

Implementation of
ENV 1973:1995

Information technology — European Subsets of ISO/IEC 10646-1

ICS 35.040

NO COPYING WITHOUT BSI PERMISSION EXCEPT AS PERMITTED BY COPYRIGHT LAW



Document Sharing Hub

Committees responsible for this Draft for Development

This Draft for Development was entrusted to Technical Committee IST/2, Character sets and information coding, upon which the following bodies were represented:

- British Computer Society
- Institution of Electrical Engineers
- Joint Information Systems Committee
- Philips Research Laboratories
- PICON
- Reuters Limited
- Co-opted members

This Draft for Development, having been prepared under the direction of the Information Systems Technology Assembly, was published under the authority of the Standards Board and comes into effect on 15 July 1996

© BSI 04-2000

The following BSI references relate to the work on this Draft for Development:
Committee reference IST/2

ISBN 0 580 25786 X

Amendments issued since publication

Amd. No.	Date	Comments

Contents

	Page
Committees responsible	Inside front cover
National foreword	ii
Foreword	2
Text of ENV 1973	3

National foreword

This Draft for Development has been prepared by Technical Committee IST/2, Character sets and information coding, and is the English Language version of ENV 1973:1995, published by the European Committee for Standardization (CEN).

This Draft for Development is published under the direction of the Information Systems Technology Assembly whose Technical Committee IST/2 has the responsibility to:

- aid enquirers to understand the text;
- present to the responsible international/European committee any enquiries on interpretation, or proposals for change, and keep UK interests informed;
- monitor related international and European developments and promulgate them in the UK.

NOTE International and European Standards, as well as overseas standards, are available from Customer Services, BSI, 389 Chiswick High Road, London W4 4AL.

This Draft for Development is not to be regarded as a British Standard.

It is being issued in the Draft for Development series of publications and is of a provisional nature because the subject matter is still under development requiring wider exposure. It should be applied on this provisional basis so that information and experience of its practical application may be obtained.

Comments arising from the use of this Draft for Development are requested so that the UK experience can be reported to the European organization responsible for its conversion to a European Standard. A review of this publication will be initiated 2 years after its publication by the European organization so that a decision can be taken on its status at the end of its three-year life. The commencement of the review period will be notified by and announcement in *BSI News*.

According to the replies received by the end of the review period, the responsible BSI Committee will decide whether to support the conversion into a European Standard, to extend the life of the prestandard or to withdraw it. Comments should be sent in writing to the Secretary of BSI Technical Committee IST/2 at BSI, 389 Chiswick High Road, London, W4 4AL, giving the document reference and clause number and proposing, where possible, an appropriate revision of the text.

Summary of pages

This document comprises a front cover, an inside front cover, pages i and ii, the ENV title page, pages 2 to 47 and a back cover.

This standard has been updated (see copyright date) and may have had amendments incorporated. This will be indicated in the amendment table on the inside front cover.

English version

Information Technology — European Subsets of ISO/IEC 10646-1

Technologies de l'information — Jeux partiels
européens de l'ISO/CEI 10646-1

Informationstechnologie — Europäische
Untermengen von ISO/IEC 10646-1

This European Prestandard (ENV) was approved by CEN on 1995-10-12 as a prospective standard for provisional application. The period of validity of this ENV is limited initially to three years. After two years the members of CEN will be requested to submit their comments, particularly on the question whether the ENV can be converted into a European Standard (EN).

CEN members are required to announce the existence of this ENV in the same way as for an EN and to make the ENV available promptly at national level in an appropriate form. It is permissible to keep conflicting national standards in force (in parallel to the ENV) until the final decision about the possible conversion of the ENV into an EN is reached.

CEN members are the national standards bodies of Austria, Belgium, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Luxembourg, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and United Kingdom.

CEN

European Committee for Standardization
Comité Européen de Normalisation
Europäisches Komitee für Normung

Central Secretariat: rue de Stassart 36, B-1050 Brussels

Foreword

This European Prestandard has been prepared by the Technical Committee CEN/TC 304 “Character Set Technology” of which the secretariat is held by STRI.

It was approved in the month of December 1995.

This standard does not cancel or replace any other standard.

There is no known identical or conflicting national standard in Europe.

According to the CEN/CENELEC Internal Regulations, the national standards organisations of the following countries are bound to implement this European Standard: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Luxembourg, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the United Kingdom.

