H/>>	<U0124>	LATIN CAPITAL LETTER H WITH CIRCUMFLEX
4133	<h/>>	<U0125>	LATIN SMALL LETTER H WITH CIRCUMFLEX
4134	<H//>	<U0126>	LATIN CAPITAL LETTER H WITH STROKE
4135	<h//>	<U0127>	LATIN SMALL LETTER H WITH STROKE
4136	<I?>	<U0128>	LATIN CAPITAL LETTER I WITH TILDE
4137	<i?>	<U0129>	LATIN SMALL LETTER I WITH TILDE
4138	<I->	<U012A>	LATIN CAPITAL LETTER I WITH MACRON
4139	<i->	<U012B>	LATIN SMALL LETTER I WITH MACRON
4140	<I(>	<U012C>	LATIN CAPITAL LETTER I WITH BREVE
4141	<i(>	<U012D>	LATIN SMALL LETTER I WITH BREVE
4142	<I; >	<U012E>	LATIN CAPITAL LETTER I WITH OGONEK
4143	<i; >	<U012F>	LATIN SMALL LETTER I WITH OGONEK
4144	<I. >	<U0130>	LATIN CAPITAL LETTER I WITH DOT ABOVE
4145	<i. >	<U0131>	LATIN SMALL LETTER DOTLESS I
4146	<IJ>	<U0132>	LATIN CAPITAL LIGATURE IJ
4147	<ij>	<U0133>	LATIN SMALL LIGATURE IJ
4148	<J/>>	<U0134>	LATIN CAPITAL LETTER J WITH CIRCUMFLEX
4149	<j/>>	<U0135>	LATIN SMALL LETTER J WITH CIRCUMFLEX
4150	<K, >	<U0136>	LATIN CAPITAL LETTER K WITH CEDILLA
4151	<k, >	<U0137>	LATIN SMALL LETTER K WITH CEDILLA
4152	<kk>	<U0138>	LATIN SMALL LETTER KRA (Greenlandic)
4153	<L' >	<U0139>	LATIN CAPITAL LETTER L WITH ACUTE
4154	<l' >	<U013A>	LATIN SMALL LETTER L WITH ACUTE
4155	<L, >	<U013B>	LATIN CAPITAL LETTER L WITH CEDILLA
4156	<l, >	<U013C>	LATIN SMALL LETTER L WITH CEDILLA
4157	<L< >	<U013D>	LATIN CAPITAL LETTER L WITH CARON
4158	<l< >	<U013E>	LATIN SMALL LETTER L WITH CARON
4159	<L. >	<U013F>	LATIN CAPITAL LETTER L WITH MIDDLE DOT
4160	<l. >	<U0140>	LATIN SMALL LETTER L WITH MIDDLE DOT
4161	<L//>	<U0141>	LATIN CAPITAL LETTER L WITH STROKE
4162	<l//>	<U0142>	LATIN SMALL LETTER L WITH STROKE
4163	<N' >	<U0143>	LATIN CAPITAL LETTER N WITH ACUTE
4164	<n' >	<U0144>	LATIN SMALL LETTER N WITH ACUTE
4165	<N, >	<U0145>	LATIN CAPITAL LETTER N WITH CEDILLA
4166	<n, >	<U0146>	LATIN SMALL LETTER N WITH CEDILLA
4167	<N< >	<U0147>	LATIN CAPITAL LETTER N WITH CARON
4168	<n< >	<U0148>	LATIN SMALL LETTER N WITH CARON
4169	<'n>	<U0149>	LATIN SMALL LETTER N PRECEDED BY APOSTROPHE
4170	<NG>	<U014A>	LATIN CAPITAL LETTER ENG (Sami)
4171	<ng>	<U014B>	LATIN SMALL LETTER ENG (Sami)
4172	<O->	<U014C>	LATIN CAPITAL LETTER O WITH MACRON
4173	<o->	<U014D>	LATIN SMALL LETTER O WITH MACRON
4174	<O(>	<U014E>	LATIN CAPITAL LETTER O WITH BREVE
4175	<o(>	<U014F>	LATIN SMALL LETTER O WITH BREVE
4176	<O" >	<U0150>	LATIN CAPITAL LETTER O WITH DOUBLE ACUTE
4177	<o" >	<U0151>	LATIN SMALL LETTER O WITH DOUBLE ACUTE
4178	<OE>	<U0152>	LATIN CAPITAL LIGATURE OE
4179	<oe>	<U0153>	LATIN SMALL LIGATURE OE
4180	<R' >	<U0154>	LATIN CAPITAL LETTER R WITH ACUTE
4181	<r' >	<U0155>	LATIN SMALL LETTER R WITH ACUTE
4182	<R, >	<U0156>	LATIN CAPITAL LETTER R WITH CEDILLA
4183	<r, >	<U0157>	LATIN SMALL LETTER R WITH CEDILLA
4184	<R< >	<U0158>	LATIN CAPITAL LETTER R WITH CARON
4185	<r< >	<U0159>	LATIN SMALL LETTER R WITH CARON
4186	<S' >	<U015A>	LATIN CAPITAL LETTER S WITH ACUTE
4187	<s' >	<U015B>	LATIN SMALL LETTER S WITH ACUTE
4188	<S/>>	<U015C>	LATIN CAPITAL LETTER S WITH CIRCUMFLEX
4189	<s/>>	<U015D>	LATIN SMALL LETTER S WITH CIRCUMFLEX
4190	<S, >	<U015E>	LATIN CAPITAL LETTER S WITH CEDILLA
4191	<s, >	<U015F>	LATIN SMALL LETTER S WITH CEDILLA
4192	<S< >	<U0160>	LATIN CAPITAL LETTER S WITH CARON
4193	<s< >	<U0161>	LATIN SMALL LETTER S WITH CARON
4194	<T, >	<U0162>	LATIN CAPITAL LETTER T WITH CEDILLA
4195	<t, >	<U0163>	LATIN SMALL LETTER T WITH CEDILLA
4196	<T< >	<U0164>	LATIN CAPITAL LETTER T WITH CARON
4197	<t< >	<U0165>	LATIN SMALL LETTER T WITH CARON
4198	<T//>	<U0166>	LATIN CAPITAL LETTER T WITH STROKE
4199	<t//>	<U0167>	LATIN SMALL LETTER T WITH STROKE
4200	<U?>	<U0168>	LATIN CAPITAL LETTER U WITH TILDE
4201	<u?>	<U0169>	LATIN SMALL LETTER U WITH TILDE
4202	<U->	<U016A>	LATIN CAPITAL LETTER U WITH MACRON
4203	<u->	<U016B>	LATIN SMALL LETTER U WITH MACRON
4204	<U(>	<U016C>	LATIN CAPITAL LETTER U WITH BREVE
4205	<u(>	<U016D>	LATIN SMALL LETTER U WITH BREVE
4206	<UO>	<U016E>	LATIN CAPITAL LETTER U WITH RING ABOVE
4207	<uO>	<U016F>	LATIN SMALL LETTER U WITH RING ABOVE
4208	<U" >	<U0170>	LATIN CAPITAL LETTER U WITH DOUBLE ACUTE
4209	<u" >	<U0171>	LATIN SMALL LETTER U WITH DOUBLE ACUTE
4210	<U; >	<U0172>	LATIN CAPITAL LETTER U WITH OGONEK
4211	<u; >	<U0173>	LATIN SMALL LETTER U WITH OGONEK
4212	<W/>>	<U0174>	LATIN CAPITAL LETTER W WITH CIRCUMFLEX
4213	<w/>>	<U0175>	LATIN SMALL LETTER W WITH CIRCUMFLEX
4214	<Y/>>	<U0176>	LATIN CAPITAL LETTER Y WITH CIRCUMFLEX
4215	<y/>>	<U0177>	LATIN SMALL LETTER Y WITH CIRCUMFLEX
4216	<Y: >	<U0178>	LATIN CAPITAL LETTER Y WITH DIAERESIS
4217	<Z' >	<U0179>	LATIN CAPITAL LETTER Z WITH ACUTE
4218	<z' >	<U017A>	LATIN SMALL LETTER Z WITH ACUTE
4219	<Z. >	<U017B>	LATIN CAPITAL LETTER Z WITH DOT ABOVE
4220	<z. >	<U017C>	LATIN SMALL LETTER Z WITH DOT ABOVE

4221	<z<>	<U017D>	LATIN CAPITAL LETTER Z WITH CARON
4222	<z<>	<U017E>	LATIN SMALL LETTER Z WITH CARON
4223	<sl>	<U017F>	LATIN SMALL LETTER LONG S
4224	<b//>	<U0180>	LATIN SMALL LETTER B WITH STROKE
4225	<B2>	<U0181>	LATIN CAPITAL LETTER B WITH HOOK
4226	<C2>	<U0187>	LATIN CAPITAL LETTER C WITH HOOK
4227	<c2>	<U0188>	LATIN SMALL LETTER C WITH HOOK
4228	<F2>	<U0191>	LATIN CAPITAL LETTER F WITH HOOK
4229	<f2>	<U0192>	LATIN SMALL LETTER F WITH HOOK
4230	<K2>	<U0198>	LATIN CAPITAL LETTER K WITH HOOK
4231	<k2>	<U0199>	LATIN SMALL LETTER K WITH HOOK
4232	<O9>	<U01A0>	LATIN CAPITAL LETTER O WITH HORN
4233	<o9>	<U01A1>	LATIN SMALL LETTER O WITH HORN
4234	<OI>	<U01A2>	LATIN CAPITAL LETTER OI
4235	<oi>	<U01A3>	LATIN SMALL LETTER OI
4236	<yr>	<U01A6>	LATIN LETTER YR
4237	<U9>	<U01AF>	LATIN CAPITAL LETTER U WITH HORN
4238	<u9>	<U01B0>	LATIN SMALL LETTER U WITH HORN
4239	<Z//>	<U01B5>	LATIN CAPITAL LETTER Z WITH STROKE
4240	<z//>	<U01B6>	LATIN SMALL LETTER Z WITH STROKE
4241	<ED>	<U01B7>	LATIN CAPITAL LETTER EZH
4242	<DZ<>	<U01C4>	LATIN CAPITAL LETTER DZ WITH CARON
4243	<Dz<>	<U01C5>	LATIN CAPITAL LETTER D WITH SMALL LETTER Z WITH CARON
4244	<dz<>	<U01C6>	LATIN SMALL LETTER DZ WITH CARON
4245	<LJ3>	<U01C7>	LATIN CAPITAL LETTER LJ
4246	<Lj3>	<U01C8>	LATIN CAPITAL LETTER L WITH SMALL LETTER J
4247	<lj3>	<U01C9>	LATIN SMALL LETTER LJ
4248	<NJ3>	<U01CA>	LATIN CAPITAL LETTER NJ
4249	<Nj3>	<U01CB>	LATIN CAPITAL LETTER N WITH SMALL LETTER J
4250	<nj3>	<U01CC>	LATIN SMALL LETTER NJ
4251	<A<>	<U01CD>	LATIN CAPITAL LETTER A WITH CARON
4252	<a<>	<U01CE>	LATIN SMALL LETTER A WITH CARON
4253	<I<>	<U01CF>	LATIN CAPITAL LETTER I WITH CARON
4254	<i<>	<U01D0>	LATIN SMALL LETTER I WITH CARON
4255	<O<>	<U01D1>	LATIN CAPITAL LETTER O WITH CARON
4256	<o<>	<U01D2>	LATIN SMALL LETTER O WITH CARON
4257	<U<>	<U01D3>	LATIN CAPITAL LETTER U WITH CARON
4258	<u<>	<U01D4>	LATIN SMALL LETTER U WITH CARON
4259	<U:~>	<U01D5>	LATIN CAPITAL LETTER U WITH DIAERESIS AND MACRON
4260	<u:~>	<U01D6>	LATIN SMALL LETTER U WITH DIAERESIS AND MACRON
4261	<U:ˆ>	<U01D7>	LATIN CAPITAL LETTER U WITH DIAERESIS AND ACUTE
4262	<u:ˆ>	<U01D8>	LATIN SMALL LETTER U WITH DIAERESIS AND ACUTE
4263	<U:ˆ>	<U01D9>	LATIN CAPITAL LETTER U WITH DIAERESIS AND CARON
4264	<u:ˆ>	<U01DA>	LATIN SMALL LETTER U WITH DIAERESIS AND CARON
4265	<U:ˆ>	<U01DB>	LATIN CAPITAL LETTER U WITH DIAERESIS AND GRAVE
4266	<u:ˆ>	<U01DC>	LATIN SMALL LETTER U WITH DIAERESIS AND GRAVE
4267	<e1>	<U01DD>	LATIN SMALL LETTER TURNED E
4268	<A1>	<U01DE>	LATIN CAPITAL LETTER A WITH DIAERESIS AND MACRON
4269	<a1>	<U01DF>	LATIN SMALL LETTER A WITH DIAERESIS AND MACRON
4270	<A7>	<U01E0>	LATIN CAPITAL LETTER A WITH DOT ABOVE AND MACRON
4271	<a7>	<U01E1>	LATIN SMALL LETTER A WITH DOT ABOVE AND MACRON
4272	<A3>	<U01E2>	LATIN CAPITAL LETTER AE WITH MACRON (ash)
4273	<a3>	<U01E3>	LATIN SMALL LETTER AE WITH MACRON (ash)
4274	<G//>	<U01E4>	LATIN CAPITAL LETTER G WITH STROKE
4275	<g//>	<U01E5>	LATIN SMALL LETTER G WITH STROKE
4276	<G<>	<U01E6>	LATIN CAPITAL LETTER G WITH CARON
4277	<g<>	<U01E7>	LATIN SMALL LETTER G WITH CARON
4278	<K<>	<U01E8>	LATIN CAPITAL LETTER K WITH CARON
4279	<k<>	<U01E9>	LATIN SMALL LETTER K WITH CARON
4280	<O;ˆ>	<U01EA>	LATIN CAPITAL LETTER O WITH OGONEK
4281	<o;ˆ>	<U01EB>	LATIN SMALL LETTER O WITH OGONEK
4282	<O1>	<U01EC>	LATIN CAPITAL LETTER O WITH OGONEK AND MACRON
4283	<o1>	<U01ED>	LATIN SMALL LETTER O WITH OGONEK AND MACRON
4284	<EZ>	<U01EE>	LATIN CAPITAL LETTER EZH WITH CARON
4285	<ez>	<U01EF>	LATIN SMALL LETTER EZH WITH CARON
4286	<j<>	<U01F0>	LATIN SMALL LETTER J WITH CARON
4287	<DZ3>	<U01F1>	LATIN CAPITAL LETTER DZ
4288	<Dz3>	<U01F2>	LATIN CAPITAL LETTER D WITH SMALL LETTER Z
4289	<dz3>	<U01F3>	LATIN SMALL LETTER DZ
4290	<G'>	<U01F4>	LATIN CAPITAL LETTER G WITH ACUTE
4291	<g'>	<U01F5>	LATIN SMALL LETTER G WITH ACUTE
4292	<AA'>	<U01FA>	LATIN CAPITAL LETTER A WITH RING ABOVE AND ACUTE
4293	<aa'>	<U01FB>	LATIN SMALL LETTER A WITH RING ABOVE AND ACUTE
4294	<AE'>	<U01FC>	LATIN CAPITAL LETTER AE WITH ACUTE (ash)
4295	<ae'>	<U01FD>	LATIN SMALL LETTER AE WITH ACUTE (ash)
4296	<O//'>	<U01FE>	LATIN CAPITAL LETTER O WITH STROKE AND ACUTE
4297	<o//'>	<U01FF>	LATIN SMALL LETTER O WITH STROKE AND ACUTE
4298	<A!!>	<U0200>	LATIN CAPITAL LETTER A WITH DOUBLE GRAVE
4299	<a!!>	<U0201>	LATIN SMALL LETTER A WITH DOUBLE GRAVE
4300	<A)ˆ>	<U0202>	LATIN CAPITAL LETTER A WITH INVERTED BREVE
4301	<a)ˆ>	<U0203>	LATIN SMALL LETTER A WITH INVERTED BREVE
4302	<E!ˆ>	<U0204>	LATIN CAPITAL LETTER E WITH DOUBLE GRAVE
4303	<e!ˆ>	<U0205>	LATIN SMALL LETTER E WITH DOUBLE GRAVE
4304	<E)ˆ>	<U0206>	LATIN CAPITAL LETTER E WITH INVERTED BREVE
4305	<e)ˆ>	<U0207>	LATIN SMALL LETTER E WITH INVERTED BREVE
4306	<I!ˆ>	<U0208>	LATIN CAPITAL LETTER I WITH DOUBLE GRAVE
4307	<i!ˆ>	<U0209>	LATIN SMALL LETTER I WITH DOUBLE GRAVE
4308	<I)ˆ>	<U020A>	LATIN CAPITAL LETTER I WITH INVERTED BREVE
4309	<i)ˆ>	<U020B>	LATIN SMALL LETTER I WITH INVERTED BREVE

4310	<O! !>	<U020C>	LATIN CAPITAL LETTER O WITH DOUBLE GRAVE
4311	<o! !>	<U020D>	LATIN SMALL LETTER O WITH DOUBLE GRAVE
4312	<O>	<U020E>	LATIN CAPITAL LETTER O WITH INVERTED BREVE
4313	<o>	<U020F>	LATIN SMALL LETTER O WITH INVERTED BREVE
4314	<R! !>	<U0210>	LATIN CAPITAL LETTER R WITH DOUBLE GRAVE
4315	<r! !>	<U0211>	LATIN SMALL LETTER R WITH DOUBLE GRAVE
4316	<R>	<U0212>	LATIN CAPITAL LETTER R WITH INVERTED BREVE
4317	<r>	<U0213>	LATIN SMALL LETTER R WITH INVERTED BREVE
4318	<U! !>	<U0214>	LATIN CAPITAL LETTER U WITH DOUBLE GRAVE
4319	<u! !>	<U0215>	LATIN SMALL LETTER U WITH DOUBLE GRAVE
4320	<U>	<U0216>	LATIN CAPITAL LETTER U WITH INVERTED BREVE
4321	<u>	<U0217>	LATIN SMALL LETTER U WITH INVERTED BREVE
4322	<rl>	<U027C>	LATIN SMALL LETTER R WITH LONG LEG
4323	<ed>	<U0292>	LATIN SMALL LETTER EZH
4324	<;S>	<U02BB>	MODIFIER LETTER TURNED COMMA
4325	<1/>	<U02C6>	MODIFIER LETTER CIRCUMFLEX ACCENT
4326	<'<>	<U02C7>	CARON (Mandarin Chinese third tone)
4327	<1->	<U02C9>	MODIFIER LETTER MACRON (Mandarin Chinese first tone)
4328	<1!>	<U02CB>	MODIFIER LETTER GRAVE ACCENT (Mandarin Chinese fourth tone)
4329	<'(>	<U02D8>	BREVE
4330	<'>	<U02D9>	DOT ABOVE (Mandarin Chinese light tone)
4331	<'0>	<U02DA>	RING ABOVE
4332	<'>	<U02DB>	OGONEK
4333	<1?>	<U02DC>	SMALL TILDE
4334	<' " >	<U02DD>	DOUBLE ACUTE ACCENT
4335	<'G>	<U0374>	GREEK NUMERAL SIGN (Dexia keraia)
4336	<,G>	<U0375>	GREEK LOWER NUMERAL SIGN (Aristeri keraia)
4337	<j3>	<U037A>	GREEK YPOGEGRAMMENI
4338	<?%>	<U037E>	GREEK QUESTION MARK (Erotimatiko)
4339	<' * >	<U0384>	GREEK TONOS
4340	<' % >	<U0385>	GREEK DIALYTIKA TONOS
4341	<A%>	<U0386>	GREEK CAPITAL LETTER ALPHA WITH TONOS
4342	<. * >	<U0387>	GREEK ANO TELEIA
4343	<E%>	<U0388>	GREEK CAPITAL LETTER EPSILON WITH TONOS
4344	<Y%>	<U0389>	GREEK CAPITAL LETTER ETA WITH TONOS
4345	<I%>	<U038A>	GREEK CAPITAL LETTER IOTA WITH TONOS
4346	<O%>	<U038C>	GREEK CAPITAL LETTER OMICRON WITH TONOS
4347	<U%>	<U038E>	GREEK CAPITAL LETTER UPSILON WITH TONOS
4348	<W%>	<U038F>	GREEK CAPITAL LETTER OMEGA WITH TONOS
4349	<i3>	<U0390>	GREEK SMALL LETTER IOTA WITH DIALYTIKA AND TONOS
4350	<A* >	<U0391>	GREEK CAPITAL LETTER ALPHA
4351	<B* >	<U0392>	GREEK CAPITAL LETTER BETA
4352	<G* >	<U0393>	GREEK CAPITAL LETTER GAMMA
4353	<D* >	<U0394>	GREEK CAPITAL LETTER DELTA
4354	<E* >	<U0395>	GREEK CAPITAL LETTER EPSILON
4355	<Z* >	<U0396>	GREEK CAPITAL LETTER ZETA
4356	<Y* >	<U0397>	GREEK CAPITAL LETTER ETA
4357	<H* >	<U0398>	GREEK CAPITAL LETTER THETA
4358	<I* >	<U0399>	GREEK CAPITAL LETTER IOTA
4359	<K* >	<U039A>	GREEK CAPITAL LETTER KAPPA
4360	<L* >	<U039B>	GREEK CAPITAL LETTER LAMDA
4361	<M* >	<U039C>	GREEK CAPITAL LETTER MU
4362	<N* >	<U039D>	GREEK CAPITAL LETTER NU
4363	<C* >	<U039E>	GREEK CAPITAL LETTER XI
4364	<O* >	<U039F>	GREEK CAPITAL LETTER OMICRON
4365	<P* >	<U03A0>	GREEK CAPITAL LETTER PI
4366	<R* >	<U03A1>	GREEK CAPITAL LETTER RHO
4367	<S* >	<U03A3>	GREEK CAPITAL LETTER SIGMA
4368	<T* >	<U03A4>	GREEK CAPITAL LETTER TAU
4369	<U* >	<U03A5>	GREEK CAPITAL LETTER UPSILON
4370	<F* >	<U03A6>	GREEK CAPITAL LETTER PHI
4371	<X* >	<U03A7>	GREEK CAPITAL LETTER CHI
4372	<Q* >	<U03A8>	GREEK CAPITAL LETTER PSI
4373	<W* >	<U03A9>	GREEK CAPITAL LETTER OMEGA
4374	<J* >	<U03AA>	GREEK CAPITAL LETTER IOTA WITH DIALYTIKA
4375	<V* >	<U03AB>	GREEK CAPITAL LETTER UPSILON WITH DIALYTIKA
4376	<a%>	<U03AC>	GREEK SMALL LETTER ALPHA WITH TONOS
4377	<e%>	<U03AD>	GREEK SMALL LETTER EPSILON WITH TONOS
4378	<y%>	<U03AE>	GREEK SMALL LETTER ETA WITH TONOS
4379	<i%>	<U03AF>	GREEK SMALL LETTER IOTA WITH TONOS
4380	<u3>	<U03B0>	GREEK SMALL LETTER UPSILON WITH DIALYTIKA AND TONOS
4381	<a* >	<U03B1>	GREEK SMALL LETTER ALPHA
4382	<b* >	<U03B2>	GREEK SMALL LETTER BETA
4383	<g* >	<U03B3>	GREEK SMALL LETTER GAMMA
4384	<d* >	<U03B4>	GREEK SMALL LETTER DELTA
4385	<e* >	<U03B5>	GREEK SMALL LETTER EPSILON
4386	<z* >	<U03B6>	GREEK SMALL LETTER ZETA
4387	<y* >	<U03B7>	GREEK SMALL LETTER ETA
4388	<h* >	<U03B8>	GREEK SMALL LETTER THETA
4389	<i* >	<U03B9>	GREEK SMALL LETTER IOTA
4390	<k* >	<U03BA>	GREEK SMALL LETTER KAPPA
4391	<l* >	<U03BB>	GREEK SMALL LETTER LAMDA
4392	<m* >	<U03BC>	GREEK SMALL LETTER MU
4393	<n* >	<U03BD>	GREEK SMALL LETTER NU
4394	<c* >	<U03BE>	GREEK SMALL LETTER XI
4395	<o* >	<U03BF>	GREEK SMALL LETTER OMICRON
4396	<p* >	<U03C0>	GREEK SMALL LETTER PI
4397	<r* >	<U03C1>	GREEK SMALL LETTER RHO
4398	<*s>	<U03C2>	GREEK SMALL LETTER FINAL SIGMA

4399	<s*>	<U03C3>	GREEK SMALL LETTER SIGMA
4400	<t*>	<U03C4>	GREEK SMALL LETTER TAU
4401	<u*>	<U03C5>	GREEK SMALL LETTER UPSILON
4402	<f*>	<U03C6>	GREEK SMALL LETTER PHI
4403	<x*>	<U03C7>	GREEK SMALL LETTER CHI
4404	<q*>	<U03C8>	GREEK SMALL LETTER PSI
4405	<w*>	<U03C9>	GREEK SMALL LETTER OMEGA
4406	<j*>	<U03CA>	GREEK SMALL LETTER IOTA WITH DIALYTIKA
4407	<v*>	<U03CB>	GREEK SMALL LETTER UPSILON WITH DIALYTIKA
4408	<o%>	<U03CC>	GREEK SMALL LETTER OMICRON WITH TONOS
4409	<u%>	<U03CD>	GREEK SMALL LETTER UPSILON WITH TONOS
4410	<w%>	<U03CE>	GREEK SMALL LETTER OMEGA WITH TONOS
4411	<b3>	<U03D0>	GREEK BETA SYMBOL
4412	<T3>	<U03DA>	GREEK LETTER STIGMA
4413	<M3>	<U03DC>	GREEK LETTER DIGAMMA
4414	<K3>	<U03DE>	GREEK LETTER KOPPA
4415	<P3>	<U03E0>	GREEK LETTER SAMPI
4416	<IO>	<U0401>	CYRILLIC CAPITAL LETTER IO
4417	<D%>	<U0402>	CYRILLIC CAPITAL LETTER DJE (Serbocroatian)
4418	<G%>	<U0403>	CYRILLIC CAPITAL LETTER GJE
4419	<IE>	<U0404>	CYRILLIC CAPITAL LETTER UKRAINIAN IE
4420	<DS>	<U0405>	CYRILLIC CAPITAL LETTER DZE
4421	<II>	<U0406>	CYRILLIC CAPITAL LETTER BYELORUSSIAN-UKRAINIAN I
4422	<YI>	<U0407>	CYRILLIC CAPITAL LETTER YI (Ukrainian)
4423	<J%>	<U0408>	CYRILLIC CAPITAL LETTER JE
4424	<LJ>	<U0409>	CYRILLIC CAPITAL LETTER LJJE
4425	<NJ>	<U040A>	CYRILLIC CAPITAL LETTER NJJE
4426	<Ts>	<U040B>	CYRILLIC CAPITAL LETTER TSHE (Serbocroatian)
4427	<KJ>	<U040C>	CYRILLIC CAPITAL LETTER KJE
4428	<V%>	<U040E>	CYRILLIC CAPITAL LETTER SHORT U (Byelorussian)
4429	<DZ>	<U040F>	CYRILLIC CAPITAL LETTER DZHE
4430	<A=>	<U0410>	CYRILLIC CAPITAL LETTER A
4431	<B=>	<U0411>	CYRILLIC CAPITAL LETTER BE
4432	<V=>	<U0412>	CYRILLIC CAPITAL LETTER VE
4433	<G=>	<U0413>	CYRILLIC CAPITAL LETTER GHE
4434	<D=>	<U0414>	CYRILLIC CAPITAL LETTER DE
4435	<E=>	<U0415>	CYRILLIC CAPITAL LETTER IE
4436	<Z%>	<U0416>	CYRILLIC CAPITAL LETTER ZHE
4437	<Z=>	<U0417>	CYRILLIC CAPITAL LETTER ZE
4438	<I=>	<U0418>	CYRILLIC CAPITAL LETTER I
4439	<J=>	<U0419>	CYRILLIC CAPITAL LETTER SHORT I
4440	<K=>	<U041A>	CYRILLIC CAPITAL LETTER KA
4441	<L=>	<U041B>	CYRILLIC CAPITAL LETTER EL
4442	<M=>	<U041C>	CYRILLIC CAPITAL LETTER EM
4443	<N=>	<U041D>	CYRILLIC CAPITAL LETTER EN
4444	<O=>	<U041E>	CYRILLIC CAPITAL LETTER O
4445	<P=>	<U041F>	CYRILLIC CAPITAL LETTER PE
4446	<R=>	<U0420>	CYRILLIC CAPITAL LETTER ER
4447	<S=>	<U0421>	CYRILLIC CAPITAL LETTER ES
4448	<T=>	<U0422>	CYRILLIC CAPITAL LETTER TE
4449	<U=>	<U0423>	CYRILLIC CAPITAL LETTER U
4450	<F=>	<U0424>	CYRILLIC CAPITAL LETTER EF
4451	<H=>	<U0425>	CYRILLIC CAPITAL LETTER HA
4452	<C=>	<U0426>	CYRILLIC CAPITAL LETTER TSE
4453	<C%>	<U0427>	CYRILLIC CAPITAL LETTER CHE
4454	<S%>	<U0428>	CYRILLIC CAPITAL LETTER SHA
4455	<Sc>	<U0429>	CYRILLIC CAPITAL LETTER SHCHA
4456	<=">	<U042A>	CYRILLIC CAPITAL LETTER HARD SIGN
4457	<Y=>	<U042B>	CYRILLIC CAPITAL LETTER YERU
4458	<% ">	<U042C>	CYRILLIC CAPITAL LETTER SOFT SIGN
4459	<JE>	<U042D>	CYRILLIC CAPITAL LETTER E
4460	<JU>	<U042E>	CYRILLIC CAPITAL LETTER YU
4461	<JA>	<U042F>	CYRILLIC CAPITAL LETTER YA
4462	<a=>	<U0430>	CYRILLIC SMALL LETTER A
4463	<b=>	<U0431>	CYRILLIC SMALL LETTER BE
4464	<v=>	<U0432>	CYRILLIC SMALL LETTER VE
4465	<g=>	<U0433>	CYRILLIC SMALL LETTER GHE
4466	<d=>	<U0434>	CYRILLIC SMALL LETTER DE
4467	<e=>	<U0435>	CYRILLIC SMALL LETTER IE
4468	<z%>	<U0436>	CYRILLIC SMALL LETTER ZHE
4469	<z=>	<U0437>	CYRILLIC SMALL LETTER ZE
4470	<i=>	<U0438>	CYRILLIC SMALL LETTER I
4471	<j=>	<U0439>	CYRILLIC SMALL LETTER SHORT I
4472	<k=>	<U043A>	CYRILLIC SMALL LETTER KA
4473	<l=>	<U043B>	CYRILLIC SMALL LETTER EL
4474	<m=>	<U043C>	CYRILLIC SMALL LETTER EM
4475	<n=>	<U043D>	CYRILLIC SMALL LETTER EN
4476	<o=>	<U043E>	CYRILLIC SMALL LETTER O
4477	<p=>	<U043F>	CYRILLIC SMALL LETTER PE
4478	<r=>	<U0440>	CYRILLIC SMALL LETTER ER
4479	<s=>	<U0441>	CYRILLIC SMALL LETTER ES
4480	<t=>	<U0442>	CYRILLIC SMALL LETTER TE
4481	<u=>	<U0443>	CYRILLIC SMALL LETTER U
4482	<f=>	<U0444>	CYRILLIC SMALL LETTER EF
4483	<h=>	<U0445>	CYRILLIC SMALL LETTER HA
4484	<c=>	<U0446>	CYRILLIC SMALL LETTER TSE
4485	<c%>	<U0447>	CYRILLIC SMALL LETTER CHE
4486	<s%>	<U0448>	CYRILLIC SMALL LETTER SHA
4487	<sc>	<U0449>	CYRILLIC SMALL LETTER SHCHA

4488	<='>	<U044A>	CYRILLIC SMALL LETTER HARD SIGN
4489	<y=>	<U044B>	CYRILLIC SMALL LETTER YERU
4490	<% '>	<U044C>	CYRILLIC SMALL LETTER SOFT SIGN
4491	<je>	<U044D>	CYRILLIC SMALL LETTER E
4492	<ju>	<U044E>	CYRILLIC SMALL LETTER YU
4493	<ja>	<U044F>	CYRILLIC SMALL LETTER YA
4494	<io>	<U0451>	CYRILLIC SMALL LETTER IO
4495	<d%>	<U0452>	CYRILLIC SMALL LETTER DJE (Serbocroatian)
4496	<g%>	<U0453>	CYRILLIC SMALL LETTER GJE
4497	<ie>	<U0454>	CYRILLIC SMALL LETTER UKRAINIAN IE
4498	<ds>	<U0455>	CYRILLIC SMALL LETTER DZE
4499	<ii>	<U0456>	CYRILLIC SMALL LETTER BYELORUSSIAN-UKRAINIAN I
4500	<yi>	<U0457>	CYRILLIC SMALL LETTER YI (Ukrainian)
4501	<j%>	<U0458>	CYRILLIC SMALL LETTER JE
4502	<lj>	<U0459>	CYRILLIC SMALL LETTER LJJE
4503	<nj>	<U045A>	CYRILLIC SMALL LETTER NJJE
4504	<ts>	<U045B>	CYRILLIC SMALL LETTER TSHE (Serbocroatian)
4505	<kj>	<U045C>	CYRILLIC SMALL LETTER KJE
4506	<v%>	<U045E>	CYRILLIC SMALL LETTER SHORT U (Byelorussian)
4507	<dz>	<U045F>	CYRILLIC SMALL LETTER DZHE
4508	<Y3>	<U0462>	CYRILLIC CAPITAL LETTER YAT
4509	<y3>	<U0463>	CYRILLIC SMALL LETTER YAT
4510	<O3>	<U046A>	CYRILLIC CAPITAL LETTER BIG YUS
4511	<o3>	<U046B>	CYRILLIC SMALL LETTER BIG YUS
4512	<F3>	<U0472>	CYRILLIC CAPITAL LETTER FITA
4513	<f3>	<U0473>	CYRILLIC SMALL LETTER FITA
4514	<V3>	<U0474>	CYRILLIC CAPITAL LETTER IZHITSA
4515	<v3>	<U0475>	CYRILLIC SMALL LETTER IZHITSA
4516	<C3>	<U0480>	CYRILLIC CAPITAL LETTER KOPPA
4517	<c3>	<U0481>	CYRILLIC SMALL LETTER KOPPA
4518	<G3>	<U0490>	CYRILLIC CAPITAL LETTER GHE WITH UPTURN
4519	<g3>	<U0491>	CYRILLIC SMALL LETTER GHE WITH UPTURN
4520	<A+>	<U05D0>	HEBREW LETTER ALEF
4521	<B+>	<U05D1>	HEBREW LETTER BET
4522	<G+>	<U05D2>	HEBREW LETTER GIMEL
4523	<D+>	<U05D3>	HEBREW LETTER DALET
4524	<H+>	<U05D4>	HEBREW LETTER HE
4525	<W+>	<U05D5>	HEBREW LETTER VAV
4526	<Z+>	<U05D6>	HEBREW LETTER ZAYIN
4527	<X+>	<U05D7>	HEBREW LETTER HET
4528	<Tj>	<U05D8>	HEBREW LETTER TET
4529	<J+>	<U05D9>	HEBREW LETTER YOD
4530	<K%>	<U05DA>	HEBREW LETTER FINAL KAF
4531	<K+>	<U05DB>	HEBREW LETTER KAF
4532	<L+>	<U05DC>	HEBREW LETTER LAMED
4533	<M%>	<U05DD>	HEBREW LETTER FINAL MEM
4534	<M+>	<U05DE>	HEBREW LETTER MEM
4535	<N%>	<U05DF>	HEBREW LETTER FINAL NUN
4536	<N+>	<U05E0>	HEBREW LETTER NUN
4537	<S+>	<U05E1>	HEBREW LETTER SAMEKH
4538	<E+>	<U05E2>	HEBREW LETTER AYIN
4539	<P%>	<U05E3>	HEBREW LETTER FINAL PE
4540	<P+>	<U05E4>	HEBREW LETTER PE
4541	<Zj>	<U05E5>	HEBREW LETTER FINAL TSADI
4542	<Zj>	<U05E6>	HEBREW LETTER TSADI
4543	<Q+>	<U05E7>	HEBREW LETTER QOF
4544	<R+>	<U05E8>	HEBREW LETTER RESH
4545	<Sh>	<U05E9>	HEBREW LETTER SHIN
4546	<T+>	<U05EA>	HEBREW LETTER TAV
4547	<,+>	<U060C>	ARABIC COMMA
4548	< ; +>	<U061B>	ARABIC SEMICOLON
4549	< ? +>	<U061F>	ARABIC QUESTION MARK
4550	<H'>	<U0621>	ARABIC LETTER HAMZA
4551	<aM>	<U0622>	ARABIC LETTER ALEF WITH MADDA ABOVE
4552	<aH>	<U0623>	ARABIC LETTER ALEF WITH HAMZA ABOVE
4553	<wH>	<U0624>	ARABIC LETTER WAW WITH HAMZA ABOVE
4554	<ah>	<U0625>	ARABIC LETTER ALEF WITH HAMZA BELOW
4555	<yH>	<U0626>	ARABIC LETTER YEH WITH HAMZA ABOVE
4556	<a+>	<U0627>	ARABIC LETTER ALEF
4557	<b+>	<U0628>	ARABIC LETTER BEH
4558	<tm>	<U0629>	ARABIC LETTER TEH MARBUTA
4559	<t+>	<U062A>	ARABIC LETTER TEH
4560	<tk>	<U062B>	ARABIC LETTER THEH
4561	<g+>	<U062C>	ARABIC LETTER JEEM
4562	<hk>	<U062D>	ARABIC LETTER HAH
4563	<x+>	<U062E>	ARABIC LETTER KHAH
4564	<d+>	<U062F>	ARABIC LETTER DAL
4565	<dk>	<U0630>	ARABIC LETTER THAL
4566	<r+>	<U0631>	ARABIC LETTER REH
4567	<z+>	<U0632>	ARABIC LETTER ZAIN
4568	<s+>	<U0633>	ARABIC LETTER SEEN
4569	<sn>	<U0634>	ARABIC LETTER SHEEN
4570	<c+>	<U0635>	ARABIC LETTER SAD
4571	<dd>	<U0636>	ARABIC LETTER DAD
4572	<tj>	<U0637>	ARABIC LETTER TAH
4573	<zH>	<U0638>	ARABIC LETTER ZAH
4574	<e+>	<U0639>	ARABIC LETTER AIN
4575	<i+>	<U063A>	ARABIC LETTER GHAIN
4576	<+>	<U0640>	ARABIC TATWEEL

4577	<f+>	<U0641>	ARABIC LETTER FEH
4578	<q+>	<U0642>	ARABIC LETTER QAF
4579	<k+>	<U0643>	ARABIC LETTER KAF
4580	<l+>	<U0644>	ARABIC LETTER LAM
4581	<m+>	<U0645>	ARABIC LETTER MEEM
4582	<n+>	<U0646>	ARABIC LETTER NOON
4583	<h+>	<U0647>	ARABIC LETTER HEH
4584	<w+>	<U0648>	ARABIC LETTER WAW
4585	<j+>	<U0649>	ARABIC LETTER ALEF MAKSURA
4586	<y+>	<U064A>	ARABIC LETTER YEH
4587	<: +>	<U064B>	ARABIC FATHATAN
4588	<" +>	<U064C>	ARABIC DAMMATAN
4589	<= +>	<U064D>	ARABIC KASRATAN
4590	< / / +>	<U064E>	ARABIC FATHA
4591	<' +>	<U064F>	ARABIC DAMMA
4592	<l+>	<U0650>	ARABIC KASRA
4593	<3+>	<U0651>	ARABIC SHADDA
4594	<0+>	<U0652>	ARABIC SUKUN
4595	<0a>	<U0660>	ARABIC-INDIC DIGIT ZERO
4596	<1a>	<U0661>	ARABIC-INDIC DIGIT ONE
4597	<2a>	<U0662>	ARABIC-INDIC DIGIT TWO
4598	<3a>	<U0663>	ARABIC-INDIC DIGIT THREE
4599	<4a>	<U0664>	ARABIC-INDIC DIGIT FOUR
4600	<5a>	<U0665>	ARABIC-INDIC DIGIT FIVE
4601	<6a>	<U0666>	ARABIC-INDIC DIGIT SIX
4602	<7a>	<U0667>	ARABIC-INDIC DIGIT SEVEN
4603	<8a>	<U0668>	ARABIC-INDIC DIGIT EIGHT
4604	<9a>	<U0669>	ARABIC-INDIC DIGIT NINE
4605	<aS>	<U0670>	ARABIC LETTER SUPERScript ALEF
4606	<p+>	<U067E>	ARABIC LETTER PEH
4607	<hH>	<U0681>	ARABIC LETTER HAH WITH HAMZA ABOVE
4608	<tc>	<U0686>	ARABIC LETTER TCHEH
4609	<zj>	<U0698>	ARABIC LETTER JEH
4610	<v+>	<U06A4>	ARABIC LETTER VEH
4611	<gf>	<U06AF>	ARABIC LETTER GAF
4612	<A-0>	<U1E00>	LATIN CAPITAL LETTER A WITH RING BELOW
4613	<a-0>	<U1E01>	LATIN SMALL LETTER A WITH RING BELOW
4614	<B .>	<U1E02>	LATIN CAPITAL LETTER B WITH DOT ABOVE
4615	<b .>	<U1E03>	LATIN SMALL LETTER B WITH DOT ABOVE
4616	<B- .>	<U1E04>	LATIN CAPITAL LETTER B WITH DOT BELOW
4617	<b- .>	<U1E05>	LATIN SMALL LETTER B WITH DOT BELOW
4618	<B_>	<U1E06>	LATIN CAPITAL LETTER B WITH LINE BELOW
4619	<b_>	<U1E07>	LATIN SMALL LETTER B WITH LINE BELOW
4620	<C, ' >	<U1E08>	LATIN CAPITAL LETTER C WITH CEDILLA AND ACUTE
4621	<c, ' >	<U1E09>	LATIN SMALL LETTER C WITH CEDILLA AND ACUTE
4622	<D .>	<U1E0A>	LATIN CAPITAL LETTER D WITH DOT ABOVE
4623	<d .>	<U1E0B>	LATIN SMALL LETTER D WITH DOT ABOVE
4624	<D- .>	<U1E0C>	LATIN CAPITAL LETTER D WITH DOT BELOW
4625	<d- .>	<U1E0D>	LATIN SMALL LETTER D WITH DOT BELOW
4626	<D_>	<U1E0E>	LATIN CAPITAL LETTER D WITH LINE BELOW
4627	<d_>	<U1E0F>	LATIN SMALL LETTER D WITH LINE BELOW
4628	<D, >	<U1E10>	LATIN CAPITAL LETTER D WITH CEDILLA
4629	<d, >	<U1E11>	LATIN SMALL LETTER D WITH CEDILLA
4630	<D- / >>	<U1E12>	LATIN CAPITAL LETTER D WITH CIRCUMFLEX BELOW
4631	<d- / >>	<U1E13>	LATIN SMALL LETTER D WITH CIRCUMFLEX BELOW
4632	<E- !>	<U1E14>	LATIN CAPITAL LETTER E WITH MACRON AND GRAVE
4633	<e- !>	<U1E15>	LATIN SMALL LETTER E WITH MACRON AND GRAVE
4634	<E- ' >	<U1E16>	LATIN CAPITAL LETTER E WITH MACRON AND ACUTE
4635	<e- ' >	<U1E17>	LATIN SMALL LETTER E WITH MACRON AND ACUTE
4636	<E- / >>	<U1E18>	LATIN CAPITAL LETTER E WITH CIRCUMFLEX BELOW
4637	<e- / >>	<U1E19>	LATIN SMALL LETTER E WITH CIRCUMFLEX BELOW
4638	<E- ?>	<U1E1A>	LATIN CAPITAL LETTER E WITH TILDE BELOW
4639	<e- ?>	<U1E1B>	LATIN SMALL LETTER E WITH TILDE BELOW
4640	<E, (>	<U1E1C>	LATIN CAPITAL LETTER E WITH CEDILLA AND BREVE
4641	<e, (>	<U1E1D>	LATIN SMALL LETTER E WITH CEDILLA AND BREVE
4642	<F .>	<U1E1E>	LATIN CAPITAL LETTER F WITH DOT ABOVE
4643	<f .>	<U1E1F>	LATIN SMALL LETTER F WITH DOT ABOVE
4644	<G- >	<U1E20>	LATIN CAPITAL LETTER G WITH MACRON
4645	<g- >	<U1E21>	LATIN SMALL LETTER G WITH MACRON
4646	<H .>	<U1E22>	LATIN CAPITAL LETTER H WITH DOT ABOVE
4647	<h .>	<U1E23>	LATIN SMALL LETTER H WITH DOT ABOVE
4648	<H- .>	<U1E24>	LATIN CAPITAL LETTER H WITH DOT BELOW
4649	<h- .>	<U1E25>	LATIN SMALL LETTER H WITH DOT BELOW
4650	<H: >	<U1E26>	LATIN CAPITAL LETTER H WITH DIAERESIS
4651	<h: >	<U1E27>	LATIN SMALL LETTER H WITH DIAERESIS
4652	<H, >	<U1E28>	LATIN CAPITAL LETTER H WITH CEDILLA
4653	<h, >	<U1E29>	LATIN SMALL LETTER H WITH CEDILLA
4654	<H- (>	<U1E2A>	LATIN CAPITAL LETTER H WITH BREVE BELOW
4655	<h- (>	<U1E2B>	LATIN SMALL LETTER H WITH BREVE BELOW
4656	<I- ?>	<U1E2C>	LATIN CAPITAL LETTER I WITH TILDE BELOW
4657	<i- ?>	<U1E2D>	LATIN SMALL LETTER I WITH TILDE BELOW
4658	<I: ' >	<U1E2E>	LATIN CAPITAL LETTER I WITH DIAERESIS AND ACUTE
4659	<i: ' >	<U1E2F>	LATIN SMALL LETTER I WITH DIAERESIS AND ACUTE
4660	<K' >	<U1E30>	LATIN CAPITAL LETTER K WITH ACUTE
4661	<k' >	<U1E31>	LATIN SMALL LETTER K WITH ACUTE
4662	<K- .>	<U1E32>	LATIN CAPITAL LETTER K WITH DOT BELOW
4663	<k- .>	<U1E33>	LATIN SMALL LETTER K WITH DOT BELOW
4664	<K_>	<U1E34>	LATIN CAPITAL LETTER K WITH LINE BELOW
4665	<k_>	<U1E35>	LATIN SMALL LETTER K WITH LINE BELOW

4666	<L- .>	<U1E36>	LATIN CAPITAL LETTER L WITH DOT BELOW
4667	<l- .>	<U1E37>	LATIN SMALL LETTER L WITH DOT BELOW
4668	<L- - .>	<U1E38>	LATIN CAPITAL LETTER L WITH DOT BELOW AND MACRON
4669	<l- - .>	<U1E39>	LATIN SMALL LETTER L WITH DOT BELOW AND MACRON
4670	<L_>	<U1E3A>	LATIN CAPITAL LETTER L WITH LINE BELOW
4671	<l_>	<U1E3B>	LATIN SMALL LETTER L WITH LINE BELOW
4672	<L- />>	<U1E3C>	LATIN CAPITAL LETTER L WITH CIRCUMFLEX BELOW
4673	<l- />>	<U1E3D>	LATIN SMALL LETTER L WITH CIRCUMFLEX BELOW
4674	<M'>	<U1E3E>	LATIN CAPITAL LETTER M WITH ACUTE
4675	<m'>	<U1E3F>	LATIN SMALL LETTER M WITH ACUTE
4676	<M .>	<U1E40>	LATIN CAPITAL LETTER M WITH DOT ABOVE
4677	<m .>	<U1E41>	LATIN SMALL LETTER M WITH DOT ABOVE
4678	<M- .>	<U1E42>	LATIN CAPITAL LETTER M WITH DOT BELOW
4679	<m- .>	<U1E43>	LATIN SMALL LETTER M WITH DOT BELOW
4680	<N .>	<U1E44>	LATIN CAPITAL LETTER N WITH DOT ABOVE
4681	<n .>	<U1E45>	LATIN SMALL LETTER N WITH DOT ABOVE
4682	<N- .>	<U1E46>	LATIN CAPITAL LETTER N WITH DOT BELOW
4683	<n- .>	<U1E47>	LATIN SMALL LETTER N WITH DOT BELOW
4684	<N_>	<U1E48>	LATIN CAPITAL LETTER N WITH LINE BELOW
4685	<n_>	<U1E49>	LATIN SMALL LETTER N WITH LINE BELOW
4686	<N- />>	<U1E4A>	LATIN CAPITAL LETTER N WITH CIRCUMFLEX BELOW
4687	<n- />>	<U1E4B>	LATIN SMALL LETTER N WITH CIRCUMFLEX BELOW
4688	<O? ' >	<U1E4C>	LATIN CAPITAL LETTER O WITH TILDE AND ACUTE
4689	<o? ' >	<U1E4D>	LATIN SMALL LETTER O WITH TILDE AND ACUTE
4690	<O? : >	<U1E4E>	LATIN CAPITAL LETTER O WITH TILDE AND DIAERESIS
4691	<o? : >	<U1E4F>	LATIN SMALL LETTER O WITH TILDE AND DIAERESIS
4692	<O- !>	<U1E50>	LATIN CAPITAL LETTER O WITH MACRON AND GRAVE
4693	<o- !>	<U1E51>	LATIN SMALL LETTER O WITH MACRON AND GRAVE
4694	<O- ' >	<U1E52>	LATIN CAPITAL LETTER O WITH MACRON AND ACUTE
4695	<o- ' >	<U1E53>	LATIN SMALL LETTER O WITH MACRON AND ACUTE
4696	<P' >	<U1E54>	LATIN CAPITAL LETTER P WITH ACUTE
4697	<p' >	<U1E55>	LATIN SMALL LETTER P WITH ACUTE
4698	<P .>	<U1E56>	LATIN CAPITAL LETTER P WITH DOT ABOVE
4699	<p .>	<U1E57>	LATIN SMALL LETTER P WITH DOT ABOVE
4700	<R .>	<U1E58>	LATIN CAPITAL LETTER R WITH DOT ABOVE
4701	<r .>	<U1E59>	LATIN SMALL LETTER R WITH DOT ABOVE
4702	<R- .>	<U1E5A>	LATIN CAPITAL LETTER R WITH DOT BELOW
4703	<r- .>	<U1E5B>	LATIN SMALL LETTER R WITH DOT BELOW
4704	<R- - .>	<U1E5C>	LATIN CAPITAL LETTER R WITH DOT BELOW AND MACRON
4705	<r- - .>	<U1E5D>	LATIN SMALL LETTER R WITH DOT BELOW AND MACRON
4706	<R_>	<U1E5E>	LATIN CAPITAL LETTER R WITH LINE BELOW
4707	<r_>	<U1E5F>	LATIN SMALL LETTER R WITH LINE BELOW
4708	<S .>	<U1E60>	LATIN CAPITAL LETTER S WITH DOT ABOVE
4709	<s .>	<U1E61>	LATIN SMALL LETTER S WITH DOT ABOVE
4710	<S- .>	<U1E62>	LATIN CAPITAL LETTER S WITH DOT BELOW
4711	<s- .>	<U1E63>	LATIN SMALL LETTER S WITH DOT BELOW
4712	<S' .>	<U1E64>	LATIN CAPITAL LETTER S WITH ACUTE AND DOT ABOVE
4713	<s' .>	<U1E65>	LATIN SMALL LETTER S WITH ACUTE AND DOT ABOVE
4714	<S< .>	<U1E66>	LATIN CAPITAL LETTER S WITH CARON AND DOT ABOVE
4715	<s< .>	<U1E67>	LATIN SMALL LETTER S WITH CARON AND DOT ABOVE
4716	<S- - .>	<U1E68>	LATIN CAPITAL LETTER S WITH DOT BELOW AND DOT ABOVE
4717	<s- - .>	<U1E69>	LATIN SMALL LETTER S WITH DOT BELOW AND DOT ABOVE
4718	<T .>	<U1E6A>	LATIN CAPITAL LETTER T WITH DOT ABOVE
4719	<t .>	<U1E6B>	LATIN SMALL LETTER T WITH DOT ABOVE
4720	<T- .>	<U1E6C>	LATIN CAPITAL LETTER T WITH DOT BELOW
4721	<t- .>	<U1E6D>	LATIN SMALL LETTER T WITH DOT BELOW
4722	<T_>	<U1E6E>	LATIN CAPITAL LETTER T WITH LINE BELOW
4723	<t_>	<U1E6F>	LATIN SMALL LETTER T WITH LINE BELOW
4724	<T- />>	<U1E70>	LATIN CAPITAL LETTER T WITH CIRCUMFLEX BELOW
4725	<t- />>	<U1E71>	LATIN SMALL LETTER T WITH CIRCUMFLEX BELOW
4726	<U- - :>	<U1E72>	LATIN CAPITAL LETTER U WITH DIAERESIS BELOW
4727	<u- - :>	<U1E73>	LATIN SMALL LETTER U WITH DIAERESIS BELOW
4728	<U- ?>	<U1E74>	LATIN CAPITAL LETTER U WITH TILDE BELOW
4729	<u- ?>	<U1E75>	LATIN SMALL LETTER U WITH TILDE BELOW
4730	<U- />>	<U1E76>	LATIN CAPITAL LETTER U WITH CIRCUMFLEX BELOW
4731	<u- />>	<U1E77>	LATIN SMALL LETTER U WITH CIRCUMFLEX BELOW
4732	<U? ' >	<U1E78>	LATIN CAPITAL LETTER U WITH TILDE AND ACUTE
4733	<u? ' >	<U1E79>	LATIN SMALL LETTER U WITH TILDE AND ACUTE
4734	<U- :>	<U1E7A>	LATIN CAPITAL LETTER U WITH MACRON AND DIAERESIS
4735	<u- :>	<U1E7B>	LATIN SMALL LETTER U WITH MACRON AND DIAERESIS
4736	<V? >	<U1E7C>	LATIN CAPITAL LETTER V WITH TILDE
4737	<v? >	<U1E7D>	LATIN SMALL LETTER V WITH TILDE
4738	<V- .>	<U1E7E>	LATIN CAPITAL LETTER V WITH DOT BELOW
4739	<v- .>	<U1E7F>	LATIN SMALL LETTER V WITH DOT BELOW
4740	<W!>	<U1E80>	LATIN CAPITAL LETTER W WITH GRAVE
4741	<w!>	<U1E81>	LATIN SMALL LETTER W WITH GRAVE
4742	<W' >	<U1E82>	LATIN CAPITAL LETTER W WITH ACUTE
4743	<w' >	<U1E83>	LATIN SMALL LETTER W WITH ACUTE
4744	<W :>	<U1E84>	LATIN CAPITAL LETTER W WITH DIAERESIS
4745	<w :>	<U1E85>	LATIN SMALL LETTER W WITH DIAERESIS
4746	<W .>	<U1E86>	LATIN CAPITAL LETTER W WITH DOT ABOVE
4747	<w .>	<U1E87>	LATIN SMALL LETTER W WITH DOT ABOVE
4748	<W- .>	<U1E88>	LATIN CAPITAL LETTER W WITH DOT BELOW
4749	<w- .>	<U1E89>	LATIN SMALL LETTER W WITH DOT BELOW
4750	<X .>	<U1E8A>	LATIN CAPITAL LETTER X WITH DOT ABOVE
4751	<x .>	<U1E8B>	LATIN SMALL LETTER X WITH DOT ABOVE
4752	<X :>	<U1E8C>	LATIN CAPITAL LETTER X WITH DIAERESIS
4753	<x :>	<U1E8D>	LATIN SMALL LETTER X WITH DIAERESIS
4754	<Y .>	<U1E8E>	LATIN CAPITAL LETTER Y WITH DOT ABOVE