Contents

	Page
Foreword	2
0 Introduction	3
1 Scope	4
2 Normative references	4
3 Definitions and abbreviations	4
4 Compliance and conformance	6
5 The Minimum European Subset (MES)	6
6 Abstract and transfer syntax	23
Annex A (normative) Implementation Conformance Statement (ICS) proforma	24
Annex B (informative) List of languages covered by the MES	25
Annex C (informative) Examples of coded character sets whose repertoires are covered by the MES	27
Annex D (informative) List of MES characters in alphabetical order	28
Annex E (informative) The Extended European Subset (EES)	44
Annex F (informative) Rationale for inclusion of collections in the EES	46
Annex G (informative) Bibliography	47

0 Introduction

0.1 Requirements

There is a growing need for IT-communication across the national boundaries of Europe, with public administrations cooperating in large data systems, and commerce and trade between countries increasingly using IT-techniques; in addition, there are legal requirements regarding the spelling of personal names for individuals moving around in Europe to be considered. This leads to extensive requirements on the character sets used in European IT-equipment.

The countries in question have a large number of official languages and officially-accepted native minority languages. These require a large number of letters. In addition, a large number of other graphic characters are required.

0.2 Towards large character sets

The Minimum Subset specified by this standard provides a first step towards the implementation of large character sets in Europe. The Extended Subset, specified in informative Annex E, provides a second step by covering all characters belonging to European scripts. The implementation of all the world's characters is the ultimate goal.

It is not assumed that CEN nor Europeans in general have the expertise to make provision for languages originating outside Europe. In other parts of the world, work regarding the definition and implementation of locally-important scripts (*e.g.* Hebrew, Arabic, Devanagari, Korean) is ongoing. To date, however, coherent specification of the scripts of Europe (*i.e.* Latin, Cyrillic, Greek, Armenian, Georgian) has not been made.

This standard is one of the results of the CEN/TC 304 work on the characters and the scripts of Europe. It is based in part on the results of another work item that specifies the characters used by the indigenous languages of Europe.

0.3 ISO/IEC 10646-1 Universal Character Set (UCS), Basic Multilingual Plane (BMP)

Prior to the UCS, there was no character set standard that included all the characters and scripts needed by Europe. The possibilities to combine several standardized character sets — 7-bit or 8-bit coded — in the same IT-systems with existing code extension techniques have proven to be impractical and insufficient.

Part 1 of ISO/IEC 10646-1 (BMP) provides a good base for European character coding, since it defines fixed code positions for almost all presently known letters and a very large number of symbols and other characters for non-specialized use.

0.4 Implementing the Universal Character Set

It is likely that various compression schemes will be used for data storage and transmission of UCS encoded data. Required storage space and communication time will thus not double compared with 1-octet codes.

Implementation of ISO/IEC 10646-1 characters requires resources for supporting *e.g.*:

- font definitions
- tables for code table translation, sorting, matching and upper/lower casing

The resources needed are material or human, *e.g.*:

- system memory space for definitions, tables and functions
- licence costs for fonts
- documentation
- working time for program development
- learning time
- convenient and standardized keyboard input methods

0.5 Migration path

Some of the implementations problems discussed above could be solved by subsets of the UCS. The work on European subsets is particularly aimed at solving the problems of outputting the full character set of the UCS.

It is estimated that implementing the full character set of the UCS will be costly in the first stages of UCS use, and that manufacturers will implement in subset-stages. To ensure that a common subset usable to the vast majority of European users be available for a reasonable price, and as a guide to manufacturers, it will be helpful to specify, to users and procurers of systems, European subsets of the UCS encompassing all characters for use in European languages plus other frequently used characters. Such subsets may also be useful with regard to further standardization work (for example, on sorting specifications), so that the work is reasonably limited and still useful in a European environment. Only a limited number of subsets should be specified.

0.6 Rationale for a Minimum European Subset

Since the number of characters needed to fulfil the official European needs is only a fraction of the more than 30 000 characters presently encoded in the Basic Multilingual Plane of ISO/IEC 10646-1, substantial advantages can be expected from the specification and implementation of a Minimum European Subset thereof. With 926 characters, this standard covers a large set of graphic characters identified as being of immediate utility in Europe.

1 Scope

This European standard specifies a minimum *limited subset* of ISO/IEC 10646-1 for use in Europe. The Minimum European Subset (MES) is a *limited subset* that includes the characters used by a number of European languages.

NOTE These languages are listed in Annex B.

It covers the repertoires of all options of EN 1923, ENV 41503:1990, ENV 41505:1991, and ENV 41508:1990. It also covers the repertoires of commonly used registered coded character sets and commonly used vendor coded character sets.

NOTE Examples of these character sets are found in Annex C.