4755	<y.>	<U1E8F>	LATIN SMALL LETTER Y WITH DOT ABOVE
4756	<Z/>>	<U1E90>	LATIN CAPITAL LETTER Z WITH CIRCUMFLEX
4757	<z/>>	<U1E91>	LATIN SMALL LETTER Z WITH CIRCUMFLEX
4758	<Z-.>	<U1E92>	LATIN CAPITAL LETTER Z WITH DOT BELOW
4759	<z-.>	<U1E93>	LATIN SMALL LETTER Z WITH DOT BELOW
4760	<Z_>	<U1E94>	LATIN CAPITAL LETTER Z WITH LINE BELOW
4761	<z_>	<U1E95>	LATIN SMALL LETTER Z WITH LINE BELOW
4762	<A-.>	<U1EA0>	LATIN CAPITAL LETTER A WITH DOT BELOW
4763	<a-.>	<U1EA1>	LATIN SMALL LETTER A WITH DOT BELOW
4764	<A2>	<U1EA2>	LATIN CAPITAL LETTER A WITH HOOK ABOVE
4765	<a2>	<U1EA3>	LATIN SMALL LETTER A WITH HOOK ABOVE
4766	<A/>'>	<U1EA4>	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND ACUTE
4767	<a/>'>	<U1EA5>	LATIN SMALL LETTER A WITH CIRCUMFLEX AND ACUTE
4768	<A/>!>	<U1EA6>	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND GRAVE
4769	<a/>!>	<U1EA7>	LATIN SMALL LETTER A WITH CIRCUMFLEX AND GRAVE
4770	<A/>2>	<U1EA8>	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND HOOK ABOVE
4771	<a/>2>	<U1EA9>	LATIN SMALL LETTER A WITH CIRCUMFLEX AND HOOK ABOVE
4772	<A/>?>	<U1EAA>	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND TILDE
4773	<a/>?>	<U1EAB>	LATIN SMALL LETTER A WITH CIRCUMFLEX AND TILDE
4774	<A/>-.>	<U1EAC>	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND DOT BELOW
4775	<a/>-.>	<U1EAD>	LATIN SMALL LETTER A WITH CIRCUMFLEX AND DOT BELOW
4776	<A(')>	<U1EAE>	LATIN CAPITAL LETTER A WITH BREVE AND ACUTE
4777	<a(')>	<U1EAF>	LATIN SMALL LETTER A WITH BREVE AND ACUTE
4778	<A(!)>	<U1EB0>	LATIN CAPITAL LETTER A WITH BREVE AND GRAVE
4779	<a(!)>	<U1EB1>	LATIN SMALL LETTER A WITH BREVE AND GRAVE
4780	<A(2)>	<U1EB2>	LATIN CAPITAL LETTER A WITH BREVE AND HOOK ABOVE
4781	<a(2)>	<U1EB3>	LATIN SMALL LETTER A WITH BREVE AND HOOK ABOVE
4782	<A(?)>	<U1EB4>	LATIN CAPITAL LETTER A WITH BREVE AND TILDE
4783	<a(?)>	<U1EB5>	LATIN SMALL LETTER A WITH BREVE AND TILDE
4784	<A(-.>	<U1EB6>	LATIN CAPITAL LETTER A WITH BREVE AND DOT BELOW
4785	<a(-.>	<U1EB7>	LATIN SMALL LETTER A WITH BREVE AND DOT BELOW
4786	<E-.>	<U1EB8>	LATIN CAPITAL LETTER E WITH DOT BELOW
4787	<e-.>	<U1EB9>	LATIN SMALL LETTER E WITH DOT BELOW
4788	<E2>	<U1EBA>	LATIN CAPITAL LETTER E WITH HOOK ABOVE
4789	<e2>	<U1EBB>	LATIN SMALL LETTER E WITH HOOK ABOVE
4790	<E?>	<U1EBC>	LATIN CAPITAL LETTER E WITH TILDE
4791	<e?>	<U1EBD>	LATIN SMALL LETTER E WITH TILDE
4792	<E/>'>	<U1EBE>	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND ACUTE
4793	<e/>'>	<U1EBF>	LATIN SMALL LETTER E WITH CIRCUMFLEX AND ACUTE
4794	<E/>!>	<U1EC0>	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND GRAVE
4795	<e/>!>	<U1EC1>	LATIN SMALL LETTER E WITH CIRCUMFLEX AND GRAVE
4796	<E/>2>	<U1EC2>	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND HOOK ABOVE
4797	<e/>2>	<U1EC3>	LATIN SMALL LETTER E WITH CIRCUMFLEX AND HOOK ABOVE
4798	<E/>?>	<U1EC4>	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND TILDE
4799	<e/>?>	<U1EC5>	LATIN SMALL LETTER E WITH CIRCUMFLEX AND TILDE
4800	<E/>-.>	<U1EC6>	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND DOT BELOW
4801	<e/>-.>	<U1EC7>	LATIN SMALL LETTER E WITH CIRCUMFLEX AND DOT BELOW
4802	<I2>	<U1EC8>	LATIN CAPITAL LETTER I WITH HOOK ABOVE
4803	<i2>	<U1EC9>	LATIN SMALL LETTER I WITH HOOK ABOVE
4804	<I-.>	<U1ECA>	LATIN CAPITAL LETTER I WITH DOT BELOW
4805	<i-.>	<U1ECB>	LATIN SMALL LETTER I WITH DOT BELOW
4806	<O-.>	<U1ECC>	LATIN CAPITAL LETTER O WITH DOT BELOW
4807	<o-.>	<U1ECD>	LATIN SMALL LETTER O WITH DOT BELOW
4808	<O2>	<U1ECE>	LATIN CAPITAL LETTER O WITH HOOK ABOVE
4809	<o2>	<U1ECF>	LATIN SMALL LETTER O WITH HOOK ABOVE
4810	<O/>'>	<U1ED0>	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND ACUTE
4811	<o/>'>	<U1ED1>	LATIN SMALL LETTER O WITH CIRCUMFLEX AND ACUTE
4812	<O/>!>	<U1ED2>	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND GRAVE
4813	<o/>!>	<U1ED3>	LATIN SMALL LETTER O WITH CIRCUMFLEX AND GRAVE
4814	<O/>2>	<U1ED4>	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND HOOK ABOVE
4815	<o/>2>	<U1ED5>	LATIN SMALL LETTER O WITH CIRCUMFLEX AND HOOK ABOVE
4816	<O/>?>	<U1ED6>	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND TILDE
4817	<o/>?>	<U1ED7>	LATIN SMALL LETTER O WITH CIRCUMFLEX AND TILDE
4818	<O/>-.>	<U1ED8>	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND DOT BELOW
4819	<o/>-.>	<U1ED9>	LATIN SMALL LETTER O WITH CIRCUMFLEX AND DOT BELOW
4820	<O9'>	<U1EDA>	LATIN CAPITAL LETTER O WITH HORN AND ACUTE
4821	<o9'>	<U1EDB>	LATIN SMALL LETTER O WITH HORN AND ACUTE
4822	<O9!>	<U1EDC>	LATIN CAPITAL LETTER O WITH HORN AND GRAVE
4823	<o9!>	<U1EDD>	LATIN SMALL LETTER O WITH HORN AND GRAVE
4824	<O92>	<U1EDE>	LATIN CAPITAL LETTER O WITH HORN AND HOOK ABOVE
4825	<o92>	<U1EDF>	LATIN SMALL LETTER O WITH HORN AND HOOK ABOVE
4826	<O9?>	<U1EE0>	LATIN CAPITAL LETTER O WITH HORN AND TILDE
4827	<o9?>	<U1EE1>	LATIN SMALL LETTER O WITH HORN AND TILDE
4828	<O9-.>	<U1EE2>	LATIN CAPITAL LETTER O WITH HORN AND DOT BELOW
4829	<o9-.>	<U1EE3>	LATIN SMALL LETTER O WITH HORN AND DOT BELOW
4830	<U-.>	<U1EE4>	LATIN CAPITAL LETTER U WITH DOT BELOW
4831	<u-.>	<U1EE5>	LATIN SMALL LETTER U WITH DOT BELOW
4832	<U2>	<U1EE6>	LATIN CAPITAL LETTER U WITH HOOK ABOVE
4833	<u2>	<U1EE7>	LATIN SMALL LETTER U WITH HOOK ABOVE
4834	<U9'>	<U1EE8>	LATIN CAPITAL LETTER U WITH HORN AND ACUTE
4835	<u9'>	<U1EE9>	LATIN SMALL LETTER U WITH HORN AND ACUTE
4836	<U9!>	<U1EEA>	LATIN CAPITAL LETTER U WITH HORN AND GRAVE
4837	<u9!>	<U1EEB>	LATIN SMALL LETTER U WITH HORN AND GRAVE
4838	<U92>	<U1EEC>	LATIN CAPITAL LETTER U WITH HORN AND HOOK ABOVE
4839	<u92>	<U1EED>	LATIN SMALL LETTER U WITH HORN AND HOOK ABOVE
4840	<U9?>	<U1EEE>	LATIN CAPITAL LETTER U WITH HORN AND TILDE
4841	<u9?>	<U1EEF>	LATIN SMALL LETTER U WITH HORN AND TILDE
4842	<U9-.>	<U1EF0>	LATIN CAPITAL LETTER U WITH HORN AND DOT BELOW
4843	<u9-.>	<U1EF1>	LATIN SMALL LETTER U WITH HORN AND DOT BELOW

4844	<Y!>	<U1EF2>	LATIN CAPITAL LETTER Y WITH GRAVE
4845	<y!>	<U1EF3>	LATIN SMALL LETTER Y WITH GRAVE
4846	<Y- >	<U1EF4>	LATIN CAPITAL LETTER Y WITH DOT BELOW
4847	<y- !>	<U1EF5>	LATIN SMALL LETTER Y WITH DOT BELOW
4848	<Y2>	<U1EF6>	LATIN CAPITAL LETTER Y WITH HOOK ABOVE
4849	<y2>	<U1EF7>	LATIN SMALL LETTER Y WITH HOOK ABOVE
4850	<Y?>	<U1EF8>	LATIN CAPITAL LETTER Y WITH TILDE
4851	<y?>	<U1EF9>	LATIN SMALL LETTER Y WITH TILDE
4852	<a*,>	<U1F00>	GREEK SMALL LETTER ALPHA WITH PSILI
4853	<a*;!>	<U1F01>	GREEK SMALL LETTER ALPHA WITH DASIA
4854	<a*;!>	<U1F02>	GREEK SMALL LETTER ALPHA WITH PSILI AND VARIA
4855	<a*;!>	<U1F03>	GREEK SMALL LETTER ALPHA WITH DASIA AND VARIA
4856	<a*,'>	<U1F04>	GREEK SMALL LETTER ALPHA WITH PSILI AND OXIA
4857	<a*,'>	<U1F05>	GREEK SMALL LETTER ALPHA WITH DASIA AND OXIA
4858	<a*,>	<U1F06>	GREEK SMALL LETTER ALPHA WITH PSILI AND PERISPOMENI
4859	<a*,>	<U1F07>	GREEK SMALL LETTER ALPHA WITH DASIA AND PERISPOMENI
4860	<A*,>	<U1F08>	GREEK CAPITAL LETTER ALPHA WITH PSILI
4861	<A*,>	<U1F09>	GREEK CAPITAL LETTER ALPHA WITH DASIA
4862	<A*;!>	<U1F0A>	GREEK CAPITAL LETTER ALPHA WITH PSILI AND VARIA
4863	<A*;!>	<U1F0B>	GREEK CAPITAL LETTER ALPHA WITH DASIA AND VARIA
4864	<A*,'>	<U1F0C>	GREEK CAPITAL LETTER ALPHA WITH PSILI AND OXIA
4865	<A*,'>	<U1F0D>	GREEK CAPITAL LETTER ALPHA WITH DASIA AND OXIA
4866	<A*,>	<U1F0E>	GREEK CAPITAL LETTER ALPHA WITH PSILI AND PERISPOMENI
4867	<A*,>	<U1F0F>	GREEK CAPITAL LETTER ALPHA WITH DASIA AND PERISPOMENI
4868	<e*,>	<U1F10>	GREEK SMALL LETTER EPSILON WITH PSILI
4869	<e*,>	<U1F11>	GREEK SMALL LETTER EPSILON WITH DASIA
4870	<e*;!>	<U1F12>	GREEK SMALL LETTER EPSILON WITH PSILI AND VARIA
4871	<e*;!>	<U1F13>	GREEK SMALL LETTER EPSILON WITH DASIA AND VARIA
4872	<e*,'>	<U1F14>	GREEK SMALL LETTER EPSILON WITH PSILI AND OXIA
4873	<e*,'>	<U1F15>	GREEK SMALL LETTER EPSILON WITH DASIA AND OXIA
4874	<E*,>	<U1F18>	GREEK CAPITAL LETTER EPSILON WITH PSILI
4875	<E*,>	<U1F19>	GREEK CAPITAL LETTER EPSILON WITH DASIA
4876	<E*;!>	<U1F1A>	GREEK CAPITAL LETTER EPSILON WITH PSILI AND VARIA
4877	<E*;!>	<U1F1B>	GREEK CAPITAL LETTER EPSILON WITH DASIA AND VARIA
4878	<E*,'>	<U1F1C>	GREEK CAPITAL LETTER EPSILON WITH PSILI AND OXIA
4879	<E*,'>	<U1F1D>	GREEK CAPITAL LETTER EPSILON WITH DASIA AND OXIA
4880	<y*,>	<U1F20>	GREEK SMALL LETTER ETA WITH PSILI
4881	<y*,>	<U1F21>	GREEK SMALL LETTER ETA WITH DASIA
4882	<y*;!>	<U1F22>	GREEK SMALL LETTER ETA WITH PSILI AND VARIA
4883	<y*;!>	<U1F23>	GREEK SMALL LETTER ETA WITH DASIA AND VARIA
4884	<y*,'>	<U1F24>	GREEK SMALL LETTER ETA WITH PSILI AND OXIA
4885	<y*,'>	<U1F25>	GREEK SMALL LETTER ETA WITH DASIA AND OXIA
4886	<y*,>	<U1F26>	GREEK SMALL LETTER ETA WITH PSILI AND PERISPOMENI
4887	<y*,>	<U1F27>	GREEK SMALL LETTER ETA WITH DASIA AND PERISPOMENI
4888	<Y*,>	<U1F28>	GREEK CAPITAL LETTER ETA WITH PSILI
4889	<Y*,>	<U1F29>	GREEK CAPITAL LETTER ETA WITH DASIA
4890	<Y*;!>	<U1F2A>	GREEK CAPITAL LETTER ETA WITH PSILI AND VARIA
4891	<Y*;!>	<U1F2B>	GREEK CAPITAL LETTER ETA WITH DASIA AND VARIA
4892	<Y*,'>	<U1F2C>	GREEK CAPITAL LETTER ETA WITH PSILI AND OXIA
4893	<Y*,'>	<U1F2D>	GREEK CAPITAL LETTER ETA WITH DASIA AND OXIA
4894	<Y*,>	<U1F2E>	GREEK CAPITAL LETTER ETA WITH PSILI AND PERISPOMENI
4895	<Y*,>	<U1F2F>	GREEK CAPITAL LETTER ETA WITH DASIA AND PERISPOMENI
4896	<i*,>	<U1F30>	GREEK SMALL LETTER IOTA WITH PSILI
4897	<i*,>	<U1F31>	GREEK SMALL LETTER IOTA WITH DASIA
4898	<i*;!>	<U1F32>	GREEK SMALL LETTER IOTA WITH PSILI AND VARIA
4899	<i*;!>	<U1F33>	GREEK SMALL LETTER IOTA WITH DASIA AND VARIA
4900	<i*,'>	<U1F34>	GREEK SMALL LETTER IOTA WITH PSILI AND OXIA
4901	<i*,'>	<U1F35>	GREEK SMALL LETTER IOTA WITH DASIA AND OXIA
4902	<i*,>	<U1F36>	GREEK SMALL LETTER IOTA WITH PSILI AND PERISPOMENI
4903	<i*,>	<U1F37>	GREEK SMALL LETTER IOTA WITH DASIA AND PERISPOMENI
4904	<I*,>	<U1F38>	GREEK CAPITAL LETTER IOTA WITH PSILI
4905	<I*,>	<U1F39>	GREEK CAPITAL LETTER IOTA WITH DASIA
4906	<I*;!>	<U1F3A>	GREEK CAPITAL LETTER IOTA WITH PSILI AND VARIA
4907	<I*;!>	<U1F3B>	GREEK CAPITAL LETTER IOTA WITH DASIA AND VARIA
4908	<I*,'>	<U1F3C>	GREEK CAPITAL LETTER IOTA WITH PSILI AND OXIA
4909	<I*,'>	<U1F3D>	GREEK CAPITAL LETTER IOTA WITH DASIA AND OXIA
4910	<I*,>	<U1F3E>	GREEK CAPITAL LETTER IOTA WITH PSILI AND PERISPOMENI
4911	<I*,>	<U1F3F>	GREEK CAPITAL LETTER IOTA WITH DASIA AND PERISPOMENI
4912	<o*,>	<U1F40>	GREEK SMALL LETTER OMICRON WITH PSILI
4913	<o*,>	<U1F41>	GREEK SMALL LETTER OMICRON WITH DASIA
4914	<o*;!>	<U1F42>	GREEK SMALL LETTER OMICRON WITH PSILI AND VARIA
4915	<o*;!>	<U1F43>	GREEK SMALL LETTER OMICRON WITH DASIA AND VARIA
4916	<o*,'>	<U1F44>	GREEK SMALL LETTER OMICRON WITH PSILI AND OXIA
4917	<o*,'>	<U1F45>	GREEK SMALL LETTER OMICRON WITH DASIA AND OXIA
4918	<O*,>	<U1F48>	GREEK CAPITAL LETTER OMICRON WITH PSILI
4919	<O*,>	<U1F49>	GREEK CAPITAL LETTER OMICRON WITH DASIA
4920	<O*;!>	<U1F4A>	GREEK CAPITAL LETTER OMICRON WITH PSILI AND VARIA
4921	<O*;!>	<U1F4B>	GREEK CAPITAL LETTER OMICRON WITH DASIA AND VARIA
4922	<O*,'>	<U1F4C>	GREEK CAPITAL LETTER OMICRON WITH PSILI AND OXIA
4923	<O*,'>	<U1F4D>	GREEK CAPITAL LETTER OMICRON WITH DASIA AND OXIA
4924	<u*,>	<U1F50>	GREEK SMALL LETTER UPSILON WITH PSILI
4925	<u*,>	<U1F51>	GREEK SMALL LETTER UPSILON WITH DASIA
4926	<u*;!>	<U1F52>	GREEK SMALL LETTER UPSILON WITH PSILI AND VARIA
4927	<u*;!>	<U1F53>	GREEK SMALL LETTER UPSILON WITH DASIA AND VARIA
4928	<u*,'>	<U1F54>	GREEK SMALL LETTER UPSILON WITH PSILI AND OXIA
4929	<u*,'>	<U1F55>	GREEK SMALL LETTER UPSILON WITH DASIA AND OXIA
4930	<u*,>	<U1F56>	GREEK SMALL LETTER UPSILON WITH PSILI AND PERISPOMENI
4931	<u*,>	<U1F57>	GREEK SMALL LETTER UPSILON WITH DASIA AND PERISPOMENI
4932	<U*,>	<U1F59>	GREEK CAPITAL LETTER UPSILON WITH DASIA

4933	<U*;!>	<U1F5B>	GREEK CAPITAL LETTER UPSILON WITH DASIA AND VARIA
4934	<U*;'>	<U1F5D>	GREEK CAPITAL LETTER UPSILON WITH DASIA AND OXIA
4935	<U*?>	<U1F5F>	GREEK CAPITAL LETTER UPSILON WITH DASIA AND PERISPOMENI
4936	<w*,>	<U1F60>	GREEK SMALL LETTER OMEGA WITH PSILI
4937	<w*,>	<U1F61>	GREEK SMALL LETTER OMEGA WITH DASIA
4938	<w*;!>	<U1F62>	GREEK SMALL LETTER OMEGA WITH PSILI AND VARIA
4939	<w*;'>	<U1F63>	GREEK SMALL LETTER OMEGA WITH DASIA AND VARIA
4940	<w*,'>	<U1F64>	GREEK SMALL LETTER OMEGA WITH PSILI AND OXIA
4941	<w*;'>	<U1F65>	GREEK SMALL LETTER OMEGA WITH DASIA AND OXIA
4942	<w*,?>	<U1F66>	GREEK SMALL LETTER OMEGA WITH PSILI AND PERISPOMENI
4943	<w*?>	<U1F67>	GREEK SMALL LETTER OMEGA WITH DASIA AND PERISPOMENI
4944	<W*,>	<U1F68>	GREEK CAPITAL LETTER OMEGA WITH PSILI
4945	<W*,>	<U1F69>	GREEK CAPITAL LETTER OMEGA WITH DASIA
4946	<W*;!>	<U1F6A>	GREEK CAPITAL LETTER OMEGA WITH PSILI AND VARIA
4947	<W*;!>	<U1F6B>	GREEK CAPITAL LETTER OMEGA WITH DASIA AND VARIA
4948	<W*,'>	<U1F6C>	GREEK CAPITAL LETTER OMEGA WITH PSILI AND OXIA
4949	<W*,'>	<U1F6D>	GREEK CAPITAL LETTER OMEGA WITH DASIA AND OXIA
4950	<W*,?>	<U1F6E>	GREEK CAPITAL LETTER OMEGA WITH PSILI AND PERISPOMENI
4951	<W*?>	<U1F6F>	GREEK CAPITAL LETTER OMEGA WITH DASIA AND PERISPOMENI
4952	<a*!>	<U1F70>	GREEK SMALL LETTER ALPHA WITH VARIA
4953	<a*!'>	<U1F71>	GREEK SMALL LETTER ALPHA WITH OXIA
4954	<e*!>	<U1F72>	GREEK SMALL LETTER EPSILON WITH VARIA
4955	<e*!'>	<U1F73>	GREEK SMALL LETTER EPSILON WITH OXIA
4956	<y*!>	<U1F74>	GREEK SMALL LETTER ETA WITH VARIA
4957	<y*!'>	<U1F75>	GREEK SMALL LETTER ETA WITH OXIA
4958	<i*!>	<U1F76>	GREEK SMALL LETTER IOTA WITH VARIA
4959	<i*!'>	<U1F77>	GREEK SMALL LETTER IOTA WITH OXIA
4960	<o*!>	<U1F78>	GREEK SMALL LETTER OMICRON WITH VARIA
4961	<o*!'>	<U1F79>	GREEK SMALL LETTER OMICRON WITH OXIA
4962	<u*!>	<U1F7A>	GREEK SMALL LETTER UPSILON WITH VARIA
4963	<u*!'>	<U1F7B>	GREEK SMALL LETTER UPSILON WITH OXIA
4964	<w*!>	<U1F7C>	GREEK SMALL LETTER OMEGA WITH VARIA
4965	<w*!'>	<U1F7D>	GREEK SMALL LETTER OMEGA WITH OXIA
4966	<a*,j>	<U1F80>	GREEK SMALL LETTER ALPHA WITH PSILI AND YPOGEGRAMMENI
4967	<a*,j>	<U1F81>	GREEK SMALL LETTER ALPHA WITH DASIA AND YPOGEGRAMMENI
4968	<a*,!j>	<U1F82>	GREEK SMALL LETTER ALPHA WITH PSILI AND VARIA AND YPOGEGRAMMENI
4969	<a*,!j>	<U1F83>	GREEK SMALL LETTER ALPHA WITH DASIA AND VARIA AND YPOGEGRAMMENI
4970	<a*,'j>	<U1F84>	GREEK SMALL LETTER ALPHA WITH PSILI AND OXIA AND YPOGEGRAMMENI
4971	<a*,'j>	<U1F85>	GREEK SMALL LETTER ALPHA WITH DASIA AND OXIA AND YPOGEGRAMMENI
4972	<a*,?j>	<U1F86>	GREEK SMALL LETTER ALPHA WITH PSILI AND PERISPOMENI AND YPOGEGRAMMENI
4973	<a*,?j>	<U1F87>	GREEK SMALL LETTER ALPHA WITH DASIA AND PERISPOMENI AND YPOGEGRAMMENI
4974	<A*,J>	<U1F88>	GREEK CAPITAL LETTER ALPHA WITH PSILI AND PROSGEGRAMMENI
4975	<A*,J>	<U1F89>	GREEK CAPITAL LETTER ALPHA WITH DASIA AND PROSGEGRAMMENI
4976	<A*,!J>	<U1F8A>	GREEK CAPITAL LETTER ALPHA WITH PSILI AND VARIA AND PROSGEGRAMMENI
4977	<A*,!J>	<U1F8B>	GREEK CAPITAL LETTER ALPHA WITH DASIA AND VARIA AND PROSGEGRAMMENI
4978	<A*,'J>	<U1F8C>	GREEK CAPITAL LETTER ALPHA WITH PSILI AND OXIA AND PROSGEGRAMMENI
4979	<A*,'J>	<U1F8D>	GREEK CAPITAL LETTER ALPHA WITH DASIA AND OXIA AND PROSGEGRAMMENI
4980	<A*,?J>	<U1F8E>	GREEK CAPITAL LETTER ALPHA WITH PSILI AND PERISPOMENI AND PROSGEGRAMMENI
4981	<A*,?J>	<U1F8F>	GREEK CAPITAL LETTER ALPHA WITH DASIA AND PERISPOMENI AND PROSGEGRAMMENI
4982	<y*,j>	<U1F90>	GREEK SMALL LETTER ETA WITH PSILI AND YPOGEGRAMMENI
4983	<y*,j>	<U1F91>	GREEK SMALL LETTER ETA WITH DASIA AND YPOGEGRAMMENI
4984	<y*,!j>	<U1F92>	GREEK SMALL LETTER ETA WITH PSILI AND VARIA AND YPOGEGRAMMENI
4985	<y*,'j>	<U1F93>	GREEK SMALL LETTER ETA WITH DASIA AND VARIA AND YPOGEGRAMMENI
4986	<y*,'j>	<U1F94>	GREEK SMALL LETTER ETA WITH PSILI AND OXIA AND YPOGEGRAMMENI
4987	<y*,'j>	<U1F95>	GREEK SMALL LETTER ETA WITH DASIA AND OXIA AND YPOGEGRAMMENI
4988	<y*,?j>	<U1F96>	GREEK SMALL LETTER ETA WITH PSILI AND PERISPOMENI AND YPOGEGRAMMENI
4989	<y*,?j>	<U1F97>	GREEK SMALL LETTER ETA WITH DASIA AND PERISPOMENI AND YPOGEGRAMMENI
4990	<Y*,J>	<U1F98>	GREEK CAPITAL LETTER ETA WITH PSILI AND PROSGEGRAMMENI
4991	<Y*,J>	<U1F99>	GREEK CAPITAL LETTER ETA WITH DASIA AND PROSGEGRAMMENI
4992	<Y*,!J>	<U1F9A>	GREEK CAPITAL LETTER ETA WITH PSILI AND VARIA AND PROSGEGRAMMENI
4993	<Y*,!J>	<U1F9B>	GREEK CAPITAL LETTER ETA WITH DASIA AND VARIA AND PROSGEGRAMMENI
4994	<Y*,'J>	<U1F9C>	GREEK CAPITAL LETTER ETA WITH PSILI AND OXIA AND PROSGEGRAMMENI
4995	<Y*,'J>	<U1F9D>	GREEK CAPITAL LETTER ETA WITH DASIA AND OXIA AND PROSGEGRAMMENI
4996	<Y*,?J>	<U1F9E>	GREEK CAPITAL LETTER ETA WITH PSILI AND PERISPOMENI AND PROSGEGRAMMENI
4997	<Y*,?J>	<U1F9F>	GREEK CAPITAL LETTER ETA WITH DASIA AND PERISPOMENI AND PROSGEGRAMMENI
4998	<w*,j>	<U1FA0>	GREEK SMALL LETTER OMEGA WITH PSILI AND YPOGEGRAMMENI
4999	<w*,j>	<U1FA1>	GREEK SMALL LETTER OMEGA WITH DASIA AND YPOGEGRAMMENI
5000	<w*,!j>	<U1FA2>	GREEK SMALL LETTER OMEGA WITH PSILI AND VARIA AND YPOGEGRAMMENI
5001	<w*,!j>	<U1FA3>	GREEK SMALL LETTER OMEGA WITH DASIA AND VARIA AND YPOGEGRAMMENI
5002	<w*,'j>	<U1FA4>	GREEK SMALL LETTER OMEGA WITH PSILI AND OXIA AND YPOGEGRAMMENI
5003	<w*,'j>	<U1FA5>	GREEK SMALL LETTER OMEGA WITH DASIA AND OXIA AND YPOGEGRAMMENI
5004	<w*,?j>	<U1FA6>	GREEK SMALL LETTER OMEGA WITH PSILI AND PERISPOMENI AND YPOGEGRAMMENI
5005	<w*,?j>	<U1FA7>	GREEK SMALL LETTER OMEGA WITH DASIA AND PERISPOMENI AND YPOGEGRAMMENI
5006	<W*,J>	<U1FA8>	GREEK CAPITAL LETTER OMEGA WITH PSILI AND PROSGEGRAMMENI
5007	<W*,J>	<U1FA9>	GREEK CAPITAL LETTER OMEGA WITH DASIA AND PROSGEGRAMMENI
5008	<W*,!J>	<U1FAA>	GREEK CAPITAL LETTER OMEGA WITH PSILI AND VARIA AND PROSGEGRAMMENI
5009	<W*,!J>	<U1FAB>	GREEK CAPITAL LETTER OMEGA WITH DASIA AND VARIA AND PROSGEGRAMMENI
5010	<W*,'J>	<U1FAC>	GREEK CAPITAL LETTER OMEGA WITH PSILI AND OXIA AND PROSGEGRAMMENI
5011	<W*,'J>	<U1FAD>	GREEK CAPITAL LETTER OMEGA WITH DASIA AND OXIA AND PROSGEGRAMMENI
5012	<W*,?J>	<U1FAE>	GREEK CAPITAL LETTER OMEGA WITH PSILI AND PERISPOMENI AND PROSGEGRAMMENI
5013	<W*,?J>	<U1FAF>	GREEK CAPITAL LETTER OMEGA WITH DASIA AND PERISPOMENI AND PROSGEGRAMMENI
5014	<a*(>	<U1FB0>	GREEK SMALL LETTER ALPHA WITH VRACHY
5015	<a*->	<U1FB1>	GREEK SMALL LETTER ALPHA WITH MACRON
5016	<a*!j>	<U1FB2>	GREEK SMALL LETTER ALPHA WITH VARIA AND YPOGEGRAMMENI
5017	<a*j>	<U1FB3>	GREEK SMALL LETTER ALPHA WITH YPOGEGRAMMENI
5018	<a*'j>	<U1FB4>	GREEK SMALL LETTER ALPHA WITH OXIA AND YPOGEGRAMMENI
5019	<a*?>	<U1FB6>	GREEK SMALL LETTER ALPHA WITH PERISPOMENI
5020	<a*?j>	<U1FB7>	GREEK SMALL LETTER ALPHA WITH PERISPOMENI AND YPOGEGRAMMENI
5021	<A*(>	<U1FB8>	GREEK CAPITAL LETTER ALPHA WITH VRACHY