This standard also specifies, informatively, an extended *selected subset* of ISO/IEC 10646-1 for use in Europe. The Extended European Subset (EES) is a *selected subset* and a superset of the MES. The Extended European Subset differs from the Minimum European Subset mainly by basing its selection of characters by script and function, rather than by use in particular languages.

NOTE The EES is found in Annex E.

2 Normative references

2.1 ISO/IEC 10646-1:1993, *Information technology — Universal Multiple-Octet Coded Character Set (UCS) — Part 1: Architecture and Basic Multilingual Plane; and COR No. 1, to ISO/IEC 10646-1 on LETTER AE (adopted in April 1995, to be published).*

2.2 ISO/IEC 8824-1:1995, *Information technology — Abstract Syntax Notation One (ASN.1) — Specification of Basic Notation (third edition).*

3 Definitions and abbreviations

3.1 Basic definitions

For the purposes of this standard the basic definitions of ISO/IEC 10646-1 section 4 apply. The following are reproduced here for ease of reference:

3.1.1

coded character

a character together with its coded representation (ISO/IEC 10646-1)

3.1.2

coded character set

a set of unambiguous rules that establishes a character set and the relationship between the characters of the set and their coded representation (ISO/IEC 10646-1)

3.1.3**combining character**

a member of an identified subset of the coded character set of ISO/IEC 10646 intended for combination with the preceding non-combining graphic character, or with a sequence of combining characters preceded by a non-combining character (ISO/IEC 10646-1)

3.2 Other definitions

Also the following definitions apply:

3.2.1**abstract syntax**

the specification of Application Layer data or application-protocol-control-information by using notation rules which are independent of the encoding technique used to represent them (ISO/IEC 8822:1994)

3.2.2**character abstract syntax**

any abstract syntax whose values are specified as the set of character strings of zero, one or more characters from some specified collection of characters (ISO/IEC 8824-1:1995)

3.2.3**character transfer syntax**

any transfer syntax for a character abstract syntax for which all bit patterns are an integral multiple of eight bits (ISO/IEC 8824-1:1995)

3.2.4**concrete syntax**

those aspects of the rules used in the formal specification of data which embody a specific representation of that data (ISO 7498-1:1994)

3.2.5**limited subset**

a limited subset consists of a list of graphic characters in the specified subset. This specification allows applications and devices that were developed using other codes to interwork with this coded character set. A claim of conformance referring to a limited subset shall list the graphic characters in the subset by the names of graphic characters or code positions as defined in ISO/IEC 10646 (ISO/IEC 10646-1 § 13.1)

3.2.6**object identifier**

a value (distinguishable from all other such values) which is associated with an information object (ISO/IEC 8824-1:1995)

3.2.7**selected subset**

a selected subset consists of a list of collections of graphic characters as defined in ISO/IEC 10646. The collections from which the selection may be made are listed in Annex A of each part of ISO/IEC 10646. A selected subset shall always automatically include the Cells 20 to 7E of Row 00 of Plane 00 of Group 00. A claim of conformance referring to a selected subset shall list the collections chosen as defined in ISO/IEC 10646 (ISO/IEC 10646-1 § 13.2)

3.2.8**transfer syntax**

that concrete syntax used in the transfer of data between open systems (ISO 7498-1:1994)

3.3 Abbreviations

The following abbreviations apply:

3.3.1 BMP: Basic Multilingual Plane

3.3.2 EES: Extended European Subset

3.3.3 MES: Minimum European Subset

3.3.4 UCS: Universal Character Set

4 Compliance and conformance

4.0 Compliance

Clause 5 of this standard is in compliance with the provisions of ISO/IEC 10646-1 clause 13 for the specification of a limited subset.

4.1 Conformance to 10646

Implementations claiming conformance to this standard shall also conform to the conformance provisions of ISO/IEC 10646-1. A claim of conformance shall be supported by a completed Implementation Conformance Statement (ICS) proforma (see Annex A).

4.2 Support of the Minimum European Subset

4.2.1 The supported subset shall be the Minimum European Subset, as identified in clause 5.

4.2.2 A sending implementation shall be capable of coding the Minimum European Subset at level 1 of ISO/IEC 10646-1 using one of the codings as specified in ISO/IEC 10646-1.

4.2.3 A receiving implementation shall be capable of supporting the coding of the Minimum European Subset at level 1.

4.3 Receiving devices

4.3.1 A receiving device shall be able to present to an internal application the coded representation of any received character within the supported subset, in such a way that the internal application can identify the character.

4.3.2 A receiving device with retransmission capability shall be able to retransmit the coded representation of any character of the supported subset without change to its coded representation unless modified by the user.