5022	<A*->	<U1FB9>	GREEK CAPITAL LETTER ALPHA WITH MACRON
5023	<A*!>	<U1FBA>	GREEK CAPITAL LETTER ALPHA WITH VARIA
5024	<A*’>	<U1FBB>	GREEK CAPITAL LETTER ALPHA WITH OXIA
5025	<A*J>	<U1FBC>	GREEK CAPITAL LETTER ALPHA WITH PROSGEGRAMMENI
5026	<)*>	<U1FBD>	GREEK KORONIS
5027	<J3>	<U1FBE>	GREEK PROSGEGRAMMENI
5028	<, ,>	<U1FBF>	GREEK PSILI
5029	<?*>	<U1FC0>	GREEK PERISPOMENI
5030	<?:>	<U1FC1>	GREEK DIALYTIKA AND PERISPOMENI
5031	<y*!j>	<U1FC2>	GREEK SMALL LETTER ETA WITH VARIA AND YPOGEGRAMMENI
5032	<y*j>	<U1FC3>	GREEK SMALL LETTER ETA WITH YPOGEGRAMMENI
5033	<y*’j>	<U1FC4>	GREEK SMALL LETTER ETA WITH OXIA AND YPOGEGRAMMENI
5034	<y*?j>	<U1FC6>	GREEK SMALL LETTER ETA WITH PERISPOMENI
5035	<y*?j>	<U1FC7>	GREEK SMALL LETTER ETA WITH PERISPOMENI AND YPOGEGRAMMENI
5036	<E*!>	<U1FC8>	GREEK CAPITAL LETTER EPSILON WITH VARIA
5037	<E*’>	<U1FC9>	GREEK CAPITAL LETTER EPSILON WITH OXIA
5038	<Y*!>	<U1FCA>	GREEK CAPITAL LETTER ETA WITH VARIA
5039	<Y*’>	<U1FCB>	GREEK CAPITAL LETTER ETA WITH OXIA
5040	<Y*J>	<U1FCC>	GREEK CAPITAL LETTER ETA WITH PROSGEGRAMMENI
5041	<, !>	<U1FCD>	GREEK PSILI AND VARIA
5042	<, ’>	<U1FCE>	GREEK PSILI AND OXIA
5043	<?>	<U1FCF>	GREEK PSILI AND PERISPOMENI
5044	<i*(>	<U1FD0>	GREEK SMALL LETTER IOTA WITH VRACHY
5045	<i*->	<U1FD1>	GREEK SMALL LETTER IOTA WITH MACRON
5046	<i*!>	<U1FD2>	GREEK SMALL LETTER IOTA WITH DIALYTIKA AND VARIA
5047	<i*’>	<U1FD3>	GREEK SMALL LETTER IOTA WITH DIALYTIKA AND OXIA
5048	<i*?>	<U1FD6>	GREEK SMALL LETTER IOTA WITH PERISPOMENI
5049	<i*?j>	<U1FD7>	GREEK SMALL LETTER IOTA WITH DIALYTIKA AND PERISPOMENI
5050	<I*(>	<U1FD8>	GREEK CAPITAL LETTER IOTA WITH VRACHY
5051	<I*->	<U1FD9>	GREEK CAPITAL LETTER IOTA WITH MACRON
5052	<I*!>	<U1FDA>	GREEK CAPITAL LETTER IOTA WITH VARIA
5053	<I*’>	<U1FDB>	GREEK CAPITAL LETTER IOTA WITH OXIA
5054	<!>	<U1FDD>	GREEK DASIA AND VARIA
5055	<!’>	<U1FDE>	GREEK DASIA AND OXIA
5056	<?>	<U1FDF>	GREEK DASIA AND PERISPOMENI
5057	<u*(>	<U1FE0>	GREEK SMALL LETTER UPSILON WITH VRACHY
5058	<u*->	<U1FE1>	GREEK SMALL LETTER UPSILON WITH MACRON
5059	<u*!>	<U1FE2>	GREEK SMALL LETTER UPSILON WITH DIALYTIKA AND VARIA
5060	<u*’>	<U1FE3>	GREEK SMALL LETTER UPSILON WITH DIALYTIKA AND OXIA
5061	<r* >	<U1FE4>	GREEK SMALL LETTER RHO WITH PSILI
5062	<r*’>	<U1FE5>	GREEK SMALL LETTER RHO WITH DASIA
5063	<u*?>	<U1FE6>	GREEK SMALL LETTER UPSILON WITH PERISPOMENI
5064	<u*?j>	<U1FE7>	GREEK SMALL LETTER UPSILON WITH DIALYTIKA AND PERISPOMENI
5065	<U*(>	<U1FE8>	GREEK CAPITAL LETTER UPSILON WITH VRACHY
5066	<U*->	<U1FE9>	GREEK CAPITAL LETTER UPSILON WITH MACRON
5067	<U*!>	<U1FEA>	GREEK CAPITAL LETTER UPSILON WITH VARIA
5068	<U*’>	<U1FEB>	GREEK CAPITAL LETTER UPSILON WITH OXIA
5069	<R*’>	<U1FEC>	GREEK CAPITAL LETTER RHO WITH DASIA
5070	<!:>	<U1FED>	GREEK DIALYTIKA AND VARIA
5071	<:’>	<U1FEE>	GREEK DIALYTIKA AND OXIA
5072	<!*>	<U1FEF>	GREEK VARIA
5073	<w*!j>	<U1FF2>	GREEK SMALL LETTER OMEGA WITH VARIA AND YPOGEGRAMMENI
5074	<w*j>	<U1FF3>	GREEK SMALL LETTER OMEGA WITH YPOGEGRAMMENI
5075	<w*’j>	<U1FF4>	GREEK SMALL LETTER OMEGA WITH OXIA AND YPOGEGRAMMENI
5076	<w*?>	<U1FF6>	GREEK SMALL LETTER OMEGA WITH PERISPOMENI
5077	<w*?j>	<U1FF7>	GREEK SMALL LETTER OMEGA WITH PERISPOMENI AND YPOGEGRAMMENI
5078	<O*!>	<U1FF8>	GREEK CAPITAL LETTER OMICRON WITH VARIA
5079	<O*’>	<U1FF9>	GREEK CAPITAL LETTER OMICRON WITH OXIA
5080	<W*!>	<U1FFA>	GREEK CAPITAL LETTER OMEGA WITH VARIA
5081	<W*’>	<U1FFB>	GREEK CAPITAL LETTER OMEGA WITH OXIA
5082	<W*J>	<U1FFC>	GREEK CAPITAL LETTER OMEGA WITH PROSGEGRAMMENI
5083	<//*>	<U1FFD>	GREEK OXIA
5084	<!’>	<U1FFE>	GREEK DASIA
5085	<1N>	<U2002>	EN SPACE
5086	<1M>	<U2003>	EM SPACE
5087	<3M>	<U2004>	THREE-PER-EM SPACE
5088	<4M>	<U2005>	FOUR-PER-EM SPACE
5089	<6M>	<U2006>	SIX-PER-EM SPACE
5090	<LR>	<U200E>	LEFT-TO-RIGHT MARK
5091	<RL>	<U200F>	RIGHT-TO-LEFT MARK
5092	<1T>	<U2009>	THIN SPACE
5093	<1H>	<U200A>	HAIR SPACE
5094	<-1>	<U2010>	HYPHEN
5095	<-N>	<U2013>	EN DASH
5096	<-M>	<U2014>	EM DASH
5097	<-3>	<U2015>	HORIZONTAL BAR
5098	<!2>	<U2016>	DOUBLE VERTICAL LINE
5099	<=2>	<U2017>	DOUBLE LOW LINE
5100	<’6>	<U2018>	LEFT SINGLE QUOTATION MARK
5101	<’9>	<U2019>	RIGHT SINGLE QUOTATION MARK
5102	<.9>	<U201A>	SINGLE LOW-9 QUOTATION MARK
5103	<’9>	<U201B>	SINGLE HIGH-REVERSED-9 QUOTATION MARK
5104	<”6>	<U201C>	LEFT DOUBLE QUOTATION MARK
5105	<”9>	<U201D>	RIGHT DOUBLE QUOTATION MARK
5106	<.:9>	<U201E>	DOUBLE LOW-9 QUOTATION MARK
5107	<”9>	<U201F>	DOUBLE HIGH-REVERSED-9 QUOTATION MARK
5108	<//->	<U2020>	DAGGER
5109	<//=>	<U2021>	DOUBLE DAGGER
5110	<sb>	<U2022>	BULLET

111	<3b>	<U2023>	TRIANGULAR BULLET
112	<.>	<U2025>	TWO DOT LEADER
113	<.3>	<U2026>	HORIZONTAL ELLIPSIS
114	<.->	<U2027>	HYPHENATION POINT
115	<linesep>	<U2028>	LINE SEPARATOR
116	<parsep>	<U2029>	PARAGRAPH SEPARATOR
117	<%0>	<U2030>	PER MILLE SIGN
118	<1'>	<U2032>	PRIME
119	<2'>	<U2033>	DOUBLE PRIME
120	<3'>	<U2034>	TRIPLE PRIME
121	<1">	<U2035>	REVERSED PRIME
122	<2">	<U2036>	REVERSED DOUBLE PRIME
123	<3">	<U2037>	REVERSED TRIPLE PRIME
124	<Ca>	<U2038>	CARET
125	<<1>	<U2039>	SINGLE LEFT-POINTING ANGLE QUOTATION MARK
126	</>1>	<U203A>	SINGLE RIGHT-POINTING ANGLE QUOTATION MARK
127	<:X>	<U203B>	REFERENCE MARK
128	<!*2>	<U203C>	DOUBLE EXCLAMATION MARK
129	<' ->	<U203E>	OVERLINE
130	<-b>	<U2043>	HYPHEN BULLET
131	</f>	<U2044>	FRACTION SLASH
132	<0s>	<U2070>	SUPERSCRIPIT ZERO
133	<4s>	<U2074>	SUPERSCRIPIT FOUR
134	<5s>	<U2075>	SUPERSCRIPIT FIVE
135	<6s>	<U2076>	SUPERSCRIPIT SIX
136	<7s>	<U2077>	SUPERSCRIPIT SEVEN
137	<8s>	<U2078>	SUPERSCRIPIT EIGHT
138	<9s>	<U2079>	SUPERSCRIPIT NINE
139	<+s>	<U207A>	SUPERSCRIPIT PLUS SIGN
140	<-s>	<U207B>	SUPERSCRIPIT MINUS
141	<=s>	<U207C>	SUPERSCRIPIT EQUALS SIGN
142	<(s>	<U207D>	SUPERSCRIPIT LEFT PARENTHESIS
143	<)s>	<U207E>	SUPERSCRIPIT RIGHT PARENTHESIS
144	<nS>	<U207F>	SUPERSCRIPIT LATIN SMALL LETTER N
145	<0s>	<U2080>	SUBSCRIPT ZERO
146	<1s>	<U2081>	SUBSCRIPT ONE
147	<2s>	<U2082>	SUBSCRIPT TWO
148	<3s>	<U2083>	SUBSCRIPT THREE
149	<4s>	<U2084>	SUBSCRIPT FOUR
150	<5s>	<U2085>	SUBSCRIPT FIVE
151	<6s>	<U2086>	SUBSCRIPT SIX
152	<7s>	<U2087>	SUBSCRIPT SEVEN
153	<8s>	<U2088>	SUBSCRIPT EIGHT
154	<9s>	<U2089>	SUBSCRIPT NINE
155	<+s>	<U208A>	SUBSCRIPT PLUS SIGN
156	<-s>	<U208B>	SUBSCRIPT MINUS
157	<=s>	<U208C>	SUBSCRIPT EQUALS SIGN
158	<(s>	<U208D>	SUBSCRIPT LEFT PARENTHESIS
159	<)s>	<U208E>	SUBSCRIPT RIGHT PARENTHESIS
160	<Ff>	<U20A3>	FRENCH FRANC SIGN
161		<U20A4>	LIRA SIGN
162	<Pt>	<U20A7>	PESETA SIGN
163	<W= >	<U20A9>	WON SIGN
164	<"7>	<U20D1>	COMBINING RIGHT HARPOON ABOVE
165	<oC>	<U2103>	DEGREE CELSIUS
166	<co>	<U2105>	CARE OF
167	<oF>	<U2109>	DEGREE FAHRENHEIT
168	<NO>	<U2116>	NUMERO SIGN
169	<PO>	<U2117>	SOUND RECORDING COPYRIGHT
170	<Rx>	<U211E>	PRESCRIPTION TAKE
171	<SM>	<U2120>	SERVICE MARK
172	<TM>	<U2122>	TRADE MARK SIGN
173	<Om>	<U2126>	OHM SIGN
174	<AO>	<U212B>	ANGSTROM SIGN
175	<Est>	<U212E>	ESTIMATED SYMBOL
176	<13>	<U2153>	VULGAR FRACTION ONE THIRD
177	<23>	<U2154>	VULGAR FRACTION TWO THIRDS
178	<15>	<U2155>	VULGAR FRACTION ONE FIFTH
179	<25>	<U2156>	VULGAR FRACTION TWO FIFTHS
180	<35>	<U2157>	VULGAR FRACTION THREE FIFTHS
181	<45>	<U2158>	VULGAR FRACTION FOUR FIFTHS
182	<16>	<U2159>	VULGAR FRACTION ONE SIXTH
183	<56>	<U215A>	VULGAR FRACTION FIVE SIXTHS
184	<18>	<U215B>	VULGAR FRACTION ONE EIGHTH
185	<38>	<U215C>	VULGAR FRACTION THREE EIGHTHS
186	<58>	<U215D>	VULGAR FRACTION FIVE EIGHTHS
187	<78>	<U215E>	VULGAR FRACTION SEVEN EIGHTHS
188	<1R>	<U2160>	ROMAN NUMERAL ONE
189	<2R>	<U2161>	ROMAN NUMERAL TWO
190	<3R>	<U2162>	ROMAN NUMERAL THREE
191	<4R>	<U2163>	ROMAN NUMERAL FOUR
192	<5R>	<U2164>	ROMAN NUMERAL FIVE
193	<6R>	<U2165>	ROMAN NUMERAL SIX
194	<7R>	<U2166>	ROMAN NUMERAL SEVEN
195	<8R>	<U2167>	ROMAN NUMERAL EIGHT
196	<9R>	<U2168>	ROMAN NUMERAL NINE
197	<aR>	<U2169>	ROMAN NUMERAL TEN
198	 	<U216A>	ROMAN NUMERAL ELEVEN
199	<cR>	<U216B>	ROMAN NUMERAL TWELVE

5200	<50R>	<U216C>	ROMAN NUMERAL FIFTY
5201	<100R>	<U216D>	ROMAN NUMERAL ONE HUNDRED
5202	<500R>	<U216E>	ROMAN NUMERAL FIVE HUNDRED
5203	<1000R>	<U216F>	ROMAN NUMERAL ONE THOUSAND
5204	<1r>	<U2170>	SMALL ROMAN NUMERAL ONE
5205	<2r>	<U2171>	SMALL ROMAN NUMERAL TWO
5206	<3r>	<U2172>	SMALL ROMAN NUMERAL THREE
5207	<4r>	<U2173>	SMALL ROMAN NUMERAL FOUR
5208	<5r>	<U2174>	SMALL ROMAN NUMERAL FIVE
5209	<6r>	<U2175>	SMALL ROMAN NUMERAL SIX
5210	<7r>	<U2176>	SMALL ROMAN NUMERAL SEVEN
5211	<8r>	<U2177>	SMALL ROMAN NUMERAL EIGHT
5212	<9r>	<U2178>	SMALL ROMAN NUMERAL NINE
5213	<ar>	<U2179>	SMALL ROMAN NUMERAL TEN
5214	 	<U217A>	SMALL ROMAN NUMERAL ELEVEN
5215	<cr>	<U217B>	SMALL ROMAN NUMERAL TWELVE
5216	<50r>	<U217C>	SMALL ROMAN NUMERAL FIFTY
5217	<100r>	<U217D>	SMALL ROMAN NUMERAL ONE HUNDRED
5218	<500r>	<U217E>	SMALL ROMAN NUMERAL FIVE HUNDRED
5219	<1000r>	<U217F>	SMALL ROMAN NUMERAL ONE THOUSAND
5220	<1000RCD>	<U2180>	ROMAN NUMERAL ONE THOUSAND C D
5221	<5000R>	<U2181>	ROMAN NUMERAL FIVE THOUSAND
5222	<10000R>	<U2182>	ROMAN NUMERAL TEN THOUSAND
5223	<->	<U2190>	LEFTWARDS ARROW
5224	<-!>	<U2191>	UPWARDS ARROW
5225	<-/>	<U2192>	RIGHTWARDS ARROW
5226	<-v>	<U2193>	DOWNWARDS ARROW
5227	<</>	<U2194>	LEFT RIGHT ARROW
5228	<UD>	<U2195>	UP DOWN ARROW
5229	<<!!>	<U2196>	NORTH WEST ARROW
5230	</////>	<U2197>	NORTH EAST ARROW
5231	<!!!/>	<U2198>	SOUTH EAST ARROW
5232	<</////>	<U2199>	SOUTH WEST ARROW
5233	<UD->	<U21A8>	UP DOWN ARROW WITH BASE
5234	</>v>	<U21C0>	RIGHTWARDS HARPOON WITH BARB UPWARDS
5235	<<=>	<U21D0>	LEFTWARDS DOUBLE ARROW
5236	<=>/>	<U21D2>	RIGHTWARDS DOUBLE ARROW
5237	<==>	<U21D4>	LEFT RIGHT DOUBLE ARROW
5238	<FA>	<U2200>	FOR ALL
5239	<dP>	<U2202>	PARTIAL DIFFERENTIAL
5240	<TE>	<U2203>	THERE EXISTS
5241	<//0>	<U2205>	EMPTY SET
5242	<DE>	<U2206>	INCREMENT
5243	<NB>	<U2207>	NABLA
5244	<(->	<U2208>	ELEMENT OF
5245	<(-)>	<U220B>	CONTAINS AS MEMBER
5246	<FP>	<U220E>	END OF PROOF
5247	<*P>	<U220F>	N-ARY PRODUCT
5248	<+Z>	<U2211>	N-ARY SUMMATION
5249	<-2>	<U2212>	MINUS SIGN
5250	<-+>	<U2213>	MINUS-OR-PLUS SIGN
5251	<.+>	<U2214>	DOT PLUS
5252	<*->	<U2217>	ASTERISK OPERATOR
5253	<Ob>	<U2218>	RING OPERATOR
5254	<Sb>	<U2219>	BULLET OPERATOR
5255	<RT>	<U221A>	SQUARE ROOT
5256	<0(>	<U221D>	PROPORTIONAL TO
5257	<00>	<U221E>	INFINITY
5258	<-L>	<U221F>	RIGHT ANGLE
5259	<-V>	<U2220>	ANGLE
5260	<PP>	<U2225>	PARALLEL TO
5261	<AN>	<U2227>	LOGICAL AND
5262	<OR>	<U2228>	LOGICAL OR
5263	<(U>	<U2229>	INTERSECTION
5264	<)U>	<U222A>	UNION
5265	<In>	<U222B>	INTEGRAL
5266	<DI>	<U222C>	DOUBLE INTEGRAL
5267	<Io>	<U222E>	CONTOUR INTEGRAL
5268	<.:>	<U2234>	THEREFORE
5269	<:.>	<U2235>	BECAUSE
5270	<:R>	<U2236>	RATIO
5271	<::>	<U2237>	PROPORTION
5272	<?1>	<U223C>	TILDE OPERATOR
5273	<CG>	<U223E>	INVERTED LAZY S
5274	<?->	<U2243>	ASYMPTOTICALLY EQUAL TO
5275	<?=>	<U2245>	APPROXIMATELY EQUAL TO
5276	<?2>	<U2248>	ALMOST EQUAL TO
5277	<=?>	<U224C>	ALL EQUAL TO
5278	<HI>	<U2253>	IMAGE OF OR APPROXIMATELY EQUAL TO
5279	<!=>	<U2260>	NOT EQUAL TO
5280	<=3>	<U2261>	IDENTICAL TO
5281	<=<>	<U2264>	LESS-THAN OR EQUAL TO
5282	</>=>	<U2265>	GREATER-THAN OR EQUAL TO
5283	<<*>	<U226A>	MUCH LESS-THAN
5284	<*/>>	<U226B>	MUCH GREATER-THAN
5285	<!<>	<U226E>	NOT LESS-THAN
5286	<!/>>	<U226F>	NOT GREATER-THAN
5287	<(C>	<U2282>	SUBSET OF
5288	<)C>	<U2283>	SUPERSET OF

5289	<(_>	<U2286>	SUBSET OF OR EQUAL TO
5290	<_)>	<U2287>	SUPERSET OF OR EQUAL TO
5291	<0.>	<U2299>	CIRCLED DOT OPERATOR
5292	<02>	<U229A>	CIRCLED RING OPERATOR
5293	<-T>	<U22A5>	UP TACK
5294	<..P>	<U22C5>	DOT OPERATOR
5295	<:3>	<U22EE>	VERTICAL ELLIPSIS
5296	<Eh>	<U2302>	HOUSE
5297	<<7>	<U2308>	LEFT CEILING
5298	</>7>	<U2309>	RIGHT CEILING
5299	<7<>	<U230A>	LEFT FLOOR
5300	<7/>>	<U230B>	RIGHT FLOOR
5301	<NI>	<U2310>	REVERSED NOT SIGN
5302	<(A>	<U2312>	ARC
5303	<TR>	<U2315>	TELEPHONE RECORDER
5304	<88>	<U2318>	PLACE OF INTEREST SIGN
5305	<Iu>	<U2320>	TOP HALF INTEGRAL
5306	<Il>	<U2321>	BOTTOM HALF INTEGRAL
5307	<<//>	<U2329>	LEFT-POINTING ANGLE BRACKET
5308	<///>>	<U232A>	RIGHT-POINTING ANGLE BRACKET
5309	<Vs>	<U2423>	OPEN BOX
5310	<1h>	<U2440>	OCR HOOK
5311	<3h>	<U2441>	OCR CHAIR
5312	<2h>	<U2442>	OCR FORK
5313	<4h>	<U2443>	OCR INVERTED FORK
5314	<1j>	<U2446>	OCR BRANCH BANK IDENTIFICATION
5315	<2j>	<U2447>	OCR AMOUNT OF CHECK
5316	<3j>	<U2448>	OCR DASH
5317	<4j>	<U2449>	OCR CUSTOMER ACCOUNT NUMBER
5318	<1-o>	<U2460>	CIRCLED DIGIT ONE
5319	<2-o>	<U2461>	CIRCLED DIGIT TWO
5320	<3-o>	<U2462>	CIRCLED DIGIT THREE
5321	<4-o>	<U2463>	CIRCLED DIGIT FOUR
5322	<5-o>	<U2464>	CIRCLED DIGIT FIVE
5323	<6-o>	<U2465>	CIRCLED DIGIT SIX
5324	<7-o>	<U2466>	CIRCLED DIGIT SEVEN
5325	<8-o>	<U2467>	CIRCLED DIGIT EIGHT
5326	<9-o>	<U2468>	CIRCLED DIGIT NINE
5327	<10-o>	<U2469>	CIRCLED NUMBER TEN
5328	<11-o>	<U246A>	CIRCLED NUMBER ELEVEN
5329	<12-o>	<U246B>	CIRCLED NUMBER TWELVE
5330	<13-o>	<U246C>	CIRCLED NUMBER THIRTEEN
5331	<14-o>	<U246D>	CIRCLED NUMBER FOURTEEN
5332	<15-o>	<U246E>	CIRCLED NUMBER FIFTEEN
5333	<16-o>	<U246F>	CIRCLED NUMBER SIXTEEN
5334	<17-o>	<U2470>	CIRCLED NUMBER SEVENTEEN
5335	<18-o>	<U2471>	CIRCLED NUMBER EIGHTEEN
5336	<19-o>	<U2472>	CIRCLED NUMBER NINETEEN
5337	<20-o>	<U2473>	CIRCLED NUMBER TWENTY
5338	<(1)>	<U2474>	PARENTHESIZED DIGIT ONE
5339	<(2)>	<U2475>	PARENTHESIZED DIGIT TWO
5340	<(3)>	<U2476>	PARENTHESIZED DIGIT THREE
5341	<(4)>	<U2477>	PARENTHESIZED DIGIT FOUR
5342	<(5)>	<U2478>	PARENTHESIZED DIGIT FIVE
5343	<(6)>	<U2479>	PARENTHESIZED DIGIT SIX
5344	<(7)>	<U247A>	PARENTHESIZED DIGIT SEVEN
5345	<(8)>	<U247B>	PARENTHESIZED DIGIT EIGHT
5346	<(9)>	<U247C>	PARENTHESIZED DIGIT NINE
5347	<(10)>	<U247D>	PARENTHESIZED NUMBER TEN
5348	<(11)>	<U247E>	PARENTHESIZED NUMBER ELEVEN
5349	<(12)>	<U247F>	PARENTHESIZED NUMBER TWELVE
5350	<(13)>	<U2480>	PARENTHESIZED NUMBER THIRTEEN
5351	<(14)>	<U2481>	PARENTHESIZED NUMBER FOURTEEN
5352	<(15)>	<U2482>	PARENTHESIZED NUMBER FIFTEEN
5353	<(16)>	<U2483>	PARENTHESIZED NUMBER SIXTEEN
5354	<(17)>	<U2484>	PARENTHESIZED NUMBER SEVENTEEN
5355	<(18)>	<U2485>	PARENTHESIZED NUMBER EIGHTEEN
5356	<(19)>	<U2486>	PARENTHESIZED NUMBER NINETEEN
5357	<(20)>	<U2487>	PARENTHESIZED NUMBER TWENTY
5358	<1.>	<U2488>	DIGIT ONE FULL STOP
5359	<2.>	<U2489>	DIGIT TWO FULL STOP
5360	<3.>	<U248A>	DIGIT THREE FULL STOP
5361	<4.>	<U248B>	DIGIT FOUR FULL STOP
5362	<5.>	<U248C>	DIGIT FIVE FULL STOP
5363	<6.>	<U248D>	DIGIT SIX FULL STOP
5364	<7.>	<U248E>	DIGIT SEVEN FULL STOP
5365	<8.>	<U248F>	DIGIT EIGHT FULL STOP
5366	<9.>	<U2490>	DIGIT NINE FULL STOP
5367	<10.>	<U2491>	NUMBER TEN FULL STOP
5368	<11.>	<U2492>	NUMBER ELEVEN FULL STOP
5369	<12.>	<U2493>	NUMBER TWELVE FULL STOP
5370	<13.>	<U2494>	NUMBER THIRTEEN FULL STOP
5371	<14.>	<U2495>	NUMBER FOURTEEN FULL STOP
5372	<15.>	<U2496>	NUMBER FIFTEEN FULL STOP
5373	<16.>	<U2497>	NUMBER SIXTEEN FULL STOP
5374	<17.>	<U2498>	NUMBER SEVENTEEN FULL STOP
5375	<18.>	<U2499>	NUMBER EIGHTEEN FULL STOP
5376	<19.>	<U249A>	NUMBER NINETEEN FULL STOP
5377	<20.>	<U249B>	NUMBER TWENTY FULL STOP