4.3.3 Coded representation of characters not contained in the supported subset may or may not be retransmitted.

4.4 Originating devices

An originating device shall allow its user to supply any characters from the supported subset and be capable of transmitting their coded representation.

4.5 Receiving devices with rendering capabilities

A receiving device with rendering capability shall be able to display to the user in a recognizable form any character of the supported subset.

5 The Minimum European Subset (MES)

The following list contains the subset characters ordered in the same sequence as they appear in ISO/IEC 10646-1. The list of UCS hexadecimal code positions is normative. The additional information, that is, the graphic symbols (glyphs) and the character names, are only informative and are provided to facilitate the usage of the list.

Glyph	UCS Hex (Normative)	Character Name (Informative)
	0020	SPACE
!	0021	EXCLAMATION MARK
"	0022	QUOTATION MARK
#	0023	NUMBER SIGN
\$	0024	DOLLAR SIGN
%	0025	PERCENT SIGN
€	0026	AMPERSAND
'	0027	APOSTROPHE
(	0028	LEFT PARENTHESIS
)	0029	RIGHT PARENTHESIS
*	002A	ASTERISK
+	002B	PLUS SIGN
,	002C	COMMA
-	002D	HYPHEN-MINUS
.	002E	FULL STOP
/	002F	SOLIDUS
0	0030	DIGIT ZERO
1	0031	DIGIT ONE
2	0032	DIGIT TWO
3	0033	DIGIT THREE
4	0034	DIGIT FOUR
5	0035	DIGIT FIVE
6	0036	DIGIT SIX
7	0037	DIGIT SEVEN
8	0038	DIGIT EIGHT
9	0039	DIGIT NINE
:	003A	COLON
;	003B	SEMICOLON
<	003C	LESS-THAN SIGN
=	003D	EQUALS SIGN
>	003E	GREATER-THAN SIGN
?	003F	QUESTION MARK
@	0040	COMMERCIAL AT
A	0041	LATIN CAPITAL LETTER A
B	0042	LATIN CAPITAL LETTER B
C	0043	LATIN CAPITAL LETTER C
D	0044	LATIN CAPITAL LETTER D
E	0045	LATIN CAPITAL LETTER E
F	0046	LATIN CAPITAL LETTER F
G	0047	LATIN CAPITAL LETTER G
H	0048	LATIN CAPITAL LETTER H
I	0049	LATIN CAPITAL LETTER I
J	004A	LATIN CAPITAL LETTER J
K	004B	LATIN CAPITAL LETTER K
L	004C	LATIN CAPITAL LETTER L
M	004D	LATIN CAPITAL LETTER M
N	004E	LATIN CAPITAL LETTER N
O	004F	LATIN CAPITAL LETTER O
P	0050	LATIN CAPITAL LETTER P
Q	0051	LATIN CAPITAL LETTER Q
R	0052	LATIN CAPITAL LETTER R

Glyph	UCS Hex (Normative)	Character Name (Informative)
S	0053	LATIN CAPITAL LETTER S
T	0054	LATIN CAPITAL LETTER T
U	0055	LATIN CAPITAL LETTER U
V	0056	LATIN CAPITAL LETTER V
W	0057	LATIN CAPITAL LETTER W
X	0058	LATIN CAPITAL LETTER X
Y	0059	LATIN CAPITAL LETTER Y
Z	005A	LATIN CAPITAL LETTER Z
[	005B	LEFT SQUARE BRACKET
\	005C	REVERSE SOLIDUS
]	005D	RIGHT SQUARE BRACKET
^	005E	CIRCUMFLEX ACCENT
ˉ	005F	LOW LINE
˘	0060	GRAVE ACCENT
a	0061	LATIN SMALL LETTER A
b	0062	LATIN SMALL LETTER B
c	0063	LATIN SMALL LETTER C
d	0064	LATIN SMALL LETTER D
e	0065	LATIN SMALL LETTER E
f	0066	LATIN SMALL LETTER F
g	0067	LATIN SMALL LETTER G
h	0068	LATIN SMALL LETTER H
i	0069	LATIN SMALL LETTER I
j	006A	LATIN SMALL LETTER J
k	006B	LATIN SMALL LETTER K
l	006C	LATIN SMALL LETTER L
m	006D	LATIN SMALL LETTER M
n	006E	LATIN SMALL LETTER N
o	006F	LATIN SMALL LETTER O
p	0070	LATIN SMALL LETTER P
q	0071	LATIN SMALL LETTER Q
r	0072	LATIN SMALL LETTER R
s	0073	LATIN SMALL LETTER S
t	0074	LATIN SMALL LETTER T
u	0075	LATIN SMALL LETTER U
v	0076	LATIN SMALL LETTER V
w	0077	LATIN SMALL LETTER W
x	0078	LATIN SMALL LETTER X
y	0079	LATIN SMALL LETTER Y
z	007A	LATIN SMALL LETTER Z
{	007B	LEFT CURLY BRACKET
	007C	VERTICAL LINE
}	007D	RIGHT CURLY BRACKET
~	007E	TILDE
	00A0	NO-BREAK SPACE
¡	00A1	INVERTED EXCLAMATION MARK
¢	00A2	CENT SIGN
£	00A3	POUND SIGN
¤	00A4	CURRENCY SIGN
¥	00A5	YEN SIGN
¦	00A6	BROKEN BAR
§	00A7	SECTION SIGN
¨	00A8	DIAERESIS
©	00A9	COPYRIGHT SIGN
ª	00AA	FEMININE ORDINAL INDICATOR
«	00AB	LEFT-POINTING DOUBLE ANGLE QUOTATION MARK
¬	00AC	NOT SIGN

Glyph	UCS Hex (Normative)	Character Name (Informative)
–	00AD	SOFT HYPHEN
®	00AE	REGISTERED SIGN
ˉ	00AF	MACRON
°	00B0	DEGREE SIGN
±	00B1	PLUS-MINUS SIGN
²	00B2	SUPERSCRIPIT TWO
³	00B3	SUPERSCRIPIT THREE
´	00B4	ACUTE ACCENT
μ	00B5	MICRO SIGN
¶	00B6	PILCROW SIGN
·	00B7	MIDDLE DOT
¸	00B8	CEDILLA
¹	00B9	SUPERSCRIPIT ONE
º	00BA	MASCULINE ORDINAL INDICATOR
»	00BB	RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK
¼	00BC	VULGAR FRACTION ONE QUARTER
½	00BD	VULGAR FRACTION ONE HALF
¾	00BE	VULGAR FRACTION THREE QUARTERS
¿	00BF	INVERTED QUESTION MARK
À	00C0	LATIN CAPITAL LETTER A WITH GRAVE
Á	00C1	LATIN CAPITAL LETTER A WITH ACUTE
Â	00C2	LATIN CAPITAL LETTER A WITH CIRCUMFLEX
Ã	00C3	LATIN CAPITAL LETTER A WITH TILDE
Ä	00C4	LATIN CAPITAL LETTER A WITH DIAERESIS
Å	00C5	LATIN CAPITAL LETTER A WITH RING ABOVE
Æ	00C6	LATIN CAPITAL LETTER AE
Ç	00C7	LATIN CAPITAL LETTER C WITH CEDILLA
È	00C8	LATIN CAPITAL LETTER E WITH GRAVE
É	00C9	LATIN CAPITAL LETTER E WITH ACUTE
Ê	00CA	LATIN CAPITAL LETTER E WITH CIRCUMFLEX
Ë	00CB	LATIN CAPITAL LETTER E WITH DIAERESIS
Ì	00CC	LATIN CAPITAL LETTER I WITH GRAVE
Í	00CD	LATIN CAPITAL LETTER I WITH ACUTE
Î	00CE	LATIN CAPITAL LETTER I WITH CIRCUMFLEX
Ï	00CF	LATIN CAPITAL LETTER I WITH DIAERESIS
Ð	00D0	LATIN CAPITAL LETTER ETH
Ñ	00D1	LATIN CAPITAL LETTER N WITH TILDE
Ò	00D2	LATIN CAPITAL LETTER O WITH GRAVE
Ó	00D3	LATIN CAPITAL LETTER O WITH ACUTE
Ô	00D4	LATIN CAPITAL LETTER O WITH CIRCUMFLEX
Õ	00D5	LATIN CAPITAL LETTER O WITH TILDE
Ö	00D6	LATIN CAPITAL LETTER O WITH DIAERESIS
×	00D7	MULTIPLICATION SIGN
Ø	00D8	LATIN CAPITAL LETTER O WITH STROKE
Ù	00D9	LATIN CAPITAL LETTER U WITH GRAVE
Ú	00DA	LATIN CAPITAL LETTER U WITH ACUTE
Û	00DB	LATIN CAPITAL LETTER U WITH CIRCUMFLEX
Ü	00DC	LATIN CAPITAL LETTER U WITH DIAERESIS
Ý	00DD	LATIN CAPITAL LETTER Y WITH ACUTE
Þ	00DE	LATIN CAPITAL LETTER THORN
ß	00DF	LATIN SMALL LETTER SHARP S
à	00E0	LATIN SMALL LETTER A WITH GRAVE
á	00E1	LATIN SMALL LETTER A WITH ACUTE
â	00E2	LATIN SMALL LETTER A WITH CIRCUMFLEX
ã	00E3	LATIN SMALL LETTER A WITH TILDE
ä	00E4	LATIN SMALL LETTER A WITH DIAERESIS
å	00E5	LATIN SMALL LETTER A WITH RING ABOVE
æ	00E6	LATIN SMALL LETTER AE