5378	<(a)>	<U249C>	PARENTHESIZED LATIN SMALL LETTER A
5379	<(b)>	<U249D>	PARENTHESIZED LATIN SMALL LETTER B
5380	<(c)>	<U249E>	PARENTHESIZED LATIN SMALL LETTER C
5381	<(d)>	<U249F>	PARENTHESIZED LATIN SMALL LETTER D
5382	<(e)>	<U24A0>	PARENTHESIZED LATIN SMALL LETTER E
5383	<(f)>	<U24A1>	PARENTHESIZED LATIN SMALL LETTER F
5384	<(g)>	<U24A2>	PARENTHESIZED LATIN SMALL LETTER G
5385	<(h)>	<U24A3>	PARENTHESIZED LATIN SMALL LETTER H
5386	<(i)>	<U24A4>	PARENTHESIZED LATIN SMALL LETTER I
5387	<(j)>	<U24A5>	PARENTHESIZED LATIN SMALL LETTER J
5388	<(k)>	<U24A6>	PARENTHESIZED LATIN SMALL LETTER K
5389	<(l)>	<U24A7>	PARENTHESIZED LATIN SMALL LETTER L
5390	<(m)>	<U24A8>	PARENTHESIZED LATIN SMALL LETTER M
5391	<(n)>	<U24A9>	PARENTHESIZED LATIN SMALL LETTER N
5392	<(o)>	<U24AA>	PARENTHESIZED LATIN SMALL LETTER O
5393	<(p)>	<U24AB>	PARENTHESIZED LATIN SMALL LETTER P
5394	<(q)>	<U24AC>	PARENTHESIZED LATIN SMALL LETTER Q
5395	<(r)>	<U24AD>	PARENTHESIZED LATIN SMALL LETTER R
5396	<(s)>	<U24AE>	PARENTHESIZED LATIN SMALL LETTER S
5397	<(t)>	<U24AF>	PARENTHESIZED LATIN SMALL LETTER T
5398	<(u)>	<U24B0>	PARENTHESIZED LATIN SMALL LETTER U
5399	<(v)>	<U24B1>	PARENTHESIZED LATIN SMALL LETTER V
5400	<(w)>	<U24B2>	PARENTHESIZED LATIN SMALL LETTER W
5401	<(x)>	<U24B3>	PARENTHESIZED LATIN SMALL LETTER X
5402	<(y)>	<U24B4>	PARENTHESIZED LATIN SMALL LETTER Y
5403	<(z)>	<U24B5>	PARENTHESIZED LATIN SMALL LETTER Z
5404	<A-o>	<U24B6>	CIRCLED LATIN CAPITAL LETTER A
5405	<B-o>	<U24B7>	CIRCLED LATIN CAPITAL LETTER B
5406	<C-o>	<U24B8>	CIRCLED LATIN CAPITAL LETTER C
5407	<D-o>	<U24B9>	CIRCLED LATIN CAPITAL LETTER D
5408	<E-o>	<U24BA>	CIRCLED LATIN CAPITAL LETTER E
5409	<F-o>	<U24BB>	CIRCLED LATIN CAPITAL LETTER F
5410	<G-o>	<U24BC>	CIRCLED LATIN CAPITAL LETTER G
5411	<H-o>	<U24BD>	CIRCLED LATIN CAPITAL LETTER H
5412	<I-o>	<U24BE>	CIRCLED LATIN CAPITAL LETTER I
5413	<J-o>	<U24BF>	CIRCLED LATIN CAPITAL LETTER J
5414	<K-o>	<U24C0>	CIRCLED LATIN CAPITAL LETTER K
5415	<L-o>	<U24C1>	CIRCLED LATIN CAPITAL LETTER L
5416	<M-o>	<U24C2>	CIRCLED LATIN CAPITAL LETTER M
5417	<N-o>	<U24C3>	CIRCLED LATIN CAPITAL LETTER N
5418	<O-o>	<U24C4>	CIRCLED LATIN CAPITAL LETTER O
5419	<P-o>	<U24C5>	CIRCLED LATIN CAPITAL LETTER P
5420	<Q-o>	<U24C6>	CIRCLED LATIN CAPITAL LETTER Q
5421	<R-o>	<U24C7>	CIRCLED LATIN CAPITAL LETTER R
5422	<S-o>	<U24C8>	CIRCLED LATIN CAPITAL LETTER S
5423	<T-o>	<U24C9>	CIRCLED LATIN CAPITAL LETTER T
5424	<U-o>	<U24CA>	CIRCLED LATIN CAPITAL LETTER U
5425	<V-o>	<U24CB>	CIRCLED LATIN CAPITAL LETTER V
5426	<W-o>	<U24CC>	CIRCLED LATIN CAPITAL LETTER W
5427	<X-o>	<U24CD>	CIRCLED LATIN CAPITAL LETTER X
5428	<Y-o>	<U24CE>	CIRCLED LATIN CAPITAL LETTER Y
5429	<Z-o>	<U24CF>	CIRCLED LATIN CAPITAL LETTER Z
5430	<a-o>	<U24D0>	CIRCLED LATIN SMALL LETTER A
5431	<b-o>	<U24D1>	CIRCLED LATIN SMALL LETTER B
5432	<c-o>	<U24D2>	CIRCLED LATIN SMALL LETTER C
5433	<d-o>	<U24D3>	CIRCLED LATIN SMALL LETTER D
5434	<e-o>	<U24D4>	CIRCLED LATIN SMALL LETTER E
5435	<f-o>	<U24D5>	CIRCLED LATIN SMALL LETTER F
5436	<g-o>	<U24D6>	CIRCLED LATIN SMALL LETTER G
5437	<h-o>	<U24D7>	CIRCLED LATIN SMALL LETTER H
5438	<i-o>	<U24D8>	CIRCLED LATIN SMALL LETTER I
5439	<j-o>	<U24D9>	CIRCLED LATIN SMALL LETTER J
5440	<k-o>	<U24DA>	CIRCLED LATIN SMALL LETTER K
5441	<l-o>	<U24DB>	CIRCLED LATIN SMALL LETTER L
5442	<m-o>	<U24DC>	CIRCLED LATIN SMALL LETTER M
5443	<n-o>	<U24DD>	CIRCLED LATIN SMALL LETTER N
5444	<o-o>	<U24DE>	CIRCLED LATIN SMALL LETTER O
5445	<p-o>	<U24DF>	CIRCLED LATIN SMALL LETTER P
5446	<q-o>	<U24E0>	CIRCLED LATIN SMALL LETTER Q
5447	<r-o>	<U24E1>	CIRCLED LATIN SMALL LETTER R
5448	<s-o>	<U24E2>	CIRCLED LATIN SMALL LETTER S
5449	<t-o>	<U24E3>	CIRCLED LATIN SMALL LETTER T
5450	<u-o>	<U24E4>	CIRCLED LATIN SMALL LETTER U
5451	<v-o>	<U24E5>	CIRCLED LATIN SMALL LETTER V
5452	<w-o>	<U24E6>	CIRCLED LATIN SMALL LETTER W
5453	<x-o>	<U24E7>	CIRCLED LATIN SMALL LETTER X
5454	<y-o>	<U24E8>	CIRCLED LATIN SMALL LETTER Y
5455	<z-o>	<U24E9>	CIRCLED LATIN SMALL LETTER Z
5456	<0-o>	<U24EA>	CIRCLED DIGIT ZERO
5457	<hh>	<U2500>	BOX DRAWINGS LIGHT HORIZONTAL
5458	<HH->	<U2501>	BOX DRAWINGS HEAVY HORIZONTAL
5459	<vv>	<U2502>	BOX DRAWINGS LIGHT VERTICAL
5460	<VV->	<U2503>	BOX DRAWINGS HEAVY VERTICAL
5461	<3->	<U2504>	BOX DRAWINGS LIGHT TRIPLE DASH HORIZONTAL
5462	<3_>	<U2505>	BOX DRAWINGS HEAVY TRIPLE DASH HORIZONTAL
5463	<3!>	<U2506>	BOX DRAWINGS LIGHT TRIPLE DASH VERTICAL
5464	<3//>	<U2507>	BOX DRAWINGS HEAVY TRIPLE DASH VERTICAL
5465	<4->	<U2508>	BOX DRAWINGS LIGHT QUADRUPLE DASH HORIZONTAL
5466	<4_>	<U2509>	BOX DRAWINGS HEAVY QUADRUPLE DASH HORIZONTAL

5467	<4!>	<U250A>	BOX DRAWINGS LIGHT QUADRUPLE DASH VERTICAL
5468	<4//>	<U250B>	BOX DRAWINGS HEAVY QUADRUPLE DASH VERTICAL
5469	<dr>	<U250C>	BOX DRAWINGS LIGHT DOWN AND RIGHT
5470	<dr->	<U250D>	BOX DRAWINGS DOWN LIGHT AND RIGHT HEAVY
5471	<Dr->	<U250E>	BOX DRAWINGS DOWN HEAVY AND RIGHT LIGHT
5472	<DR->	<U250F>	BOX DRAWINGS HEAVY DOWN AND RIGHT
5473	<d!>	<U2510>	BOX DRAWINGS LIGHT DOWN AND LEFT
5474	<d!->	<U2511>	BOX DRAWINGS DOWN LIGHT AND LEFT HEAVY
5475	<D!->	<U2512>	BOX DRAWINGS DOWN HEAVY AND LEFT LIGHT
5476	<LD->	<U2513>	BOX DRAWINGS HEAVY DOWN AND LEFT
5477	<ur>	<U2514>	BOX DRAWINGS LIGHT UP AND RIGHT
5478	<uR->	<U2515>	BOX DRAWINGS UP LIGHT AND RIGHT HEAVY
5479	<Ur->	<U2516>	BOX DRAWINGS UP HEAVY AND RIGHT LIGHT
5480	<UR->	<U2517>	BOX DRAWINGS HEAVY UP AND RIGHT
5481		<U2518>	BOX DRAWINGS LIGHT UP AND LEFT
5482	<uL->	<U2519>	BOX DRAWINGS UP LIGHT AND LEFT HEAVY
5483	<Ul->	<U251A>	BOX DRAWINGS UP HEAVY AND LEFT LIGHT
5484	<UL->	<U251B>	BOX DRAWINGS HEAVY UP AND LEFT
5485	<vr>	<U251C>	BOX DRAWINGS LIGHT VERTICAL AND RIGHT
5486	<vR->	<U251D>	BOX DRAWINGS VERTICAL LIGHT AND RIGHT HEAVY
5487	<Udr>	<U251E>	BOX DRAWINGS UP HEAVY AND RIGHT DOWN LIGHT
5488	<uDr>	<U251F>	BOX DRAWINGS DOWN HEAVY AND RIGHT UP LIGHT
5489	<Vr->	<U2520>	BOX DRAWINGS VERTICAL HEAVY AND RIGHT LIGHT
5490	<UdR>	<U2521>	BOX DRAWINGS DOWN LIGHT AND RIGHT UP HEAVY
5491	<uDR>	<U2522>	BOX DRAWINGS UP LIGHT AND RIGHT DOWN HEAVY
5492	<VR->	<U2523>	BOX DRAWINGS HEAVY VERTICAL AND RIGHT
5493	<v!>	<U2524>	BOX DRAWINGS LIGHT VERTICAL AND LEFT
5494	<v!->	<U2525>	BOX DRAWINGS VERTICAL LIGHT AND LEFT HEAVY
5495	<Ud!>	<U2526>	BOX DRAWINGS UP HEAVY AND LEFT DOWN LIGHT
5496	<uD!>	<U2527>	BOX DRAWINGS DOWN HEAVY AND LEFT UP LIGHT
5497	<V!->	<U2528>	BOX DRAWINGS VERTICAL HEAVY AND LEFT LIGHT
5498	<Ud!>	<U2529>	BOX DRAWINGS DOWN LIGHT AND LEFT UP HEAVY
5499	<uD!>	<U252A>	BOX DRAWINGS UP LIGHT AND LEFT DOWN HEAVY
5500	<VL->	<U252B>	BOX DRAWINGS HEAVY VERTICAL AND LEFT
5501	<dh>	<U252C>	BOX DRAWINGS LIGHT DOWN AND HORIZONTAL
5502	<dLr>	<U252D>	BOX DRAWINGS LEFT HEAVY AND RIGHT DOWN LIGHT
5503	<d!R>	<U252E>	BOX DRAWINGS RIGHT HEAVY AND LEFT DOWN LIGHT
5504	<dH->	<U252F>	BOX DRAWINGS DOWN LIGHT AND HORIZONTAL HEAVY
5505	<Dh->	<U2530>	BOX DRAWINGS DOWN HEAVY AND HORIZONTAL LIGHT
5506	<DLr>	<U2531>	BOX DRAWINGS RIGHT LIGHT AND LEFT DOWN HEAVY
5507	<D!R>	<U2532>	BOX DRAWINGS LEFT LIGHT AND RIGHT DOWN HEAVY
5508	<DH->	<U2533>	BOX DRAWINGS HEAVY DOWN AND HORIZONTAL
5509	<uh>	<U2534>	BOX DRAWINGS LIGHT UP AND HORIZONTAL
5510	<uLr>	<U2535>	BOX DRAWINGS LEFT HEAVY AND RIGHT UP LIGHT
5511	<u!R>	<U2536>	BOX DRAWINGS RIGHT HEAVY AND LEFT UP LIGHT
5512	<uH->	<U2537>	BOX DRAWINGS UP LIGHT AND HORIZONTAL HEAVY
5513	<Uh->	<U2538>	BOX DRAWINGS UP HEAVY AND HORIZONTAL LIGHT
5514	<ULr>	<U2539>	BOX DRAWINGS RIGHT LIGHT AND LEFT UP HEAVY
5515	<U!R>	<U253A>	BOX DRAWINGS LEFT LIGHT AND RIGHT UP HEAVY
5516	<UH->	<U253B>	BOX DRAWINGS HEAVY UP AND HORIZONTAL
5517	<vh>	<U253C>	BOX DRAWINGS LIGHT VERTICAL AND HORIZONTAL
5518	<vLr>	<U253D>	BOX DRAWINGS LEFT HEAVY AND RIGHT VERTICAL LIGHT
5519	<v!R>	<U253E>	BOX DRAWINGS RIGHT HEAVY AND LEFT VERTICAL LIGHT
5520	<vH->	<U253F>	BOX DRAWINGS VERTICAL LIGHT AND HORIZONTAL HEAVY
5521	<Udh>	<U2540>	BOX DRAWINGS UP HEAVY AND DOWN HORIZONTAL LIGHT
5522	<udh>	<U2541>	BOX DRAWINGS DOWN HEAVY AND UP HORIZONTAL LIGHT
5523	<Vh->	<U2542>	BOX DRAWINGS VERTICAL HEAVY AND HORIZONTAL LIGHT
5524	<UdLr>	<U2543>	BOX DRAWINGS LEFT UP HEAVY AND RIGHT DOWN LIGHT
5525	<Ud!R>	<U2544>	BOX DRAWINGS RIGHT UP HEAVY AND LEFT DOWN LIGHT
5526	<uD!R>	<U2545>	BOX DRAWINGS LEFT DOWN HEAVY AND RIGHT UP LIGHT
5527	<uD!R>	<U2546>	BOX DRAWINGS RIGHT DOWN HEAVY AND LEFT UP LIGHT
5528	<UdH>	<U2547>	BOX DRAWINGS DOWN LIGHT AND UP HORIZONTAL HEAVY
5529	<uDh>	<U2548>	BOX DRAWINGS UP LIGHT AND DOWN HORIZONTAL HEAVY
5530	<VLr>	<U2549>	BOX DRAWINGS RIGHT LIGHT AND LEFT VERTICAL HEAVY
5531	<V!R>	<U254A>	BOX DRAWINGS LEFT LIGHT AND RIGHT VERTICAL HEAVY
5532	<VH->	<U254B>	BOX DRAWINGS HEAVY VERTICAL AND HORIZONTAL
5533	<HH>	<U2550>	BOX DRAWINGS DOUBLE HORIZONTAL
5534	<VV>	<U2551>	BOX DRAWINGS DOUBLE VERTICAL
5535	<dR>	<U2552>	BOX DRAWINGS DOWN SINGLE AND RIGHT DOUBLE
5536	<Dr>	<U2553>	BOX DRAWINGS DOWN DOUBLE AND RIGHT SINGLE
5537	<DR>	<U2554>	BOX DRAWINGS DOUBLE DOWN AND RIGHT
5538	<d!>	<U2555>	BOX DRAWINGS DOWN SINGLE AND LEFT DOUBLE
5539	<D!>	<U2556>	BOX DRAWINGS DOWN DOUBLE AND LEFT SINGLE
5540	<LD>	<U2557>	BOX DRAWINGS DOUBLE DOWN AND LEFT
5541	<uR>	<U2558>	BOX DRAWINGS UP SINGLE AND RIGHT DOUBLE
5542	<Ur>	<U2559>	BOX DRAWINGS UP DOUBLE AND RIGHT SINGLE
5543	<UR>	<U255A>	BOX DRAWINGS DOUBLE UP AND RIGHT
5544		<U255B>	BOX DRAWINGS UP SINGLE AND LEFT DOUBLE
5545		<U255C>	BOX DRAWINGS UP DOUBLE AND LEFT SINGLE
5546		<U255D>	BOX DRAWINGS DOUBLE UP AND LEFT
5547	<vR>	<U255E>	BOX DRAWINGS VERTICAL SINGLE AND RIGHT DOUBLE
5548	<Vr>	<U255F>	BOX DRAWINGS VERTICAL DOUBLE AND RIGHT SINGLE
5549	<VR>	<U2560>	BOX DRAWINGS DOUBLE VERTICAL AND RIGHT
5550	<vL>	<U2561>	BOX DRAWINGS VERTICAL SINGLE AND LEFT DOUBLE
5551	<Vl>	<U2562>	BOX DRAWINGS VERTICAL DOUBLE AND LEFT SINGLE
5552	<VL>	<U2563>	BOX DRAWINGS DOUBLE VERTICAL AND LEFT
5553	<dH>	<U2564>	BOX DRAWINGS DOWN SINGLE AND HORIZONTAL DOUBLE
5554	<Dh>	<U2565>	BOX DRAWINGS DOWN DOUBLE AND HORIZONTAL SINGLE
5555	<DH>	<U2566>	BOX DRAWINGS DOUBLE DOWN AND HORIZONTAL

5556	<uH>	<U2567>	BOX DRAWINGS UP SINGLE AND HORIZONTAL DOUBLE
5557	<Uh>	<U2568>	BOX DRAWINGS UP DOUBLE AND HORIZONTAL SINGLE
5558	<UH>	<U2569>	BOX DRAWINGS DOUBLE UP AND HORIZONTAL
5559	<vH>	<U256A>	BOX DRAWINGS VERTICAL SINGLE AND HORIZONTAL DOUBLE
5560	<Vh>	<U256B>	BOX DRAWINGS VERTICAL DOUBLE AND HORIZONTAL SINGLE
5561	<VH>	<U256C>	BOX DRAWINGS DOUBLE VERTICAL AND HORIZONTAL
5562	<FD>	<U2571>	BOX DRAWINGS LIGHT DIAGONAL UPPER RIGHT TO LOWER LEFT
5563	<BD>	<U2572>	BOX DRAWINGS LIGHT DIAGONAL UPPER LEFT TO LOWER RIGHT
5564	<TB>	<U2580>	UPPER HALF BLOCK
5565	<LB>	<U2584>	LOWER HALF BLOCK
5566	<FB>	<U2588>	FULL BLOCK
5567	<LB>	<U258C>	LEFT HALF BLOCK
5568	<RB>	<U2590>	RIGHT HALF BLOCK
5569	<.S>	<U2591>	LIGHT SHADE
5570	<:S>	<U2592>	MEDIUM SHADE
5571	<?S>	<U2593>	DARK SHADE
5572	<fS>	<U25A0>	BLACK SQUARE
5573	<oS>	<U25A1>	WHITE SQUARE
5574	<RO>	<U25A2>	WHITE SQUARE WITH ROUNDED CORNERS
5575	<Rr>	<U25A3>	WHITE SQUARE CONTAINING BLACK SMALL SQUARE
5576	<RF>	<U25A4>	SQUARE WITH HORIZONTAL FILL
5577	<RY>	<U25A5>	SQUARE WITH VERTICAL FILL
5578	<RH>	<U25A6>	SQUARE WITH ORTHOGONAL CROSSHATCH FILL
5579	<RZ>	<U25A7>	SQUARE WITH UPPER LEFT TO LOWER RIGHT FILL
5580	<RK>	<U25A8>	SQUARE WITH UPPER RIGHT TO LOWER LEFT FILL
5581	<RX>	<U25A9>	SQUARE WITH DIAGONAL CROSSHATCH FILL
5582	<sB>	<U25AA>	BLACK SMALL SQUARE
5583	<SR>	<U25AC>	BLACK RECTANGLE
5584	<Or>	<U25AD>	WHITE RECTANGLE
5585	<UT>	<U25B2>	BLACK UP-POINTING TRIANGLE
5586	<uT>	<U25B3>	WHITE UP-POINTING TRIANGLE
5587	<Tr>	<U25B7>	WHITE RIGHT-POINTING TRIANGLE
5588	<PR>	<U25BA>	BLACK RIGHT-POINTING POINTER
5589	<Dt>	<U25BC>	BLACK DOWN-POINTING TRIANGLE
5590	<dT>	<U25BD>	WHITE DOWN-POINTING TRIANGLE
5591	<Tl>	<U25C1>	WHITE LEFT-POINTING TRIANGLE
5592	<PL>	<U25C4>	BLACK LEFT-POINTING POINTER
5593	<Db>	<U25C6>	BLACK DIAMOND
5594	<Dw>	<U25C7>	WHITE DIAMOND
5595	<LZ>	<U25CA>	LOZENGE
5596	<Om>	<U25CB>	WHITE CIRCLE
5597	<Oo>	<U25CE>	BULLSEYE
5598	<OM>	<U25CF>	BLACK CIRCLE
5599		<U25D0>	CIRCLE WITH LEFT HALF BLACK
5600	<OR>	<U25D1>	CIRCLE WITH RIGHT HALF BLACK
5601	<Sn>	<U25D8>	INVERSE BULLET
5602	<Ic>	<U25D9>	INVERSE WHITE CIRCLE
5603	<Fd>	<U25E2>	BLACK LOWER RIGHT TRIANGLE
5604	<Bd>	<U25E3>	BLACK LOWER LEFT TRIANGLE
5605	<Ci>	<U25EF>	LARGE CIRCLE
5606	<*2>	<U2605>	BLACK STAR
5607	<*1>	<U2606>	WHITE STAR
5608	<TEL>	<U260E>	BLACK TELEPHONE
5609	<tel>	<U260F>	WHITE TELEPHONE
5610	<<H>	<U261C>	WHITE LEFT POINTING INDEX
5611	</>H>	<U261E>	WHITE RIGHT POINTING INDEX
5612	<Ou>	<U263A>	WHITE SMILING FACE
5613	<OU>	<U263B>	BLACK SMILING FACE
5614	<SU>	<U263C>	WHITE SUN WITH RAYS
5615	<Fm>	<U2640>	FEMALE SIGN
5616	<Ml>	<U2642>	MALE SIGN
5617	<cS>	<U2660>	BLACK SPADE SUIT
5618	<cH>	<U2661>	WHITE HEART SUIT
5619	<cD>	<U2662>	WHITE DIAMOND SUIT
5620	<cC>	<U2663>	BLACK CLUB SUIT
5621	<cS->	<U2664>	WHITE SPADE SUIT
5622	<cH->	<U2665>	BLACK HEART SUIT
5623	<cD->	<U2666>	BLACK DIAMOND SUIT
5624	<cC->	<U2667>	WHITE CLUB SUIT
5625	<Md>	<U2669>	QUARTER NOTE
5626	<M8>	<U266A>	EIGHTH NOTE
5627	<M2>	<U266B>	BEAMED EIGHTH NOTES
5628	<M16>	<U266C>	BEAMED SIXTEENTH NOTES
5629	<Mb>	<U266D>	MUSIC FLAT SIGN
5630	<Mx>	<U266E>	MUSIC NATURAL SIGN
5631	<MX>	<U266F>	MUSIC SHARP SIGN
5632	<OK>	<U2713>	CHECK MARK
5633	<XX>	<U2717>	BALLOT X
5634	<-X>	<U2720>	MALTESE CROSS
5635	<IS>	<U3000>	IDEOGRAPHIC SPACE
5636	< , _>	<U3001>	IDEOGRAPHIC COMMA
5637	< . _>	<U3002>	IDEOGRAPHIC FULL STOP
5638	<+ ">	<U3003>	DITTO MARK
5639	<JIS>	<U3004>	JAPANESE INDUSTRIAL STANDARD SYMBOL
5640	<* _>	<U3005>	IDEOGRAPHIC ITERATION MARK
5641	< ; _>	<U3006>	IDEOGRAPHIC CLOSING MARK
5642	<0 _>	<U3007>	IDEOGRAPHIC NUMBER ZERO
5643	<<+>	<U300A>	LEFT DOUBLE ANGLE BRACKET
5644	</>+>	<U300B>	RIGHT DOUBLE ANGLE BRACKET

5645	<<'>	<U300C>	LEFT CORNER BRACKET
5646	</>'>	<U300D>	RIGHT CORNER BRACKET
5647	<<">	<U300E>	LEFT WHITE CORNER BRACKET
5648	</>">	<U300F>	RIGHT WHITE CORNER BRACKET
5649	<(" >	<U3010>	LEFT BLACK LENTICULAR BRACKET
5650	<)" >	<U3011>	RIGHT BLACK LENTICULAR BRACKET
5651	<=T>	<U3012>	POSTAL MARK
5652	<= >	<U3013>	GETA MARK
5653	<(' >	<U3014>	LEFT TORTOISE SHELL BRACKET
5654	<)' >	<U3015>	RIGHT TORTOISE SHELL BRACKET
5655	<(I >	<U3016>	LEFT WHITE LENTICULAR BRACKET
5656	<)'I >	<U3017>	RIGHT WHITE LENTICULAR BRACKET
5657	<-?>	<U301C>	WAVE DASH
5658	<=T:)>	<U3020>	POSTAL MARK FACE
5659	<A5>	<U3041>	HIRAGANA LETTER SMALL A
5660	<a5>	<U3042>	HIRAGANA LETTER A
5661	<I5>	<U3043>	HIRAGANA LETTER SMALL I
5662	<i5>	<U3044>	HIRAGANA LETTER I
5663	<U5>	<U3045>	HIRAGANA LETTER SMALL U
5664	<u5>	<U3046>	HIRAGANA LETTER U
5665	<E5>	<U3047>	HIRAGANA LETTER SMALL E
5666	<e5>	<U3048>	HIRAGANA LETTER E
5667	<O5>	<U3049>	HIRAGANA LETTER SMALL O
5668	<o5>	<U304A>	HIRAGANA LETTER O
5669	<ka>	<U304B>	HIRAGANA LETTER KA
5670	<ga>	<U304C>	HIRAGANA LETTER GA
5671	<ki>	<U304D>	HIRAGANA LETTER KI
5672	<gi>	<U304E>	HIRAGANA LETTER GI
5673	<ku>	<U304F>	HIRAGANA LETTER KU
5674	<gu>	<U3050>	HIRAGANA LETTER GU
5675	<ke>	<U3051>	HIRAGANA LETTER KE
5676	<ge>	<U3052>	HIRAGANA LETTER GE
5677	<ko>	<U3053>	HIRAGANA LETTER KO
5678	<go>	<U3054>	HIRAGANA LETTER GO
5679	<sa>	<U3055>	HIRAGANA LETTER SA
5680	<za>	<U3056>	HIRAGANA LETTER ZA
5681	<si>	<U3057>	HIRAGANA LETTER SI
5682	<zi>	<U3058>	HIRAGANA LETTER ZI
5683	<su>	<U3059>	HIRAGANA LETTER SU
5684	<zu>	<U305A>	HIRAGANA LETTER ZU
5685	<se>	<U305B>	HIRAGANA LETTER SE
5686	<ze>	<U305C>	HIRAGANA LETTER ZE
5687	<so>	<U305D>	HIRAGANA LETTER SO
5688	<zo>	<U305E>	HIRAGANA LETTER ZO
5689	<ta>	<U305F>	HIRAGANA LETTER TA
5690	<da>	<U3060>	HIRAGANA LETTER DA
5691	<ti>	<U3061>	HIRAGANA LETTER TI
5692	<di>	<U3062>	HIRAGANA LETTER DI
5693	<tU>	<U3063>	HIRAGANA LETTER SMALL TU
5694	<tu>	<U3064>	HIRAGANA LETTER TU
5695	<du>	<U3065>	HIRAGANA LETTER DU
5696	<te>	<U3066>	HIRAGANA LETTER TE
5697	<de>	<U3067>	HIRAGANA LETTER DE
5698	<to>	<U3068>	HIRAGANA LETTER TO
5699	<do>	<U3069>	HIRAGANA LETTER DO
5700	<na>	<U306A>	HIRAGANA LETTER NA
5701	<ni>	<U306B>	HIRAGANA LETTER NI
5702	<nu>	<U306C>	HIRAGANA LETTER NU
5703	<ne>	<U306D>	HIRAGANA LETTER NE
5704	<no>	<U306E>	HIRAGANA LETTER NO
5705	<ha>	<U306F>	HIRAGANA LETTER HA
5706	<ba>	<U3070>	HIRAGANA LETTER BA
5707	<pa>	<U3071>	HIRAGANA LETTER PA
5708	<hi>	<U3072>	HIRAGANA LETTER HI
5709	<bi>	<U3073>	HIRAGANA LETTER BI
5710	<pi>	<U3074>	HIRAGANA LETTER PI
5711	<hu>	<U3075>	HIRAGANA LETTER HU
5712	<bu>	<U3076>	HIRAGANA LETTER BU
5713	<pu>	<U3077>	HIRAGANA LETTER PU
5714	<he>	<U3078>	HIRAGANA LETTER HE
5715	<be>	<U3079>	HIRAGANA LETTER BE
5716	<pe>	<U307A>	HIRAGANA LETTER PE
5717	<ho>	<U307B>	HIRAGANA LETTER HO
5718	<bo>	<U307C>	HIRAGANA LETTER BO
5719	<po>	<U307D>	HIRAGANA LETTER PO
5720	<ma>	<U307E>	HIRAGANA LETTER MA
5721	<mi>	<U307F>	HIRAGANA LETTER MI
5722	<mu>	<U3080>	HIRAGANA LETTER MU
5723	<me>	<U3081>	HIRAGANA LETTER ME
5724	<mo>	<U3082>	HIRAGANA LETTER MO
5725	<yA>	<U3083>	HIRAGANA LETTER SMALL YA
5726	<ya>	<U3084>	HIRAGANA LETTER YA
5727	<yU>	<U3085>	HIRAGANA LETTER SMALL YU
5728	<yu>	<U3086>	HIRAGANA LETTER YU
5729	<yO>	<U3087>	HIRAGANA LETTER SMALL YO
5730	<yo>	<U3088>	HIRAGANA LETTER YO
5731	<ra>	<U3089>	HIRAGANA LETTER RA
5732	<ri>	<U308A>	HIRAGANA LETTER RI
5733	<ru>	<U308B>	HIRAGANA LETTER RU