Glyph	UCS Hex (Normative)	Character Name (Informative)
ç	00E7	LATIN SMALL LETTER C WITH CEDILLA
è	00E8	LATIN SMALL LETTER E WITH GRAVE
é	00E9	LATIN SMALL LETTER E WITH ACUTE
ê	00EA	LATIN SMALL LETTER E WITH CIRCUMFLEX
ë	00EB	LATIN SMALL LETTER E WITH DIAERESIS
ì	00EC	LATIN SMALL LETTER I WITH GRAVE
í	00ED	LATIN SMALL LETTER I WITH ACUTE
î	00EE	LATIN SMALL LETTER I WITH CIRCUMFLEX
ï	00EF	LATIN SMALL LETTER I WITH DIAERESIS
ð	00F0	LATIN SMALL LETTER ETH
ñ	00F1	LATIN SMALL LETTER N WITH TILDE
ò	00F2	LATIN SMALL LETTER O WITH GRAVE
ó	00F3	LATIN SMALL LETTER O WITH ACUTE
ô	00F4	LATIN SMALL LETTER O WITH CIRCUMFLEX
õ	00F5	LATIN SMALL LETTER O WITH TILDE
ö	00F6	LATIN SMALL LETTER O WITH DIAERESIS
÷	00F7	DIVISION SIGN
ø	00F8	LATIN SMALL LETTER O WITH STROKE
ù	00F9	LATIN SMALL LETTER U WITH GRAVE
ú	00FA	LATIN SMALL LETTER U WITH ACUTE
û	00FB	LATIN SMALL LETTER U WITH CIRCUMFLEX
ü	00FC	LATIN SMALL LETTER U WITH DIAERESIS
ý	00FD	LATIN SMALL LETTER Y WITH ACUTE
þ	00FE	LATIN SMALL LETTER THORN
ÿ	00FF	LATIN SMALL LETTER Y WITH DIAERESIS
Ā	0100	LATIN CAPITAL LETTER A WITH MACRON
ā	0101	LATIN SMALL LETTER A WITH MACRON
Ă	0102	LATIN CAPITAL LETTER A WITH BREVE
ă	0103	LATIN SMALL LETTER A WITH BREVE
Ą	0104	LATIN CAPITAL LETTER A WITH OGONEK
ą	0105	LATIN SMALL LETTER A WITH OGONEK
Ć	0106	LATIN CAPITAL LETTER C WITH ACUTE
ć	0107	LATIN SMALL LETTER C WITH ACUTE
Ĉ	0108	LATIN CAPITAL LETTER C WITH CIRCUMFLEX
ĉ	0109	LATIN SMALL LETTER C WITH CIRCUMFLEX
Č	010A	LATIN CAPITAL LETTER C WITH DOT ABOVE
č	010B	LATIN SMALL LETTER C WITH DOT ABOVE
Č	010C	LATIN CAPITAL LETTER C WITH CARON
č	010D	LATIN SMALL LETTER C WITH CARON
Ď	010E	LATIN CAPITAL LETTER D WITH CARON
ď	010F	LATIN SMALL LETTER D WITH CARON
Ð	0110	LATIN CAPITAL LETTER D WITH STROKE
ð	0111	LATIN SMALL LETTER D WITH STROKE
Ē	0112	LATIN CAPITAL LETTER E WITH MACRON
ē	0113	LATIN SMALL LETTER E WITH MACRON
Ĕ	0114	LATIN CAPITAL LETTER E WITH BREVE
ĕ	0115	LATIN SMALL LETTER E WITH BREVE
Ė	0116	LATIN CAPITAL LETTER E WITH DOT ABOVE
ė	0117	LATIN SMALL LETTER E WITH DOT ABOVE
Ę	0118	LATIN CAPITAL LETTER E WITH OGONEK
ę	0119	LATIN SMALL LETTER E WITH OGONEK
Ě	011A	LATIN CAPITAL LETTER E WITH CARON
ě	011B	LATIN SMALL LETTER E WITH CARON
Ĝ	011C	LATIN CAPITAL LETTER G WITH CIRCUMFLEX
ĝ	011D	LATIN SMALL LETTER G WITH CIRCUMFLEX
Ğ	011E	LATIN CAPITAL LETTER G WITH BREVE
ğ	011F	LATIN SMALL LETTER G WITH BREVE

Glyph	UCS Hex (Normative)	Character Name (Informative)
Ġ	0120	LATIN CAPITAL LETTER G WITH DOT ABOVE
ġ	0121	LATIN SMALL LETTER G WITH DOT ABOVE
Ģ	0122	LATIN CAPITAL LETTER G WITH CEDILLA
ģ	0123	LATIN SMALL LETTER G WITH CEDILLA
Ĥ	0124	LATIN CAPITAL LETTER H WITH CIRCUMFLEX
ĥ	0125	LATIN SMALL LETTER H WITH CIRCUMFLEX
Ħ	0126	LATIN CAPITAL LETTER H WITH STROKE
ħ	0127	LATIN SMALL LETTER H WITH STROKE
İ	0128	LATIN CAPITAL LETTER I WITH TILDE
ĩ	0129	LATIN SMALL LETTER I WITH TILDE
Ī	012A	LATIN CAPITAL LETTER I WITH MACRON
ī	012B	LATIN SMALL LETTER I WITH MACRON
Ĭ	012C	LATIN CAPITAL LETTER I WITH BREVE
ĭ	012D	LATIN SMALL LETTER I WITH BREVE
Į	012E	LATIN CAPITAL LETTER I WITH OGONEK
į	012F	LATIN SMALL LETTER I WITH OGONEK
Ĳ	0130	LATIN CAPITAL LETTER I WITH DOT ABOVE
ı	0131	LATIN SMALL LETTER DOTLESS I
Ĵ	0132	LATIN CAPITAL LIGATURE IJ
ĵ	0133	LATIN SMALL LIGATURE IJ
Ĵ	0134	LATIN CAPITAL LETTER J WITH CIRCUMFLEX
ĵ	0135	LATIN SMALL LETTER J WITH CIRCUMFLEX
Ķ	0136	LATIN CAPITAL LETTER K WITH CEDILLA
ķ	0137	LATIN SMALL LETTER K WITH CEDILLA
ķ	0138	LATIN SMALL LETTER KRA
Ĺ	0139	LATIN CAPITAL LETTER L WITH ACUTE
ĺ	013A	LATIN SMALL LETTER L WITH ACUTE
Ľ	013B	LATIN CAPITAL LETTER L WITH CEDILLA
ļ	013C	LATIN SMALL LETTER L WITH CEDILLA
Ļ	013D	LATIN CAPITAL LETTER L WITH CARON
ļ	013E	LATIN SMALL LETTER L WITH CARON
Ł	013F	LATIN CAPITAL LETTER L WITH MIDDLE DOT
ł	0140	LATIN SMALL LETTER L WITH MIDDLE DOT
Ł	0141	LATIN CAPITAL LETTER L WITH STROKE
ł	0142	LATIN SMALL LETTER L WITH STROKE
Ń	0143	LATIN CAPITAL LETTER N WITH ACUTE
ń	0144	LATIN SMALL LETTER N WITH ACUTE
Ñ	0145	LATIN CAPITAL LETTER N WITH CEDILLA
ñ	0146	LATIN SMALL LETTER N WITH CEDILLA
Ñ	0147	LATIN CAPITAL LETTER N WITH CARON
ň	0148	LATIN SMALL LETTER N WITH CARON
ŋ	0149	LATIN SMALL LETTER N PRECEDED BY APOSTROPHE
Ŋ	014A	LATIN CAPITAL LETTER ENG
ŋ	014B	LATIN SMALL LETTER ENG
Ō	014C	LATIN CAPITAL LETTER O WITH MACRON
ō	014D	LATIN SMALL LETTER O WITH MACRON
Ö	014E	LATIN CAPITAL LETTER O WITH BREVE
ö	014F	LATIN SMALL LETTER O WITH BREVE
Ȫ	0150	LATIN CAPITAL LETTER O WITH DOUBLE ACUTE
ȫ	0151	LATIN SMALL LETTER O WITH DOUBLE ACUTE
Œ	0152	LATIN CAPITAL LIGATURE OE
œ	0153	LATIN SMALL LIGATURE OE
Ŕ	0154	LATIN CAPITAL LETTER R WITH ACUTE
ŗ	0155	LATIN SMALL LETTER R WITH ACUTE
Ŗ	0156	LATIN CAPITAL LETTER R WITH CEDILLA
ŗ	0157	LATIN SMALL LETTER R WITH CEDILLA
Ř	0158	LATIN CAPITAL LETTER R WITH CARON
ř	0159	LATIN SMALL LETTER R WITH CARON