5734	<re>	<U308C>	HIRAGANA LETTER RE
5735	<ro>	<U308D>	HIRAGANA LETTER RO
5736	<wA>	<U308E>	HIRAGANA LETTER SMALL WA
5737	<wa>	<U308F>	HIRAGANA LETTER WA
5738	<wi>	<U3090>	HIRAGANA LETTER WI
5739	<we>	<U3091>	HIRAGANA LETTER WE
5740	<wo>	<U3092>	HIRAGANA LETTER WO
5741	<n5>	<U3093>	HIRAGANA LETTER N
5742	<vu>	<U3094>	HIRAGANA LETTER VU
5743	<"5>	<U309B>	KATAKANA-HIRAGANA VOICED SOUND MARK
5744	<05>	<U309C>	KATAKANA-HIRAGANA SEMI-VOICED SOUND MARK
5745	<*5>	<U309D>	HIRAGANA ITERATION MARK
5746	<+5>	<U309E>	HIRAGANA VOICED ITERATION MARK
5747	<a6>	<U30A1>	KATAKANA LETTER SMALL A
5748	<A6>	<U30A2>	KATAKANA LETTER A
5749	<i6>	<U30A3>	KATAKANA LETTER SMALL I
5750	<I6>	<U30A4>	KATAKANA LETTER I
5751	<u6>	<U30A5>	KATAKANA LETTER SMALL U
5752	<U6>	<U30A6>	KATAKANA LETTER U
5753	<e6>	<U30A7>	KATAKANA LETTER SMALL E
5754	<E6>	<U30A8>	KATAKANA LETTER E
5755	<o6>	<U30A9>	KATAKANA LETTER SMALL O
5756	<O6>	<U30AA>	KATAKANA LETTER O
5757	<Ka>	<U30AB>	KATAKANA LETTER KA
5758	<Ga>	<U30AC>	KATAKANA LETTER GA
5759	<Ki>	<U30AD>	KATAKANA LETTER KI
5760	<Gi>	<U30AE>	KATAKANA LETTER GI
5761	<Ku>	<U30AF>	KATAKANA LETTER KU
5762	<Gu>	<U30B0>	KATAKANA LETTER GU
5763	<Ke>	<U30B1>	KATAKANA LETTER KE
5764	<Ge>	<U30B2>	KATAKANA LETTER GE
5765	<Ko>	<U30B3>	KATAKANA LETTER KO
5766	<Go>	<U30B4>	KATAKANA LETTER GO
5767	<Sa>	<U30B5>	KATAKANA LETTER SA
5768	<Za>	<U30B6>	KATAKANA LETTER ZA
5769	<Si>	<U30B7>	KATAKANA LETTER SI
5770	<Zi>	<U30B8>	KATAKANA LETTER ZI
5771	<Su>	<U30B9>	KATAKANA LETTER SU
5772	<Zu>	<U30BA>	KATAKANA LETTER ZU
5773	<Se>	<U30BB>	KATAKANA LETTER SE
5774	<Ze>	<U30BC>	KATAKANA LETTER ZE
5775	<So>	<U30BD>	KATAKANA LETTER SO
5776	<Zo>	<U30BE>	KATAKANA LETTER ZO
5777	<Ta>	<U30BF>	KATAKANA LETTER TA
5778	<Da>	<U30C0>	KATAKANA LETTER DA
5779	<Ti>	<U30C1>	KATAKANA LETTER TI
5780	<Di>	<U30C2>	KATAKANA LETTER DI
5781	<Tu>	<U30C3>	KATAKANA LETTER SMALL TU
5782	<Tu>	<U30C4>	KATAKANA LETTER TU
5783	<Du>	<U30C5>	KATAKANA LETTER DU
5784	<Te>	<U30C6>	KATAKANA LETTER TE
5785	<De>	<U30C7>	KATAKANA LETTER DE
5786	<To>	<U30C8>	KATAKANA LETTER TO
5787	<Do>	<U30C9>	KATAKANA LETTER DO
5788	<Na>	<U30CA>	KATAKANA LETTER NA
5789	<Ni>	<U30CB>	KATAKANA LETTER NI
5790	<Nu>	<U30CC>	KATAKANA LETTER NU
5791	<Ne>	<U30CD>	KATAKANA LETTER NE
5792	<No>	<U30CE>	KATAKANA LETTER NO
5793	<Ha>	<U30CF>	KATAKANA LETTER HA
5794	<Ba>	<U30D0>	KATAKANA LETTER BA
5795	<Pa>	<U30D1>	KATAKANA LETTER PA
5796	<Hi>	<U30D2>	KATAKANA LETTER HI
5797	<Bi>	<U30D3>	KATAKANA LETTER BI
5798	<Pi>	<U30D4>	KATAKANA LETTER PI
5799	<Hu>	<U30D5>	KATAKANA LETTER HU
5800	<Bu>	<U30D6>	KATAKANA LETTER BU
5801	<Pu>	<U30D7>	KATAKANA LETTER PU
5802	<He>	<U30D8>	KATAKANA LETTER HE
5803	<Be>	<U30D9>	KATAKANA LETTER BE
5804	<Pe>	<U30DA>	KATAKANA LETTER PE
5805	<Ho>	<U30DB>	KATAKANA LETTER HO
5806	<Bo>	<U30DC>	KATAKANA LETTER BO
5807	<Po>	<U30DD>	KATAKANA LETTER PO
5808	<Ma>	<U30DE>	KATAKANA LETTER MA
5809	<Mi>	<U30DF>	KATAKANA LETTER MI
5810	<Mu>	<U30E0>	KATAKANA LETTER MU
5811	<Me>	<U30E1>	KATAKANA LETTER ME
5812	<Mo>	<U30E2>	KATAKANA LETTER MO
5813	<YA>	<U30E3>	KATAKANA LETTER SMALL YA
5814	<Ya>	<U30E4>	KATAKANA LETTER YA
5815	<YU>	<U30E5>	KATAKANA LETTER SMALL YU
5816	<Yu>	<U30E6>	KATAKANA LETTER YU
5817	<YO>	<U30E7>	KATAKANA LETTER SMALL YO
5818	<Yo>	<U30E8>	KATAKANA LETTER YO
5819	<Ra>	<U30E9>	KATAKANA LETTER RA
5820	<Ri>	<U30EA>	KATAKANA LETTER RI
5821	<Ru>	<U30EB>	KATAKANA LETTER RU
5822	<Re>	<U30EC>	KATAKANA LETTER RE

5823	<Ro>	<U30ED>	KATAKANA LETTER RO
5824	<WA>	<U30EE>	KATAKANA LETTER SMALL WA
5825	<Wa>	<U30EF>	KATAKANA LETTER WA
5826	<Wi>	<U30F0>	KATAKANA LETTER WI
5827	<We>	<U30F1>	KATAKANA LETTER WE
5828	<Wo>	<U30F2>	KATAKANA LETTER WO
5829	<N6>	<U30F3>	KATAKANA LETTER N
5830	<Vu>	<U30F4>	KATAKANA LETTER VU
5831	<KA>	<U30F5>	KATAKANA LETTER SMALL KA
5832	<KE>	<U30F6>	KATAKANA LETTER SMALL KE
5833	<Va>	<U30F7>	KATAKANA LETTER VA
5834	<Vi>	<U30F8>	KATAKANA LETTER VI
5835	<Ve>	<U30F9>	KATAKANA LETTER VE
5836	<Vo>	<U30FA>	KATAKANA LETTER VO
5837	<.6>	<U30FB>	KATAKANA MIDDLE DOT
5838	<-6>	<U30FC>	KATAKANA-HIRAGANA PROLONGED SOUND MARK
5839	<*6>	<U30FD>	KATAKANA ITERATION MARK
5840	<+6>	<U30FE>	KATAKANA VOICED ITERATION MARK
5841	<b4>	<U3105>	BOPOMOFO LETTER B
5842	<p4>	<U3106>	BOPOMOFO LETTER P
5843	<m4>	<U3107>	BOPOMOFO LETTER M
5844	<f4>	<U3108>	BOPOMOFO LETTER F
5845	<d4>	<U3109>	BOPOMOFO LETTER D
5846	<t4>	<U310A>	BOPOMOFO LETTER T
5847	<n4>	<U310B>	BOPOMOFO LETTER N
5848	<l4>	<U310C>	BOPOMOFO LETTER L
5849	<g4>	<U310D>	BOPOMOFO LETTER G
5850	<k4>	<U310E>	BOPOMOFO LETTER K
5851	<h4>	<U310F>	BOPOMOFO LETTER H
5852	<j4>	<U3110>	BOPOMOFO LETTER J
5853	<q4>	<U3111>	BOPOMOFO LETTER Q
5854	<x4>	<U3112>	BOPOMOFO LETTER X
5855	<zh>	<U3113>	BOPOMOFO LETTER ZH
5856	<ch>	<U3114>	BOPOMOFO LETTER CH
5857	<sh>	<U3115>	BOPOMOFO LETTER SH
5858	<r4>	<U3116>	BOPOMOFO LETTER R
5859	<z4>	<U3117>	BOPOMOFO LETTER Z
5860	<c4>	<U3118>	BOPOMOFO LETTER C
5861	<s4>	<U3119>	BOPOMOFO LETTER S
5862	<a4>	<U311A>	BOPOMOFO LETTER A
5863	<o4>	<U311B>	BOPOMOFO LETTER O
5864	<e4>	<U311C>	BOPOMOFO LETTER E
5865	<eh4>	<U311D>	BOPOMOFO LETTER EH
5866	<ai>	<U311E>	BOPOMOFO LETTER AI
5867	<ei>	<U311F>	BOPOMOFO LETTER EI
5868	<au>	<U3120>	BOPOMOFO LETTER AU
5869	<ou>	<U3121>	BOPOMOFO LETTER OU
5870	<an>	<U3122>	BOPOMOFO LETTER AN
5871	<en>	<U3123>	BOPOMOFO LETTER EN
5872	<aN>	<U3124>	BOPOMOFO LETTER ANG
5873	<eN>	<U3125>	BOPOMOFO LETTER ENG
5874	<er>	<U3126>	BOPOMOFO LETTER ER
5875	<i4>	<U3127>	BOPOMOFO LETTER I
5876	<u4>	<U3128>	BOPOMOFO LETTER U
5877	<iu>	<U3129>	BOPOMOFO LETTER IU
5878	<v4>	<U312A>	BOPOMOFO LETTER V
5879	<nG>	<U312B>	BOPOMOFO LETTER NG
5880	<gn>	<U312C>	BOPOMOFO LETTER GN
5881	<(JU)>	<U321C>	PARENTHESESIZED HANGUL CIEUC U
5882	<1c>	<U3220>	PARENTHESESIZED IDEOGRAPH ONE
5883	<2c>	<U3221>	PARENTHESESIZED IDEOGRAPH TWO
5884	<3c>	<U3222>	PARENTHESESIZED IDEOGRAPH THREE
5885	<4c>	<U3223>	PARENTHESESIZED IDEOGRAPH FOUR
5886	<5c>	<U3224>	PARENTHESESIZED IDEOGRAPH FIVE
5887	<6c>	<U3225>	PARENTHESESIZED IDEOGRAPH SIX
5888	<7c>	<U3226>	PARENTHESESIZED IDEOGRAPH SEVEN
5889	<8c>	<U3227>	PARENTHESESIZED IDEOGRAPH EIGHT
5890	<9c>	<U3228>	PARENTHESESIZED IDEOGRAPH NINE
5891	<10c>	<U3229>	PARENTHESESIZED IDEOGRAPH TEN
5892	<KSC>	<U327F>	KOREAN STANDARD SYMBOL
5893	<am>	<U33C2>	SQUARE AM
5894	<pm>	<U33D8>	SQUARE PM
5895	<ff>	<UFB00>	LATIN SMALL LIGATURE FF
5896	<fi>	<UFB01>	LATIN SMALL LIGATURE FI
5897	<fl>	<UFB02>	LATIN SMALL LIGATURE FL
5898	<ffi>	<UFB03>	LATIN SMALL LIGATURE FFI
5899	<ffl>	<UFB04>	LATIN SMALL LIGATURE FFL
5900	<St>	<UFB05>	LATIN SMALL LIGATURE LONG S T
5901	<st>	<UFB06>	LATIN SMALL LIGATURE ST
5902	<3+i>	<UFE7D>	ARABIC SHADDA MEDIAL FORM
5903	<aM.>	<UFE82>	ARABIC LETTER ALEF WITH MADDA ABOVE FINAL FORM
5904	<aH.>	<UFE84>	ARABIC LETTER ALEF WITH HAMZA ABOVE FINAL FORM
5905	<aH.>	<UFE88>	ARABIC LETTER ALEF WITH HAMZA BELOW FINAL FORM
5906	<a+>	<UFE8D>	ARABIC LETTER ALEF ISOLATED FORM
5907	<a+.>	<UFE8E>	ARABIC LETTER ALEF FINAL FORM
5908	<b+>	<UFE8F>	ARABIC LETTER BEH ISOLATED FORM
5909	<b+.>	<UFE90>	ARABIC LETTER BEH FINAL FORM
5910	<b+,>	<UFE91>	ARABIC LETTER BEH INITIAL FORM
5911	<b+i>	<UFE92>	ARABIC LETTER BEH MEDIAL FORM

5912	<tm->	<UFE93>	ARABIC LETTER TEH MARBUTA ISOLATED FORM
5913	<tm.>	<UFE94>	ARABIC LETTER TEH MARBUTA FINAL FORM
5914	<t+>	<UFE95>	ARABIC LETTER TEH ISOLATED FORM
5915	<t+.>	<UFE96>	ARABIC LETTER TEH FINAL FORM
5916	<t+,>	<UFE97>	ARABIC LETTER TEH INITIAL FORM
5917	<t+; >	<UFE98>	ARABIC LETTER TEH MEDIAL FORM
5918	<tk->	<UFE99>	ARABIC LETTER THEH ISOLATED FORM
5919	<tk.>	<UFE9A>	ARABIC LETTER THEH FINAL FORM
5920	<tk,>	<UFE9B>	ARABIC LETTER THEH INITIAL FORM
5921	<tk; >	<UFE9C>	ARABIC LETTER THEH MEDIAL FORM
5922	<g+>	<UFE9D>	ARABIC LETTER JEEM ISOLATED FORM
5923	<g+.>	<UFE9E>	ARABIC LETTER JEEM FINAL FORM
5924	<g+,>	<UFE9F>	ARABIC LETTER JEEM INITIAL FORM
5925	<g+; >	<UFEA0>	ARABIC LETTER JEEM MEDIAL FORM
5926	<hk->	<UFEA1>	ARABIC LETTER HAH ISOLATED FORM
5927	<hk.>	<UFEA2>	ARABIC LETTER HAH FINAL FORM
5928	<hk,>	<UFEA3>	ARABIC LETTER HAH INITIAL FORM
5929	<hk; >	<UFEA4>	ARABIC LETTER HAH MEDIAL FORM
5930	<x+>	<UFEA5>	ARABIC LETTER KHAH ISOLATED FORM
5931	<x+.>	<UFEA6>	ARABIC LETTER KHAH FINAL FORM
5932	<x+,>	<UFEA7>	ARABIC LETTER KHAH INITIAL FORM
5933	<x+; >	<UFEA8>	ARABIC LETTER KHAH MEDIAL FORM
5934	<d+>	<UFEA9>	ARABIC LETTER DAL ISOLATED FORM
5935	<d+.>	<UFEAA>	ARABIC LETTER DAL FINAL FORM
5936	<dk->	<UFEAB>	ARABIC LETTER THAL ISOLATED FORM
5937	<dk.>	<UFEAC>	ARABIC LETTER THAL FINAL FORM
5938	<r+>	<UFEAD>	ARABIC LETTER REH ISOLATED FORM
5939	<r+.>	<UFEAE>	ARABIC LETTER REH FINAL FORM
5940	<z+>	<UFEAF>	ARABIC LETTER ZAIN ISOLATED FORM
5941	<z+.>	<UFEB0>	ARABIC LETTER ZAIN FINAL FORM
5942	<s+>	<UFEB1>	ARABIC LETTER SEEN ISOLATED FORM
5943	<s+.>	<UFEB2>	ARABIC LETTER SEEN FINAL FORM
5944	<s+,>	<UFEB3>	ARABIC LETTER SEEN INITIAL FORM
5945	<s+; >	<UFEB4>	ARABIC LETTER SEEN MEDIAL FORM
5946	<sn->	<UFEB5>	ARABIC LETTER SHEEN ISOLATED FORM
5947	<sn.>	<UFEB6>	ARABIC LETTER SHEEN FINAL FORM
5948	<sn,>	<UFEB7>	ARABIC LETTER SHEEN INITIAL FORM
5949	<sn; >	<UFEB8>	ARABIC LETTER SHEEN MEDIAL FORM
5950	<c+>	<UFEB9>	ARABIC LETTER SAD ISOLATED FORM
5951	<c+.>	<UFEBA>	ARABIC LETTER SAD FINAL FORM
5952	<c+,>	<UFEBB>	ARABIC LETTER SAD INITIAL FORM
5953	<c+; >	<UFEBC>	ARABIC LETTER SAD MEDIAL FORM
5954	<dd->	<UFEBD>	ARABIC LETTER DAD ISOLATED FORM
5955	<dd.>	<UFEBE>	ARABIC LETTER DAD FINAL FORM
5956	<dd,>	<UFEBF>	ARABIC LETTER DAD INITIAL FORM
5957	<dd; >	<UFEC0>	ARABIC LETTER DAD MEDIAL FORM
5958	<tj->	<UFE C1>	ARABIC LETTER TAH ISOLATED FORM
5959	<tj.>	<UFE C2>	ARABIC LETTER TAH FINAL FORM
5960	<tj,>	<UFE C3>	ARABIC LETTER TAH INITIAL FORM
5961	<tj; >	<UFE C4>	ARABIC LETTER TAH MEDIAL FORM
5962	<zH->	<UFE C5>	ARABIC LETTER ZAH ISOLATED FORM
5963	<zH.>	<UFE C6>	ARABIC LETTER ZAH FINAL FORM
5964	<zH,>	<UFE C7>	ARABIC LETTER ZAH INITIAL FORM
5965	<zH; >	<UFE C8>	ARABIC LETTER ZAH MEDIAL FORM
5966	<e+>	<UFE C9>	ARABIC LETTER AIN ISOLATED FORM
5967	<e+.>	<UFE CA>	ARABIC LETTER AIN FINAL FORM
5968	<e+,>	<UFE CB>	ARABIC LETTER AIN INITIAL FORM
5969	<e+; >	<UFE CC>	ARABIC LETTER AIN MEDIAL FORM
5970	<i+>	<UFE CD>	ARABIC LETTER GHAIN ISOLATED FORM
5971	<i+.>	<UFE CE>	ARABIC LETTER GHAIN FINAL FORM
5972	<i+,>	<UFE CF>	ARABIC LETTER GHAIN INITIAL FORM
5973	<i+; >	<UFE D0>	ARABIC LETTER GHAIN MEDIAL FORM
5974	<f+>	<UFE D1>	ARABIC LETTER FEH ISOLATED FORM
5975	<f+.>	<UFE D2>	ARABIC LETTER FEH FINAL FORM
5976	<f+,>	<UFE D3>	ARABIC LETTER FEH INITIAL FORM
5977	<f+; >	<UFE D4>	ARABIC LETTER FEH MEDIAL FORM
5978	<q+>	<UFE D5>	ARABIC LETTER QAF ISOLATED FORM
5979	<q+.>	<UFE D6>	ARABIC LETTER QAF FINAL FORM
5980	<q+,>	<UFE D7>	ARABIC LETTER QAF INITIAL FORM
5981	<q+; >	<UFE D8>	ARABIC LETTER QAF MEDIAL FORM
5982	<k+>	<UFE D9>	ARABIC LETTER KAF ISOLATED FORM
5983	<k+.>	<UFE DA>	ARABIC LETTER KAF FINAL FORM
5984	<k+,>	<UFE DB>	ARABIC LETTER KAF INITIAL FORM
5985	<k+; >	<UFE DC>	ARABIC LETTER KAF MEDIAL FORM
5986	<l+>	<UFE DD>	ARABIC LETTER LAM ISOLATED FORM
5987	<l+.>	<UFE DE>	ARABIC LETTER LAM FINAL FORM
5988	<l+,>	<UFE DF>	ARABIC LETTER LAM INITIAL FORM
5989	<l+; >	<UFE E0>	ARABIC LETTER LAM MEDIAL FORM
5990	<m+>	<UFE E1>	ARABIC LETTER MEEM ISOLATED FORM
5991	<m+.>	<UFE E2>	ARABIC LETTER MEEM FINAL FORM
5992	<m+,>	<UFE E3>	ARABIC LETTER MEEM INITIAL FORM
5993	<m+; >	<UFE E4>	ARABIC LETTER MEEM MEDIAL FORM
5994	<n+>	<UFE E5>	ARABIC LETTER NOON ISOLATED FORM
5995	<n+.>	<UFE E6>	ARABIC LETTER NOON FINAL FORM
5996	<n+,>	<UFE E7>	ARABIC LETTER NOON INITIAL FORM
5997	<n+; >	<UFE E8>	ARABIC LETTER NOON MEDIAL FORM
5998	<h+>	<UFE E9>	ARABIC LETTER HEH ISOLATED FORM
5999	<h+.>	<UFE EA>	ARABIC LETTER HEH FINAL FORM
6000	<h+,>	<UFE EB>	ARABIC LETTER HEH INITIAL FORM

6001	<h+i>	<UFEEC>	ARABIC LETTER HEH MEDIAL FORM
6002	<w+>	<UFEED>	ARABIC LETTER WAW ISOLATED FORM
6003	<w+.>	<UFEEE>	ARABIC LETTER WAW FINAL FORM
6004	<j+>	<UFEFF>	ARABIC LETTER ALEF MAKSURA ISOLATED FORM
6005	<j+.>	<UFEF0>	ARABIC LETTER ALEF MAKSURA FINAL FORM
6006	<y+>	<UFEF1>	ARABIC LETTER YEH ISOLATED FORM
6007	<y+.>	<UFEF2>	ARABIC LETTER YEH FINAL FORM
6008	<y+,>	<UFEF3>	ARABIC LETTER YEH INITIAL FORM
6009	<y+i>	<UFEF4>	ARABIC LETTER YEH MEDIAL FORM
6010	<lm->	<UFEF5>	ARABIC LIGATURE LAM WITH ALEF WITH MADDA ABOVE ISOLATED FORM
6011	<lm.>	<UFEF6>	ARABIC LIGATURE LAM WITH ALEF WITH MADDA ABOVE FINAL FORM
6012	<lh->	<UFEF7>	ARABIC LIGATURE LAM WITH ALEF WITH HAMZA ABOVE ISOLATED FORM
6013	<lh.>	<UFEF8>	ARABIC LIGATURE LAM WITH ALEF WITH HAMZA ABOVE FINAL FORM
6014	<lh->	<UFEF9>	ARABIC LIGATURE LAM WITH ALEF WITH HAMZA BELOW ISOLATED FORM
6015	<lh.>	<UFEFA>	ARABIC LIGATURE LAM WITH ALEF WITH HAMZA BELOW FINAL FORM
6016	<la->	<UFEFB>	ARABIC LIGATURE LAM WITH ALEF ISOLATED FORM
6017	<la.>	<UFEFC>	ARABIC LIGATURE LAM WITH ALEF FINAL FORM
6018	<H->	<U0023>	NUMBER SIGN
6019	<!S>	<U0024>	DOLLAR SIGN
6020	<@>	<U0040>	COMMERCIAL AT
6021	<Oa>	<U0040>	COMMERCIAL AT
6022	<!C>	<U00A2>	CENT SIGN
6023	<L->	<U00A3>	POUND SIGN
6024	<Xo>	<U00A4>	CURRENCY SIGN
6025	<Y->	<U00A5>	YEN SIGN
6026	<!B>	<U00A6>	BROKEN BAR
6027	<So>	<U00A7>	SECTION SIGN
6028	<7!>	<U00AC>	NOT SIGN
6029	<9I>	<U00B6>	PILCROW SIGN
6030	<_>	<U2500>	BOX DRAWINGS LIGHT HORIZONTAL
6031	<=>	<U2501>	BOX DRAWINGS HEAVY HORIZONTAL
6032	<_ >	<U2502>	BOX DRAWINGS LIGHT VERTICAL
6033	<_v/>>	<U250C>	BOX DRAWINGS LIGHT DOWN AND RIGHT
6034	<_v<w>	<U2510>	BOX DRAWINGS LIGHT DOWN AND LEFT
6035	<_A/>>	<U2514>	BOX DRAWINGS LIGHT UP AND RIGHT
6036	<_A<>	<U2518>	BOX DRAWINGS LIGHT UP AND LEFT
6037	<_!/>>	<U251C>	BOX DRAWINGS LIGHT VERTICAL AND RIGHT
6038	<_!<>	<U2524>	BOX DRAWINGS LIGHT VERTICAL AND LEFT
6039	<_v->	<U252C>	BOX DRAWINGS LIGHT DOWN AND HORIZONTAL
6040	<_A->	<U2534>	BOX DRAWINGS LIGHT UP AND HORIZONTAL
6041	<_!->	<U253C>	BOX DRAWINGS LIGHT VERTICAL AND HORIZONTAL
6042	<_//>//>	<U2571>	BOX DRAWINGS LIGHT DIAGONAL UPPER RIGHT TO LOWER LEFT
6043	<_<>	<U2572>	BOX DRAWINGS LIGHT DIAGONAL UPPER LEFT TO LOWER RIGHT
6044	<_./>//>	<U25E2>	BLACK LOWER RIGHT TRIANGLE
6045	<_.<>	<U25E3>	BLACK LOWER LEFT TRIANGLE
6046	<_d!>	<U266A>	EIGHTH NOTE
6047			
6048			

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.

6072
6073
6074
6075
6076
6077
6078
6079
6080
6081
6082
6083
6084
6085
6086
6087
6088
6089
6090
6091
6092
6093
6094
6095
6096
6097
6098
6099
6100
6101
6102
6103
6104
6105
6106
6107
6108
6109
6110
6111
6112
6113
6114
6115
6116
6117
6118
6119
6120
6121
6122

Annex A (informative)

Differences from the ISO/IEC 9945-2 standard

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 introduced to reorder sections.

10. Symbolic elipses (both decimal and hexadecimal) has been introduced as a notation.

- 6123 11. The "print" CTYPE class includes automatically all "graph" characters.
6124
- 6125 12. The <Uxxxx> and <Uxxxxxxxx> notations have been introduced as predefined
6126 symbolic character names, together with a number of symbolic character names derived
6127 from POSIX and the Internet.
6128
- 6129 13. New categories LC_IDENTIFICATION, LC_PAPER, LC_NAME, LC_ADDRESS,
6130 and LC_TELEPHONE, have been introduced.
6131
- 6132 14. The LC_CTYPE has got support for new classes, via the new keywords class and
6133 map, which corresponds to the C standard library functions iswctype() and towctrans()
6134 respectively.
6135
- 6136 15. The "digit" keyword now supports digits for multiple scripts.
6137
- 6138 16. The LC_MONETARY category provides support for multiple currencies, such as the
6139 native currency and the Euro in some European countries.
6140
- 6141 17. The LC_TIME has got a number of enhancements to cater for alternate calendars, and
6142 timezone information may be given.
6143
- 6144 18. The charmap specification has been enhanced to support ISO 2022.

6145
6146
6147
6148
6149
6150
6151
6152
6153
6154
6155
6156
6157
6158
6159
6160
6161
6162
6163
6164
6165
6166
6167
6168
6169
6170
6171
6172
6173
6174
6175
6176
6177
6178
6179
6180
6181
6182
6183
6184
6185
6186
6187
6188
6189
6190
6191
6192
6193
6194
6195

Annex B (informative)

Rationale

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

6196 The LC_CTYPE category primarily is used to define the encoding-independent aspects of
6197 a character set, such as character classification. In addition, certain encoding-dependent
6198 characteristics are also defined for an application via the LC_CTYPE category. This
6199 Technical Report does not mandate that the encoding used in the FDCC-set is the same as
6200 the one used by the application, because an application may decide that it is advantageous
6201 to define a FDCC-set in a system-wide encoding rather than having multiple, logically
6202 identical FDCC-sets in different encodings, and to convert from the application encoding
6203 to the system-wide encoding on usage. Other applications could require encoding-depen-
6204 dent FDCC-sets. In either case, the LC_CTYPE attributes that are directly dependent on
6205 the encoding, such as "mb_cur_max" and the display width of characters, are not user-
6206 specifiable in a locale source, and are consequently not defined as keywords.
6207

6208 As the LC_CTYPE character classes are based on the C Standard character-class
6209 definition, the category does not support multicharacter elements. For instance, the
6210 German character <sharp-s> is traditionally classified as a lowercase letter. There is no
6211 corresponding uppercase letter; in proper capitalization of German text the <sharp-s> will
6212 be replaced by SS; i.e., by two characters. This kind of conversion is outside the scope of
6213 the "toupper" and "tolower" keywords.
6214

6215 The character classes "digit", "xdigit", "lower", "upper", and "space" have a set of
6216 automatically included characters. These only need to be specified if the character values
6217 (i.e. encoding) differs from the application default values. The definition of character class
6218 "digit" allows alternate digits (e.g., Hindi) to be specified here. The definition of character
6219 class "xdigit" requires that the characters included in character class "digit" are included
6220 here also, and allows for different symbols for the hexadecimal digits 10 through 15.
6221

6222 The "combining" and "combining-level3" classes are an IT-enablement of ISO/IEC 10646
6223 definitions of combining characters. These can be used to check identifiers for consistence
6224 with the guidelines given in TR 10176 annex A.
6225

6227 **B.1.3 LC_COLLATE Rationale.**

6228
6229 The LC_COLLATE category governs the collation order in the FDCC-set, and may thus
6230 be useful for the processing of the ISO/IEC 14651 string ordering and comparison
6231 standard, the C Standard strxfrm() and strcoll() functions, as well as a number of ISO/IEC
6232 9945-2:1993 POSIX utilities.
6233

6234 The rules governing collation depends to some extent on the use. At least five different
6235 levels of increasingly complex collation rules can be distinguished:
6236

- 6237 (1) Byte/machine code order. This is the historical collation order in the UNIX
6238 system and many proprietary operating systems. Collation is here done
6239 character by character, without any regard to context. The primary virtue is that
6240 it usually is quite fast, and also completely deterministic; it works well when
6241 the native machine collation sequence matches the user expectations.
- 6242 (2) Character order. On this level, collation is also done character by character,
6243 without regard to context. The order between characters is, however, not deter-
6244 mined by the code values, but on the user's expectations of the correct order
6245 between characters. In addition, such a (simple) collation order can specify that

- 6246 certain characters collate equal (e.g., upper and lowercase letters).
- 6247 (3) String ordering. On this level, entire strings are compared based on relatively
- 6248 straightforward rules. At this level, several "passes" may be required to deter-
- 6249 mine the order between two strings. Characters may be ignored in some passes,
- 6250 but not in others; the strings may be compared in different directions; and
- 6251 simple string substitutions may be made before strings are compared. This level
- 6252 is best described as "dictionary" ordering; it is based on the spelling, not the
- 6253 pronunciation, or meaning, of the words.
- 6254 (4) Text search ordering. This is a further refinement of the previous level, best de-
- 6255 scribed as "telephone book ordering"; some common homonyms (words spelled
- 6256 differently but with same pronunciation) are collated together; numbers are
- 6257 collated as if spelled with words, and so on.
- 6258 (5) Semantic level ordering. Words and strings are collated based on their meaning;
- 6259 entire words (such as "the") are eliminated, the ordering is not deterministic.
- 6260 This may requires special software, and is highly dependent on the intended
- 6261 use.
- 6262

6263 While the historical collation order formally is at level 1, for the English language it

6264 corresponds roughly to elements at level 2. The user expects to see the output from the

6265 "ls" utility sorted very much as it would be in a dictionary. While telephone book ordering

6266 would be an optimal goal for standard collation, this was ruled out as the order would be

6267 language dependent. Furthermore, a requirement was that the order must be determined

6268 solely from the text string and the collation rules; no external information (e.g., "pronu-

6269 nciation dictionaries") could be required.

6270

6271 As a result, the goal for the collation support is at level 3. This also matches the re-

6272 quirements for the Canadian collation order standard, as well as other, known collation

6273 requirements for alphabetic scripts. It specifically rules out collation based on pronun-

6274 ciation rules, or based on semantic analysis of the text. The syntax for the LC_COLLATE

6275 category source is the result of a cooperative effort between representatives for many

6276 countries and organizations working with international issues, such as UniForum, X/Open,

6277 and ISO, and it meets the requirements for level 3, and has been verified to produce the

6278 correct result with examples based on Canadian and Danish collation order.

6279

6280 The directives that can be specified in an operand to the order_start keyword are based on

6281 the requirements specified in several proposed standards and in customary use. The

6282 following is a rephrasing of rules defined for "lexical ordering in English and French" by

6283 the Canadian Standards Association (text in brackets is rephrased):

6284

- 6285 (1) Once special characters (punctuation) have been removed from original strings,
- 6286 the ordering is determined by scanning forward (left to right) [disregarding case
- 6287 and diacriticals].
- 6288 (2) In case of equivalence, special characters are once again removed from original
- 6289 strings and the ordering is determined scanning backward (starting from the
- 6290 rightmost character of the string and back), character by character, (disregarding
- 6291 case but considering diacriticals).
- 6292 (3) In case of repeated equivalence, special characters are removed again from
- 6293 original strings and the ordering is determined scanning forward, character by
- 6294 character, (considering both case and diacriticals).
- 6295 (4) If there is still an ordering equivalence after rules (1) through (3) have been
- 6296 applied, then only special characters and the position they occupy in the string

6297 are considered to determine ordering. The string that has a special character in
6298 the lowest position comes first. If two strings have a special character in the
6299 same position, the character [with the lowest collation value] comes first. In
6300 case of equality, the other special characters are considered until there is a
6301 difference or all special characters have been exhausted.

6302
6303 It is estimated that the Technical Report covers the requirements for all European
6304 languages, and no particular problems are anticipated for Cyrillic or Middle Eastern
6305 scripts.

6306
6307 The Far East (particularly Japanese/Chinese) collations are often based on contextual
6308 information. In Japan, collations of strings containing CJK characters (ideograms) are
6309 often done considering some related information such as pronunciation, which needs a
6310 bulk dictionary (and some common sense). Such collation, in general, falls outside the
6311 desired goal of this Technical Report, and this Technical Report can support only a
6312 restricted of collations used in Japan. There are, however, several other collation rules
6313 (stroke/radical, or "most common pronunciation") which can be supported with the
6314 mechanism described here. Previous drafts contained a substitute statement, which
6315 performed a regular expression style replacement before string compares. It has been
6316 withdrawn based on balloter objections that it was not required for the types of ordering
6317 this Technical Report is aimed at.

6318
6319 The character (and collating element) order is defined by the order in which characters and
6320 elements are specified between the `order_start` and `order_end` keywords. This character
6321 order is used in range expressions in regular expressions. Weights assigned to the charac-
6322 ters and elements define the collation sequence; in the absence of weights, the character
6323 order is also the collation sequence.

6324
6325 The position keyword was introduced to provide the capability to consider, in a compare,
6326 the relative position of non-IGNORED characters. As an example, consider the two strings
6327 "o-ring" and "or-ing". Assuming the hyphen is IGNORED on the first pass, the two strings
6328 will compare equal, and the position of the hyphen is immaterial. On second pass, all
6329 characters except the hyphen are IGNORED, and in the normal case the two strings would
6330 again compare equal. By taking position into account, the first collates before the second.

6331 6332 **B.1.3.1 "reorder-after" rationale**

6333
6334 Much work has been done on FDCC-sets, making them quite general. The ISO/IEC 9945-
6335 2:1993 POSIX standard introduced a "copy" command for all categories of the POSIX
6336 locale. This is useful for many purposes and it ensures that two FDCC-sets are equivalent
6337 for this category. A further step in building on previous FDCC-set work is defined in this
6338 Technical Report.

6339
6340 Collating sequences often vary a bit from country to country, and from language to
6341 language, but generally much of the collating sequence is the same. For example the
6342 Danish sequence is for the most part the same as the German or English collation, but for
6343 about a dozen letters it differs. The same can be said for Swedish or Hungarian: generally
6344 the Latin collating sequence is the same, but a few characters are different.

6345
6346 This Technical Report defines a FDCC-set defined on the character repertoire of the

6347 ISO/IEC 10646 standard, in a character set independent way. The intention is that some of
6348 the information from this FDCC-set will be acceptable in many cultures, and that it can
6349 serve as the basis for modifications in other cultures, to obtain a culturally acceptable
6350 specification. Using the "reorder-after" construct will also help improve the overview of
6351 what the changes really are for implementers and other users.

6352
6353 An example of the use of the "reorder-after" construct is the following. A default
6354 international ordering for the Latin alphabet may be adequate for Danish, with the
6355 exception of the collation rules for the letters Û, ü, Æ, æ, Ä, ä, Ø, ø, Ö, ö, Å and å. By
6356 applying the "reorder-after" construct, the Danish specification can be made more easily
6357 by copying and reordering the existing international specification, rather than specifying
6358 collation parameters for all Latin letters (with or without diacritics). There is no obligation
6359 for Denmark to take this approach, but the "reorder-after" construct provides the
6360 mechanism for doing so if it is deemed desirable.

6361
 6362
 6363
 6364
 6365
 6366
 6367
 6368
 6369
 6370
 6371
 6372
 6373
 6374
 6375
 6376
 6377
 6378
 6379
 6380
 6381
 6382
 6383
 6384
 6385
 6386
 6387
 6388
 6389
 6390
 6391
 6392
 6393
 6394
 6395
 6396
 6397
 6398
 6399
 6400
 6401
 6402
 6403
 6404
 6405
 6406
 6407
 6408
 6409
 6410
 6411
 6412
 6413

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

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
    }
  }
}

```

6414 **B.1.3.3 Sample FDCC-set specification for Danish**

```

6415
6416 escape_char /
6417 comment_char %
6418 repertoiremap "i18nrep"
6419 charset "ISO_8859-1:1987"
6420 % Distribution and use is free, also
6421 % for commercial purposes.
6422
6423 LC_VERSION
6424 title "Danish language FDCC-set for Denmark"
6425 source "Danish Standards Association"
6426 address "Kollegievej 6, DK-2920 Charlottenlund, Danmark"
6427 contact "Keld Simonsen"
6428 email "Keld.Simonsen@dkuug.dk"
6429 tel "+45 - 3996-6101"
6430 fax "+45 - 3996-6202"
6431 language "da"
6432 territory "DK"
6433 revision "4.2"
6434 date "1997-12-22"
6435
6436 category i18n:1998;LC_IDENTIFICATION
6437 category i18n:1998;LC_CTYPE
6438 category i18n:1998;LC_COLLATE
6439 category i18n:1998;LC_TIME
6440 category posix:1993;LC_NUMERIC
6441 category i18n:1998;LC_MONETARY
6442 category posix:1993;LC_MESSAGES
6443 category i18n:1998;LC_PAPER
6444 category i18n:1998;LC_NAME
6445 category i18n:1998;LC_ADDRESS
6446 category i18n:1998;LC_TELEPHONE
6447
6448 END LC_VERSION
6449
6450 LC_CTYPE
6451 copy "i18n"
6452 END LC_CTYPE
6453
6454 LC_COLLATE
6455 % The ordering algorithm is in accordance
6456 % with Danish Standard DS 377 (1980)
6457 % and the Danish Orthography Dictionary
6458 % (Retskrivningsordbogen, 2. udgave, 1996).
6459 % It is also in accordance with
6460 % Greenlandic orthography.
6461
6462 collating-element <A-A> from "<A><A>"
6463 collating-element <A-a> from "<A><a>"
6464 collating-element <a-A> from "<a><A>"
6465 collating-element <a-a> from "<a><a>"
6466 copy i18n
6467 reorder-after <CAPITAL>
6468 <CAPITAL>
6469 <CAPITAL-SMALL>
6470 <SMALL-CAPITAL>
6471 <SMALL>
6472 reorder-after <q8>
6473 <kk> <Q>;<SPECIAL>;<SMALL>;IGNORE
6474 reorder-after <t8>
6475 <TH> "<T><H>";"<TH><TH>";"<CAPITAL><CAPITAL>";IGNORE
6476 <th> "<T><H>";"<TH><TH>";"<SMALL><SMALL>";IGNORE
6477 reorder-after <y8>
6478 % <U:> and <U"> are treated as <Y> in Danish
6479 <U:> <Y>;<U:>;<CAPITAL>;IGNORE
6480 <u:> <Y>;<U:>;<SMALL>;IGNORE
6481 <U"> <Y>;<U">;<CAPITAL>;IGNORE
6482 <u"> <Y>;<U">;<SMALL>;IGNORE
6483 reorder-after <z8>
6484 % <AE> is a separate letter in Danish

```

```

6485 <AE> <AE>;<NONE>;<CAPITAL>;IGNORE
6486 <ae> <AE>;<NONE>;<SMALL>;IGNORE
6487 <AE'> <AE>;<ACUTE>;<CAPITAL>;IGNORE
6488 <ae'> <AE>;<ACUTE>;<SMALL>;IGNORE
6489 <A3> <AE>;<MACRON>;<CAPITAL>;IGNORE
6490 <a3> <AE>;<MACRON>;<SMALL>;IGNORE
6491 <A:> <AE>;<SPECIAL>;<CAPITAL>;IGNORE
6492 <a:> <AE>;<SPECIAL>;<SMALL>;IGNORE
6493 % <O//> is a separate letter in Danish
6494 <O//> <O//>;<NONE>;<CAPITAL>;IGNORE
6495 <o//> <O//>;<NONE>;<SMALL>;IGNORE
6496 <O//'> <O//>;<ACUTE>;<CAPITAL>;IGNORE
6497 <o//'> <O//>;<ACUTE>;<SMALL>;IGNORE
6498 <O:> <O//>;<DIAERESIS>;<CAPITAL>;IGNORE
6499 <o:> <O//>;<DIAERESIS>;<SMALL>;IGNORE
6500 <O"> <O//>;<DOUBLE-ACUTE>;<CAPITAL>;IGNORE
6501 <o"> <O//>;<DOUBLE-ACUTE>;<SMALL>;IGNORE
6502 % <AA> is a separate letter in Danish
6503 <AA> <AA>;<NONE>;<CAPITAL>;IGNORE
6504 <aa> <AA>;<NONE>;<SMALL>;IGNORE
6505 <A-A> <AA>;<A-A>;<CAPITAL>;IGNORE
6506 <A-a> <AA>;<A-A>;<CAPITAL-SMALL>;IGNORE
6507 <a-A> <AA>;<A-A>;<SMALL-CAPITAL>;IGNORE
6508 <a-a> <AA>;<A-A>;<SMALL>;IGNORE
6509 <AA'> <AA>;<AA'>;<CAPITAL>;IGNORE
6510 <aa'> <AA>;<AA'>;<SMALL>;IGNORE
6511 reorder-end
6512 END LC_COLLATE
6513
6514 LC_MONETARY
6515 int_curr_symbol " <D><K><K><SP>"
6516 currency_symbol "<k><r>"
6517 mon_decimal_point "<,>"
6518 mon_thousands_sep "<.>"
6519 mon_grouping 3;3
6520 positive_sign ""
6521 negative_sign "<->"
6522 int_frac_digits 2
6523 frac_digits 2
6524 p_cs_precedes 1
6525 p_sep_by_space 2
6526 n_cs_precedes 1
6527 n_sep_by_space 2
6528 p_sign_posn 4
6529 n_sign_posn 4
6530 END LC_MONETARY
6531
6532 LC_NUMERIC
6533 decimal_point "<,>"
6534 thousands_sep "<.>"
6535 grouping 3;3
6536 END LC_NUMERIC
6537
6538 LC_TIME
6539 abday " <m><a><n>" ; /
6540 " <t><i><r>" ; " <o><n><s>" ; /
6541 " <t><o><r>" ; " <f><r><e>" ; /
6542 " <l><o//><r>" ; " <s><o//><n>"
6543 day " <m><a><n><d><a><g>" ; /
6544 " <t><i><r><s><d><a><g>" ; /
6545 " <o><n><s><d><a><g>" ; /
6546 " <t><o><r><s><d><a><g>" ; /
6547 " <f><r><e><d><a><g>" ; /
6548 " <l><o//><r><d><a><g>" /
6549 " <s><o//><n><d><a><g>" ;
6550 week 7;19971201;4
6551 abmon " <j><a><n>" ; " <f><e><b>" ; /
6552 " <m><a><r>" ; " <a><p><r>" ; /
6553 " <m><a><j>" ; " <j><u><n>" ; /
6554 " <j><u><l>" ; " <a><u><g>" ; /
6555 " <s><e><p>" ; " <o><k><t>" ; /

```

```

6556          " <n><o><v>" ; " <d><e><c>"
6557 mon       " <j><a><n><u><a><r>" ; /
6558          " <f><e><b><r><u><a><r>" ; /
6559          " <m><a><r><t><s>" ; /
6560          " <a><p><r><i><l>" ; /
6561          " <m><a><j>" ; /
6562          " <j><u><n><i>" ; /
6563          " <j><u><l><i>" ; /
6564          " <a><u><g><u><s><t>" ; /
6565          " <s><e><p><t><e><m><b><e><r>" ; /
6566          " <o><k><t><o><b><e><r>" ; /
6567          " <n><o><v><e><m><b><e><r>" ; /
6568          " <d><e><c><e><m><b><e><r>"
6569 d_t_fmt     " <%><a><SP><%><F><SP><%><T><SP><%><Z>"
6570 d_fmt       " <%><O><d><.><SP><%><B><SP><%><Y>"
6571 atl_digits  " <0><.>; <1><.>; <2><.>; <3><.>; <4><.>; /
6572           <5><.>; <6><.>; <7><.>; <8><.>; <9><.>; /
6573           <1><0><.>; <1><1><.>; <1><2><.>; <1><3><.>; <1><4><.>; /
6574           <1><5><.>; <1><6><.>; <1><7><.>; <1><8><.>; <1><9><.>; /
6575           <2><0><.>; <2><1><.>; <2><2><.>; <2><3><.>; <2><4><.>; /
6576           <2><5><.>; <2><6><.>; <2><7><.>; <2><8><.>; <2><9><.>; /
6577           <3><0><.>; <3><1><.>"
6578 t_fmt       " <%><T>"
6579 am_pm       " ; "
6580 t_fmt_ampm  " "
6581 timezone    " <C><E><T><-><1><C><E><T><SP><D><S><T>< , ><M><3><.><5><.><0>/
6582           < , ><M><1><0><.><5><.><0>"
6583 END LC_TIME
6584
6585 LC_MESSAGES
6586 yesexpr     " << (><1><J><j><Y><y>< ) />><.>< * >"
6587 noexpr      " << (><0><N><n>< ) />><.>< * >"
6588 END LC_MESSAGES
6589
6590 LC_PAPER
6591 copy "i18n"
6592 END LC_PAPER
6593
6594 LC_NAME
6595 name_fmt    " <%><p><%><t><%><g><%><t><%><m><%><t><%><f>"
6596 name_gen    ""
6597 name_mr     " <h><r>"
6598 name_mrs    " <f><r><u>"
6599 name_miss   " <f><r><o></><k><e><n>"
6600 name_ms     " <f><r>"
6601 END LC_NAME
6602
6603 LC_ADDRESS
6604 country_name " <D><a><n><m><a><r><k>"
6605 country_post " <D><K>"
6606 country_ab2  " <D><K>"
6607 country_ab3  " <D><N><K>"
6608 country_num  208
6609 country_car  " <D><K>"
6610 country_isbn " <8><7>"
6611 lang_ab     " <d><a>"
6612 lang_term    " <d><a><n>"
6613 postal_fmt   " <%><a><%><N><%><f><%><N><%><d><%><N><%><b><%><N><%>/
6614           <%><s><SP><%><h><SP><%><e><SP><%><r><%><N>/
6615           <%><C><-><%><z><SP><%><T><%><N><%><c><%><N>"
6616 END LC_ADDRESS

```



```

6617 LC_TELEPHONE
6618 tel_int_fmt    "<+><%><c><SP><%><a><SP><%><l>"
6619 tel_dom_fmt    "<%><l>"
6620 int_select     "<0><0>"
6621 int_prefix     "<4><5>"
6622 END LC_TELEPHONE

```

6623
6624
6625

B.1.4 LC_MONETARY Rationale.

6626 The currency symbol does not appear in LC_MONETARY because it is not defined in the
6627 C Standard's C locale. The C Standard limits the size of decimal points and thousands
6628 delimiters to single-byte values. In FDCC-sets based on multibyte coded character sets this
6629 cannot be enforced, obviously; this Technical Report does not prohibit such characters, but
6630 makes the behaviour unspecified (in the text "In contexts where other standards . . .").
6631

6632 The grouping specification is based on, but not identical to, the C Standard . The "-1"
6633 signals that no further grouping shall be performed, the equivalent of (CHAR_MAX) in
6634 the C Standard).
6635

6636 The FDCC-set definition is an extension of the C Standard localeconv() specification. In
6637 particular, rules on how currency_symbol is treated are extended to also cover int_
6638 curr_symbol, and p_set_by_space and n_sep_by_space have been augmented with the
6639 value 2, which places a space between the sign and the symbol (if they are adjacent;
6640 otherwise it should be treated as a 0). The following table shows the result of various
6641 combinations:
6642

		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\$+

6652
6653
6654
6655
6656
6657
6658
6659

6660 The following is an example of the interpretation of the mon_grouping keyword.
6661 Assuming that the value to be formatted is 123456789 and the mon_thousands_sep is "",
6662 then the following table shows the result. The third column shows the equivalent C
6663 Standard string that would be used to accommodate this grouping. It is the responsibility
6664 of the utility to perform mappings of the formats in this clause to those used by language
6665 bindings such as the C Standard .
6666

	Mon_grouping	Formatted Value	C String
6667	3;-1	123456'789	"\3\177"
6668	3	123'456'789	"\3"
6669	3;2;-1	1234'56'789	"\3\2\177"
6670	3;2	12'34'56'789	"\3\2"
6671	-1	123456789	"177"

In these examples, the octal value of (CHAR_MAX) is 177.

The multiple currency support is specified such that a FDCC-set can be used without change during the transition period in a static environment. For example in the case of the Euro currency as being employed in a number of European countries, there is no need to change the FDCC-set when shifting from one currency to two concurrent currencies; and there is no need to change FDCC-set, when changing to the Euro as the only currency. Also the same application call can be made to be valid for countries with a single currency and countries with dual currencies. The specifications can also be used without change of the FDCC-set on an installation, when converting from one national currency to another, for example when removing some zeroes to form a new currency.

The following example illustrates the support for multiple currencies; the example is for the Euro in Germany:

```

LC_MONETARY
valid_from          " ";          "19990101"
valid_to            "20020630";      " "
conversion_rate     1;              195/100
int_curr_symbol     "<D><E><M><SP>";    "<E><U><R><SP>"
currency_symbol     "<D><M>";        "<E><U><R>"
mon_decimal_point   "<,>"
mon_thousands_sep  "<.>"
mon_grouping        3;3
positive_sign       " "
negative_sign       "<->"
int_frac_digits     2;              2
frac_digits         2;              2
p_cs_precedes       1;              1
p_sep_by_space      2;              2
n_cs_precedes       1;              1
n_sep_by_space      2;              2
p_sign_posn         4;              4
n_sign_posn         4;              4

END LC_MONETARY
    
```

B.1.5 LC_NUMERIC Rationale.

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

B.1.6 LC_TIME Rationale.

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

In some FDCC-sets the field descriptors for weekday and month names will be given with an initial small letter. Programs using these fields may need to adjust the capitalization if the output is going to be used at the beginning of a sentence.

6726 The field descriptors corresponding to the optional keywords consist of a modifier
6727 followed by a traditional field descriptor (for instance %Ex). If the optional keywords are
6728 not supported by the application or are unspecified for the current FDCC-set, these field
6729 descriptors shall be treated as the traditional field descriptor. For instance, assume the
6730 following keywords:

```
6731         alt_digits "0th";"1st";"2nd";"3rd";"4th";"5th";"6th";"7th";"8th";"9th";"10th"  
6732         d_fmt "The %Od day of %B in %Y"
```

6734
6735 On 7/4/1776, the %x field descriptor would result in "The 4th day of July in 1776," while
6736 7/14/1789 would come out as "The 14 day of July in 1789." It can be noted that the above
6737 example is for illustrative purposes only; the %o modifier is primarily intended to provide
6738 for Kanji or Hindi digits in date formats. While it is clear that an alternate year format is
6739 required, there is no consensus on the format or the requirements. As a result, while these
6740 keywords are reserved, the details are left unspecified. It is expected that National
6741 Standards Bodies will provide specifications.

6742

6743

6744 **B.1.7 LC_MESSAGES Rationale.**

6745

6746 The LC_MESSAGES category is described in clause 4 as affecting the language used by
6747 utilities for their output. The mechanism used by the application to accomplish this, other
6748 than the responses shown here in the FDCC-set definition, is not specified by this version
6749 of this Technical Report. The internationalization working group is developing an interface
6750 that would allow applications (and, presumably some of the standard utilities) to access
6751 messages from various message catalogs, tailored to a user's LC_MESSAGES value.

6752

6753

6754 **B.1.8 LC_PAPER Rationale.**

6755

6756 The LC_PAPER category gives information to prepare output on a printer. Only the
6757 physical measurements of the height and width is available, as this is the information most
6758 often available in various document handling applications.

6759

6760

6761 **B.1.9 LC_NAME Rationale.**

6762

6763 The LC_NAME category gives information to prepare a text for addressing a person, for
6764 example as a part of a postal address on an envelope, or as a saluting line in a letter.
6765 The information is intended to be given to an API that has the various naming information
6766 as parameters and yields a formatted string as the return value.

6767

6768

6769 **B.1.10 LC_ADDRESS Rationale.**

6770

6771 The LC_ADDRESS category gives information to prepare a text for writing an address,
6772 for example as a part of a postal address on an envelope. The information is intended to
6773 be given to an API that has the various address information as parameters and yields a
6774 formatted string as the return value.

6775

6776
6777
6778
6779
6780
6781
6782
6783
6784
6785
6786
6787
6788
6789
6790
6791
6792
6793
6794
6795
6796
6797
6798
6799
6800
6801
6802
6803
6804
6805
6806
6807
6808
6809
6810
6811
6812
6813
6814
6815
6816
6817
6818
6819
6820
6821
6822
6823
6824
6825
6826

B.1.11 LC_TELEPHONE Rationale.

The LC_TELEPHONE category gives information to prepare a text for writing a telephone number. The information is intended to be given to an API that has the various information on a telephone number as parameters and yields a formatted string as the return value. Both an international and a domestic formatting possibility is available.

B.2 Character Set Rationale.

This Technical Report poses no requirement that multiple character sets or code sets be supported, leaving this as a marketing differentiation for implementors. Although multiple charmaps are supported, it is the responsibility of the application to provide the file(s); if only one is provided, only that one will be accessible.

The character set description text provides the capability to describe character set attributes (such as collation order or character classes) independent of character set encoding, and using only the characters in the portable character set. This makes it possible to create "generic" FDCC-set source texts for all code sets that share the portable character set (such as the ISO/IEC 8859 family or IBM Extended ASCII).

Applications are free to describe more than one code set in a character set description text. For example, if an application defines ISO/IEC 8859-1 as the primary code set, and ISO/IEC 8859-2 as an alternate set, with each character from the alternate code set preceded in data by a shift code, a character set description text could contain a complete description of the primary set and those characters from the secondary that are not identical, the encoding of the latter including the shift code.

Applications are free to choose their own symbolic names, as long as the names identified by this Technical Report are also defined; this provides support for already existing "character names".

The charmap was introduced to resolve problems with the portability of, especially, FDCC-set sources. While the portable character set (in Table 1) is a constant across all FDCC-sets for a particular application, this is not true for the extended character set. However, the particular coded character set used for an application does not necessarily imply different characteristics or collation: on the contrary, these attributes should in many cases be identical, regardless of codeset. The charmap provides the capability to define a common FDCC-set definition for multiple codesets (the same FDCC-set source can be used for codesets with different extended characters; the ability in the charmap to define "empty" names allows for characters missing in certain codesets).

In addition, some implementors have expressed an interest in using the charmap to define certain other characteristics of codesets, such as the <mb_cur_max> value for the particular codeset. (Note that <mb_cur_max> has to be equal to or lower than the C Standard {MB_LEN_MAX}, which is the application limit). Such extensions are not described here; but may be added in a later revision of this Technical Report.

The <escape_char> declaration was added at the request of the international community to ease the creation of portable charmaps on terminals not implementing the default backslash escape. (This approach was adopted because this is a new interface invented by

6827 ISO/IEC 9945-2:1993 POSIX. Historical interfaces, such as the shell command language
6828 and awk, have not been modified to accommodate this type of terminal.)

6829
6830 The octal number notation was selected to match those of POSIX "awk" and "tr" utilities
6831 and is consistent with that used by the POSIX localedef utility.

6832
6833 The charmap capability implements a facility available at some X/Open compatible
6834 applications. Its prime virtue is to support "generic" collation sequence source definitions.
6835 An implementor or an applications developer can produce a template definition that can be
6836 used to produce several codeset-dependent "compiled" FDCC-set definitions. The facility
6837 also removes any dependency in many source definitions on characters outside the
6838 character set defined in this clause.

6839
6840 The charmap allows specification of more than one encoding of a character. This allows
6841 for encodings that can encode items in more than one way. For example, an item can be
6842 encoded once as a fully composed character and again as a base character plus combining
6843 character. This would allow either representation to be recognized. As only the first
6844 occurrence of the character may be output, this technique could be used to normalize a
6845 character stream.

6846
6847 The ISO 2022 support introduced gives the possibility to refer other definitions via
6848 charmaps, so the full encoding does not have to be replicated. It supports shifting with G0,
6849 G1, G2 and G3 sets, and also general shifting of coded character sets via escape
6850 sequences.

6851

6852

6853 **B.3 Repertoiremap Rationale.**

6854

6855 The repertoiremap was introduced to make FDCC-sets independent of the availability of
6856 charmaps. With the repertoiremap it is possible to use a FDCC-set encoded with one set of
6857 symbolic character names, together with charmaps with other symbolic character naming
6858 schemes, provided there are repertoiremaps available for both naming schemes.

6859

6860 Repertoiremaps are also useful to describe repertoires of characters, to be used for
6861 example for transliteration.

6862
6863
6864
6865
6866
6867
6868
6869
6870
6871
6872
6873
6874
6875
6876
6877
6878
6879
6880
6881
6882
6883
6884
6885
6886
6887
6888
6889
6890
6891
6892
6893
6894
6895
6896
6897
6898
6899
6900
6901
6902
6903
6904
6905
6906
6907
6908
6909
6910
6911
6912
6913
6914
6915
6916
6917
6918
6919
6920
6921
6922

Annex C
(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 grammar rules are common to all categorie)
char = <any character except those that makes an End
Of Line>
graphic_char = <any character except control_characters and
space> ;
space = ' ' | <TAB> ;
EOL = <anything that makes an End Of Line (EOL) in
the operating system employed> ;
| comment EOL ;
comment_char = <defined by the 'comment_char' keyword> ;
escape_char = <defined by the 'escape_char' keyword> ;
charsymbol = simple_symbol | ucs_symbol ;
collsymbol = simple_symbol ;
collelement = simple_symbol ;
sectionsymbol = simple_symbol ;
octdigit = '0' | '1' | '2' | '3' | '4' | '5' | '6' | '7' ;
digit = '0' | '1' | '2' | '3' | '4' | '5' | '6' | '7' | '8' | '9' ;
hex_upper = 'A' | 'B' | 'C' | 'D' | 'E' | 'F' digit ;
hexdigit = hex_upper | 'a' | 'b' | 'c' | 'd' | 'e' | 'f' ;
letter = 'a' | 'b' | 'c' | 'd' | 'e' | 'f' | 'g' | 'h' | 'i' | 'j' | 'k'
| 'l' | 'm' | 'n' | 'o' | 'p' | 'q' | 'r' | 's' |
| 't' | 'u' | 'v' | 'w' | 'x' | 'y' | 'z' | 'A' | 'B' | 'C' | 'D' | 'E'
| 'F' | 'G' | 'H' | 'I' | 'J' | 'K' | 'L' | 'M' | 'N' | 'O' | 'P' |
| 'Q' | 'R' | 'S' | 'T' | 'U' | 'V' | 'W' | 'X' | 'Y' | 'Z' ;
portable_graph = letter digit | '!' | '"' | '#' | '$' | '%' | '&'
| "'" | '(' | ')' | '*' | '+' | ',' | '-' | '.' | '/' | ':' | ';'
| '<' | '=' | '>' | '?' | '@' | '[' | '\' | ']' | '^' | '_'
| '`' | '{' | '|' | '}' | '~' ;
portable_char = portable_graph | '<NUL>' | '<ALERT>'
| '<BACKSPACE>' | '<TAB>' | '<CARRIAGE_RETURN>'
| '<NEWLINE>' | '<VERTICAL_TAB>' | '<FORM_FEED>' ;
octal_char = escape_char octdigit octdigit octdigit* ;
hex_char = escape_char 'x' hexdigit hexdigit hexdigit* ;
decimal_char = escape_char 'd' digit digit digit* ;
number = digit digit* ;
id_part = letter | digit | '-' | '_' ;
```

```

6923 four_digit_hex_string = hex_upper hex_upper hex_upper hex_upper ;
6924 identifier = letter id_part* ;
6925 simple_symbol = space* '<' portable_graph+ '>' ;
6926 ucs_symbol = space* '<U' four_digit_hex_string
6927 [ four_digit_hex_string ] '>' ;
6928 quoted_string = ''' char_symbol* ''' ;
6929 quoted_nonempty_string = ''' char_symbol [ char_symbol* ] ''' ;
6930 char_symbol = char | charsymbol
6931 | octal_char | hex_char | decimal_char ;
6932 elem_list = elem elem* ;
6933 elem = char_symbol | collsymbol | collelement ;
6934 symb_list = collsymbol+ ;
6935 FDCC_set_name = FDCC-name | ''' FDCC-name ''' ;
6936 copy_FDCC_set = 'copy' FDCC_set_name EOL ;
6937 FDCC-name = portable_graph+ ;
6938 semicolon = ';' ;
6939 comment = comment_char char* ;
6940
6941 % The following is the overall FDCC-set grammar
6942 FDCC_set_definition = [ global_statement* ] category+ ;
6943 global_statement = 'escape_char' character EOL
6944 | 'comment_char' character EOL
6945 | 'repertoiremap' quoted_string EOL
6946 | 'charmap' quoted_string EOL ;
6947 category = lc_identification | lc_ctype | lc_collate
6948 | lc_monetary | lc_numeric | lc_time
6949 | lc_messages | lc_paper | lc_telephone
6950 | lc_name | lc_address ;
6951
6952 % The following is the LC_IDENTIFICATION category grammar
6953 lc_ident = ident_head ident_keyword* ident_tail
6954 | ident_head copy_FDCC_set ident_tail ;
6955 ident_head = 'LC_IDENTIFICATION' EOL ;
6956 ident_keyword = ident_keyword_string quoted_string EOL ;
6957 ident_keyword_string = 'title' | 'source' | 'address' | 'contact'
6958 | 'email' | 'tel' | 'fax' | 'language'
6959 | 'territory' | 'audience' | 'application'
6960 | 'abbreviation' | 'revision' | 'date' ;
6961 ident_tail = 'END' 'LC_IDENTIFICATION' EOL ;
6962
6963
6964 % The following is the LC_CTYPE category grammar
6965 lc_ctype = ctype_head ctype_keyword* [ translit ]
6966 ctype_tail
6967 | ctype_head copy_FDCC_set ctype_tail ;
6968 ctype_head = 'LC_CTYPE' EOL ;
6969 ctype_keyword = charclass_keyword charclass_list EOL
6970 | charconv_keyword charconv_list EOL ;
6971 charclass_keyword = 'upper' | 'lower' | 'alpha' | 'digit'
6972 | 'punct' | 'xdigit' | 'space' | 'print'
6973 | 'graph' | 'blank' | 'cntrl' | 'outdigit'
6974 | 'class' class_name semicolon ;
6975 class_name = "combining" | "combining_level3"
6976 | ''' identifier ''' ;
6977 charclass_list = charclass_list semicolon char_symbol
6978 | charclass_list semicolon ctype_abs_ellipsis
6979 semicolon char_symbol
6980 | charclass_list semicolon charsymbol
6981 ctype_symbolic_ellipses charsymbol
6982 | char_symbol ;
6983 charconv_keyword = 'toupper' | 'tolower'
6984 | 'map' ''' identifier ''' semicolon ;
6985 charconv_list = charconv_list semicolon charconv_entry
6986 | charconv_entry ;
6987 charconv_entry = '(' char_symbol ',' char_symbol ')' ;
6988 ctype_symbolic_ellipses = '...' | '....' | '..(2)..' ;
6989 ctype_abs_ellipses = '...' ;
6990 translit = translit_start [translit_include]
6991 [default_missing] translit_statement*
6992 translit_end ;
6993 translit_start = 'translit_start' EOL ;

```

```

6994 translit_include = 'include' FDCC_set_name semicolon
6995 quoted_nonempty_string EOL ;
6996 default_missing = 'default_missing' quoted_string EOL ;
6997 translit_ignore = 'translit_ignore' charclass_list EOL ;
6998 translit_statement = char_or_string char_or_string [ semicolon
6999 char_or_string ]* EOL ;
7000 translit_end = 'translit_end' EOL ;
7001 ctype_tail = 'END' 'LC_TYPE' EOL ;
7002
7003 % The following is the LC_COLLATE category grammar
7004 lc_collate = collate_head collate_keywords collate_tail ;
7005 collate_head = 'LC_COLLATE' EOL ;
7006 collate_keywords = [ opt_statement* ] order_statements ;
7007 opt_statement = 'collating-symbol' collsymbol* EOL
7008 | 'collating-element' collelement
7009 collelem_string EOL
7010 | 'section-symbol' sectionsymbol EOL
7011 | 'copy' FDCC_set_name EOL
7012 | 'col_weight_max number EOL
7013 | 'symbol-equivalence' collsymbol collsymbol ;
7014 collelem_string = "" char_symbol char_symbol char_symbol* "" ;
7015 order_statements = order_start collation_order order_end ;
7016 order_start = 'order_start' collsymbol [ semicolon
7017 order_opts ] EOL
7018 | 'order_start' [ order_opts ] EOL ;
7019 order_opts = order_opt [ semicolon order_opt ] ;
7020 order_opt = order_opt [ ',' opt_word ] ;
7021 opt_word = 'forward' | 'backward' | 'position' ;
7022 collation_order = collation_statement* ;
7023 collation_statement = collsymbol EOL
7024 | collating_element [ weight_list ] EOL ;
7025 collation_element = char_symbol | collelement
7026 | ellipses | 'UNDEFINED' ;
7027 weight_list = weight_symbol [ semicolon weight_symbol ]* ;
7028 weight_symbol = (* empty *)
7029 | char_symbol
7030 | collsymbol
7031 | "" elem_list ""
7032 | "" symb_list "" | 'IGNORE' ;
7033 ellipses = '...' | '..' | '....' ;
7034 reorder_after = 'reorder-after' collsymbol EOL ;
7035 reorder_end = 'reorder-end' EOL ;
7036 reorder_section_after = 'reorder-section-after' sectionsymbol
7037 sectionsymbol EOL ;
7038 reorder_section_end = 'reorder-section-end' EOL ;
7039 order_end = 'order_end' EOL
7040 collate_tail = 'END' 'LC_COLLATE' EOL ;
7041
7042 % The following is the LC_MESSAGES category grammar
7043 lc_messages = messages_head messages_keyword* messages_tail
7044 | messages_head copy_FDCC_set messages_tail ;
7045 messages_head = 'LC_MESSAGES' EOL ;
7046 messages_keyword = 'yesexpr' "" extended_reg_expr "" EOL
7047 | 'yesexpr' "" extended_reg_expr "" EOL ;
7048 messages_tail = 'END' 'LC_MESSAGES' EOL ;
7049
7050 % The following is the LC_MONETARY category grammar
7051 lc_monetary = monetary_head monetary_keyword* monetary_tail
7052 | monetary_head copy_FDCC_set monetary_tail ;
7053 monetary_head = 'LC_MONETARY' EOL ;
7054 monetary_keyword = mon_keyword_string quoted_string EOL
7055 | mon_keyword_strings mon_string_list EOL
7056 | mon_keyword_char mon_number_list EOL
7057 | mon_keyword_date mon_date_list EOL
7058 | 'conversion_rate' mon_conv_list EOL
7059 | 'mon_grouping' mon_group_list EOL ;
7060 mon_keyword_string = 'mon_decimal_point' | 'mon_thousands_sep'
7061 | 'positive_sign' | 'negative_sign' ;
7062 mon_keyword_strings = 'int_curr_symbol' | 'currency_symbol' ;
7063 mon_keyword_char = 'int_frac_digits' | 'frac_digits'
7064

```



```

7065 | 'p_cs_precedes' | 'p_sep_by_space'
7066 | 'n_cs_precedes' | 'n_sep_by_space'
7067 | 'int_p_cs_precedes' | 'int_p_sep_by_space'
7068 | 'int_n_cs_precedes' | 'int_n_sep_by_space'
7069 | 'p_sign_posn' | 'n_sign_posn'
7070 | 'int_p_sign_posn' | 'int_n_sign_posn' ;
7071 mon_keyword_date = 'valid_from' | 'valid_to' ;
7072 mon_date_list = mon_date | mon_date_list ';' mon_date ;
7073 mon_date = ''' [ '-' ] 8 * digit ''' ;
7074 mon_group_list = number | mon_group_list ';' number ;
7075 mon_string_list = quoted_string [ ';' quoted_string ] * ;
7076 mon_number_list = mon_number | mon_number_list ';' mon_number ;
7077 mon_number = number | -1 ;
7078 mon_conv_list = mon_pair | mon_conv_list ';' mon_pair ;
7079 mon_pair = number '/' number ;
7080 monetary_tail = 'END' 'LC_MONETARY' EOL ;
7081
7082 % The following is the LC_NUMERIC category grammar
7083 lc_numeric = numeric_head numeric_keyword* numeric_tail
7084 | numeric_head copy_FDCC_set numeric_tail ;
7085 numeric_head = 'LC_NUMERIC' EOL ;
7086 numeric_keyword = num_keyword_string quoted_string EOL
7087 | num_keyword_grouping num_group_list EOL ;
7088 num_keyword_string = 'decimal_point' | 'thousands_sep' ;
7089 num_keyword_grouping = 'grouping' ;
7090 num_group_list = number
7091 | num_group_list semicolon number ;
7092 numeric_tail = 'END' 'LC_NUMERIC' EOL ;
7093
7094 % The following is the LC_TIME category grammar
7095 lc_time = time_head time_keyword* time_tail
7096 | time_head copy_FDCC_set time_tail ;
7097 time_head = 'LC_TIME' EOL ;
7098 time_keyword = time_keyword_name time_list EOL
7099 | time_keyword_fmt quoted_string EOL
7100 | time_keyword_opt time_list EOL
7101 | 'week' number semicolon mon_date semicolon
7102 number EOL
7103 | time_keyword_num number EOL
7104 | 'timezone' time_list EOL ;
7105 time_keyword_name = 'abday' | 'day' | 'abmon' | 'mon' | 'am_pm' ;
7106 time_keyword_fmt = 'd_t_fmt' | 'd_fmt' | 't_fmt' | 't_fmt_ampm' ;
7107 time_keyword_opt = 'era' | 'era_year' | 'era_d_fmt' | 'alt_digits'
7108 ;
7109 time_keyword_week = 'week' ;
7110 time_keyword_num = 'first_weekday' | 'first_workday'
7111 | 'cal_direction' ;
7112 time_list = time_list semicolon quoted_string
7113 | quoted_string ;
7114 time_tail = 'END' 'LC_TIME' EOL ;
7115
7116 % The following is the LC_PAPER category grammar
7117 lc_paper = paper_head paper_keyword* paper_tail
7118 | paper_head copy_FDCC_set paper_tail ;
7119 paper_head = 'LC_PAPER' EOL ;
7120 paper_keyword = paper_keyword_num number EOL ;
7121 paper_keyword_num = 'height' | 'width' ;
7122 paper_tail = 'END' 'LC_PAPER' EOL ;
7123
7124 % The following is the LC_NAME category grammar
7125 lc_name = name_head name_keyword* name_tail
7126 | name_head copy_FDCC_set name_tail ;
7127 name_head = 'LC_NAME' EOL ;
7128 name_keyword = name_keyword_string quoted_string EOL ;
7129 name_keyword_string = 'name_fmt' | 'name_gen' | 'name_mr'
7130 | 'name_mrs' | 'name_ms' | 'name_miss'
7131 | 'name_ms' ;
7132 name_tail = 'END' 'LC_NAME' EOL ;
7133
7134 % The following is the LC_ADDRESS category grammar
7135 lc_address = address_head address_keyword* address_tail

```

```
7136 | address_head copy_FDCC_set address_tail ;
7137 address_head = 'LC_ADDRESS' EOL ;
7138 address_keyword = address_keyword_string quoted_string EOL
7139 | address_keyword_num number EOL ;
7140 address_keyword_string = 'postal_fmt' | 'country_name' |
7141 'country_post' | 'country_ab2' | 'country_ab3'
7142 | 'country_car' | 'country_isbn' | 'lang_name' |
7143 'lang_ab' | 'lang_term' | 'lang_lib' ;
7144 address_keyword_num = "country_num" ;
7145 address_tail = 'END' 'LC_ADDRESS' EOL ;
7146
7147 % The following is the LC_TELEPHONE category grammar
7148 lc_tel = tel_head tel_keyword* tel_tail
7149 | tel_head copy_FDCC_set tel_tail ;
7150 tel_head = 'LC_TELEPHONE' EOL ;
7151 tel_keyword = tel_keyword_string quoted_string EOL ;
7152 tel_keyword_string = 'tel_int_fmt' | 'tel_dom_fmt' | 'int_select'
7153 | 'int_prefix' ;
7154 tel_tail = 'END' 'LC_TELEPHONE' EOL ;
7155
```

		Annex D	
		(informative)	
		Index	
7156			
7157			
7158			
7159			
7160			
7161	abbreviation	4.2	code_set_name 5.1
7162	abday	4.7	coded character 3.1.3
7163	abmon	4.7	col_weight_max 4.4, 4.4.3
7164	absolute ellipses	4.3	collating-element 4.4
7165	address	4.2	collating statements 4.4.1
7166	addresses	4.11	collating-symbol 4.4.6
7167	addset	5.1	collating element 3.1.13
7168	alpha	4.3.1	collating sequence 3.1.15
7169	alt_digits	4.7	collating-element 4.4.5
7170	am_pm	4.7	collating-symbol 4.4
7171	application	4.2	collation 3.1.12
7172	audience	4.2	combining 4.3.1
7173	blank	4.3.1	combining_level3 4.3.1
7174	block_separator	4.3.1	comment_char 4.1.4.1, 5.1
7175	byte	3.1.1	conformance 7
7176	cal_direction	4.7	contact 4.2
7177	category	4.2	continuation line 4.1.2
7178	category names	4.1	control characters 4.3.1
7179	category trailer	4.1	conversion_rate 4.5
7180	category header	4.1	copy 4.1.3, 4.2, 4.3.1, 4.4.2, 4.5, 4.6, 4.7, 4.8, 4.9,
7181	category body	4.1	
7182	character	3.1.2	4.10, 4.11, 4.12
7183	character, graphic	4.3.1	country_ab2 4.11
7184	character, special	4.3.1	country_ab3 4.11
7185	character representation	4.1.1	country_car 4.11
7186	character, native digit	4.3.1	country_isbn 4.11
7187	character, hexadecimal digit	4.3.1	country_name 4.11
7188	character, multibyte	4.1.1	country_num 4.11
7189	character, decimal constant	4.1.1	country_post 4.11
7190	character, hexadecimal constant	4.1.1	cultural convention 3.1.5
7191	character, space	4.3.1	currency_symbol 4.5
7192	character, octal constant	4.1.1	d_fmt 4.7
7193	character, control	4.3.1	d_t_fmt 4.7
7194	character, blank	4.3.1	date field descriptors 4.7.1
7195	character, digit	4.3.1	date 4.2
7196	character, punctuation	4.3.1	day 4.7
7197	character, printable	3.1.10	decimal_point 4.6
7198	character class	3.1.9	default_missing 4.3.2
7199	character, coded	3.1.3	definitions 3.1
7200	Character set rationale	B.2	digit 4.3.1
7201	charmap text	5.1	ellipses 4.3, 4.4.1, 5.1
7202	charmap	5, 4.1.4.4, 3.1.7	ellipses, absolute 4.3, 4.4.1
7203	charmap rationale	B.2	ellipses, symbolic 4.3, 4.4.1, 5.1
7204	class	4.3.1	email 4.2
7205	cntrl	4.3.1	equivalence class 3.1.16

7206	era	4.7	LC_NAME rationale	B.1.9
7207	era_d_fmt	4.7	LC_NUMERIC	4.6
7208	era_year	4.7	LC_NUMERIC rationale	B.1.5
7209	escape_char	4.1.4.2, 5.1, 6	LC_PAPER	4.9
7210	esqseq	5.1	LC_PAPER rationale	B.1.8
7211	euro	B.1.3	LC_TELEPHONE	4.12
7212	extended regular expression	4.8	LC_TELEPHONE rationale	B.1.11
7213	fax	4.2	LC_TIME	4.7
7214	FDCC-set, definition	4.1	LC_TIME rationale	B.1.6
7215	FDCC-set	4f	LC_X_	4
7216	FDCC-set	3.1.6	line continuation	4.1.4
7217	FDCC-set rationale	B.1	lower	4.3.1
7218	first_weekday	4.7	map	4.3.1
7219	first_workday	4.7	mb_cur_max	5.1
7220	frac_digits	4.5	mb_cur_min	5.1
7221	graph	4.3.1	messages	4.8
7222	graphic chracters	4.3.1	modified date field descriptors	4.7.2
7223	grouping	4.6	mon	4.7
7224	height	4.9	mon_decimal_point	4.5
7225	include	4.3.2	mon_grouping	4.5
7226	include	5.1	mon_thousands_sep	4.5
7227	include	4.3.2.2	monetary	4.5
7228	int_curr_symbol	4.5	multicharacter collating element	3.1.14
7229	int_frac_digits	4.5	n_cs_precedes	4.5
7230	int_n_cs_precedes	4.5	n_sep_by_space	4.5
7231	int_n_sep_by_space	4.5	n_sign_posn	4.5
7232	int_n_sign_posn	4.5	name formatting	4.10
7233	int_p_cs_precedes	4.5	name_fmt	4.10
7234	int_p_sep_by_space	4.5	name_gen	4.10
7235	int_p_sign_posn	4.5	name_miss	4.10
7236	int_prefix	4.12	name_mr	4.10
7237	int_select	4.12	name_mrs	4.10
7238	keywords	4.1	name_ms	4.10
7239	lang_ab	4.11	negative_sign	4.5
7240	lang_lib	4.11	noexpr	4.8
7241	lang_name	4.11	notations	3.2
7242	lang_term	4.11	numeric	4.6
7243	language	4.2	operands	4.1
7244	LC_ADDRESS	4.11	order_end	4.4.9, 4.4
7245	LC_ADDRESS rationale	B.1.10	order_start	4.4, 4.4.8
7246	LC_COLLATE	4.4	outdigit	4.3.1
7247	LC_COLLATE rationale	B.1.3	p_cs_precedes	4.5
7248	LC_CTYPE	4.3	p_sep_by_space	4.5
7249	LC_CTYPE rationale	B.1.2	p_sign_posn	4.5
7250	LC_IDENTIFICATION	4.2	paper format	4.9
7251	LC_IDENTIFICATION rationale	B.1.1	portable character set	3.2.4
7252	LC_MESSAGES	4.8	positive_sign	4.5
7253	LC_MESSAGES rationale	B.1.7	POSIX	1
7254	LC_MONETARY	4.5	POSIX differences	A
7255	LC_MONETARY rationale	B.1.4	POSIX conformance	4.2
7256	LC_NAME	4.10	postal addresses	4.11

7257	postal_fmt	4.11	valid_from	4.5
7258	pre-category statements	4.1.4	valid_to	4.5
7259	print	4.3.1	visible glyph portable characters	3.2.4
7260	printable character	3.1.10	week	4.7
7261	punct	4.3.1	white space	3.1.11
7262	punctuation characters	4.3.1	width	4.9
7263	redefine	4.3.2	xdigit	4.3.1
7264	references	2	yesexpr	4.8
7265	reorder-section-end	4.4.13		
7266	reorder-section-after	4.4.12		
7267	reorder-section-after	4.4		
7268	reorder-after	4.4		
7269	reorder-end	4.4		
7270	reorder-section-end	4.4		
7271	reorder-after	4.4.10		
7272	reorder-end	4.4.11		
7273	reorder-after rationale	B.1.2.1		
7274	repertoire rationale	B.3		
7275	repertoire	6		
7276	repertoiremap	6, 3.1.8, 5.1, 4.1.4.3		
7277	revision	4.2		
7278	scope	1		
7279	section	4.4, 4.4.4		
7280	source	4.2		
7281	space	4.3.1		
7282	special characters	4.3.1		
7283	symbol-equivalence	4.4, 4.4.7		
7284	symbolic ellipses	4.3, 5.1		
7285	symbolic name	4.1.1		
7286	syntax format	3.2.1		
7287	t_fmt	4.7		
7288	t_fmt_ampm	4.7		
7289	tel	4.2		
7290	tel_dom_fmt	4.12		
7291	tel_int_fmt	4.12		
7292	telephone numbers	4.12		
7293	territory	4.2		
7294	text file	3.1.4		
7295	thousands_sep	4.6		
7296	timezone	4.7		
7297	title	4.2		
7298	tolower	4.3.1		
7299	tosymmetric	4.3.1		
7300	toupper	4.3.1		
7301	translit_end	4.3.2		
7302	translit_ignore	4.3.2		
7303	translit_start	4.3.2		
7304	transliteration	4.3.2		
7305	transliteration statements	4.3.2.1		
7306	upper	4.3.1		

7307

BIBLIOGRAPHY

- 7308
7309
7310 The following specifications are considered relevant to this Technical Report, in addition
7311 to the normative references.
7312
7313 CEPT, CEPT-MAILCODE, *Country code for mail*.
7314
7315 ISO 646, *Information technology - ISO 7-bit coded character set for information inter-*
7316 *change*.
7317
7318 ISO/IEC 9899, *Information technology - Programming language C*.
7319
7320 ISO/IEC 14977, *Information technology - Syntactic metalanguage - Extended BNF*.
7321
7322 The Unicode Consortium: *The Unicode Standard, Version 2.0*, Addison Wesley
7323 Developers Press, July 1996. ISBN 0-201-48345-9.
7324
7325 IBM: *National Language Design Guide Volume 2 - National Language Support Reference*
7326 *Manual*, IBM SE09-8002-03, August 1994.
7327
7328 STRÍ: *Nordic Cultural Requirements on Information Technology (Summary report)*, STRÍ
7329 TS3, Libris, Reykjavík, Iceland 1992. ISBN 9979-9004-3-1.
7330