Glyph	UCS Hex (Normative)	Character Name (Informative)
Š	015A	LATIN CAPITAL LETTER S WITH ACUTE
š	015B	LATIN SMALL LETTER S WITH ACUTE
Ŝ	015C	LATIN CAPITAL LETTER S WITH CIRCUMFLEX
ŝ	015D	LATIN SMALL LETTER S WITH CIRCUMFLEX
Ș	015E	LATIN CAPITAL LETTER S WITH CEDILLA
ș	015F	LATIN SMALL LETTER S WITH CEDILLA
Š	0160	LATIN CAPITAL LETTER S WITH CARON
š	0161	LATIN SMALL LETTER S WITH CARON
Ť	0162	LATIN CAPITAL LETTER T WITH CEDILLA
ť	0163	LATIN SMALL LETTER T WITH CEDILLA
Ť	0164	LATIN CAPITAL LETTER T WITH CARON
ť	0165	LATIN SMALL LETTER T WITH CARON
Ƨ	0166	LATIN CAPITAL LETTER T WITH STROKE
Ƨ	0167	LATIN SMALL LETTER T WITH STROKE
Ū	0168	LATIN CAPITAL LETTER U WITH TILDE
ū	0169	LATIN SMALL LETTER U WITH TILDE
Ū	016A	LATIN CAPITAL LETTER U WITH MACRON
ū	016B	LATIN SMALL LETTER U WITH MACRON
Ů	016C	LATIN CAPITAL LETTER U WITH BREVE
ů	016D	LATIN SMALL LETTER U WITH BREVE
Ů	016E	LATIN CAPITAL LETTER U WITH RING ABOVE
ů	016F	LATIN SMALL LETTER U WITH RING ABOVE
Ů	0170	LATIN CAPITAL LETTER U WITH DOUBLE ACUTE
ů	0171	LATIN SMALL LETTER U WITH DOUBLE ACUTE
U	0172	LATIN CAPITAL LETTER U WITH OGONEK
u	0173	LATIN SMALL LETTER U WITH OGONEK
Ŵ	0174	LATIN CAPITAL LETTER W WITH CIRCUMFLEX
ŵ	0175	LATIN SMALL LETTER W WITH CIRCUMFLEX
Ŷ	0176	LATIN CAPITAL LETTER Y WITH CIRCUMFLEX
ŷ	0177	LATIN SMALL LETTER Y WITH CIRCUMFLEX
ÿ	0178	LATIN CAPITAL LETTER Y WITH DIAERESIS
Ž	0179	LATIN CAPITAL LETTER Z WITH ACUTE
ž	017A	LATIN SMALL LETTER Z WITH ACUTE
Ž	017B	LATIN CAPITAL LETTER Z WITH DOT ABOVE
ž	017C	LATIN SMALL LETTER Z WITH DOT ABOVE
Ž	017D	LATIN CAPITAL LETTER Z WITH CARON
ž	017E	LATIN SMALL LETTER Z WITH CARON
ſ	017F	LATIN SMALL LETTER LONG S
f	0192	LATIN SMALL LETTER F WITH HOOK
Ʒ	01B7	LATIN CAPITAL LETTER EZH
Đ	01C4	LATIN CAPITAL LETTER DZ WITH CARON
đ	01C6	LATIN SMALL LETTER DZ WITH CARON
Ł	01C7	LATIN CAPITAL LETTER LJ
ł	01C9	LATIN SMALL LETTER LJ
Ń	01CA	LATIN CAPITAL LETTER NJ
ń	01CC	LATIN SMALL LETTER NJ
Ä	01DE	LATIN CAPITAL LETTER A WITH DIAERESIS AND MACRON
ä	01DF	LATIN SMALL LETTER A WITH DIAERESIS AND MACRON
Ġ	01E4	LATIN CAPITAL LETTER G WITH STROKE
ġ	01E5	LATIN SMALL LETTER G WITH STROKE
Ģ	01E6	LATIN CAPITAL LETTER G WITH CARON
ģ	01E7	LATIN SMALL LETTER G WITH CARON
Ķ	01E8	LATIN CAPITAL LETTER K WITH CARON
ķ	01E9	LATIN SMALL LETTER K WITH CARON
Š	01EE	LATIN CAPITAL LETTER EZH WITH CARON
š	01EF	LATIN SMALL LETTER EZH WITH CARON
Ʊ	01F1	LATIN CAPITAL LETTER DZ