

Chapter 7

Code Charts

The code charts that follow present the individual characters of the Unicode Standard. The character repertoire is laid out in successive character blocks, described in *Chapter 6, Character Block Descriptions*. A character block always begins on the left-hand page with the character chart giving an overview of the character block as a whole.

A character names list follows the character chart (except for CJK ideographs and Hangul syllables, as discussed in this chapter). The character names list itemizes every character in the block and provides supplementary information in many cases.

An index to distinctive character names is at the back of this book; a full set of character names (including earlier Version 1.0 names) are in the *Unicode Character Database* on the CD-ROM.

7.1 Character Names List

The following illustration identifies the components of typical entries in the character names list.

code image entry

| | | | |
|------|---|--|-------------------------------|
| 00AF | ̄ | MACRON | (Unicode name) |
| | | = SPACING MACRON | (Version 1.0 name) |
| | | = overline | (alternative name) |
| | | = APL overbar | |
| | | • this is a spacing character | (informative note) |
| | | → 02C3 ˘ modifier letter macron | (cross reference) |
| | | → 0304 ̄ combining macron | |
| | | → 0305 ̄ combining overline | |
| | | ≈ 0020 SP + 0304 ̄ | (compatibility decomposition) |
| 00E5 | å | LATIN SMALL LETTER A WITH RING ABOVE | |
| | | = LATIN SMALL LETTER A RING | |
| | | • Danish, Norwegian, Swedish | |
| | | ≡ 0061 a + 030A ̊ | (canonical decomposition) |

Images in the Code Charts and Character Lists

In all cases, a representative glyph is printed to represent a character. The image shown in a grid cell of a panel should not be considered to be the prescriptive form of a character; it is merely intended as a typical representation of the character encoded by the value. For example, U+0061 LATIN SMALL LETTER A can be represented by **a** or *a*.

Combining characters are shown with a dotted circle. A character that is shown as a name or abbreviation surrounded by a dashed box has no visible manifestation on its own.

Cross References

Cross referenced characters (preceded by →) have various characteristics: explicit inequality, the other member of a case pair, or some other linguistic relationship.

- **Explicit inequality.** The two characters are not identical, although the glyphs that depict them are identical or very close.

003A : COLON
 → 0589 : armenian full stop
 → 2236 : ratio

- **Other linguistic relationship.** These relationships include transliterations (such as between Serbian and Croatian), typographically unrelated characters used to represent the same sound, and so on.

01C9 lj LATIN SMALL LETTER LJ
 → 0459 њ cyrillic small letter lje
 ≈ 006C l + 006A j

Case Form Mappings

When a case mapping corresponds *solely* to a difference based on SMALL versus CAPITAL in the names of the characters, the case mapping is not given in the names list but is given only in the *Unicode Character Database* on the CD-ROM.

0041 A LATIN CAPITAL LETTER A
 01F2 Dz LATIN CAPITAL LETTER D WITH SMALL LETTER Z
 ≈ 0044 D + 007A z

When the case mapping cannot be predicted from the name, the information is given in a note.

00DF ß LATIN SMALL LETTER SHARP S
 = ess-zed
 • German
 • uppercase is "SS"
 → 03B2 β greek small letter beta

Decompositions

The decomposition sequence (one or more letters) given for a character is either its canonical decomposition or its compatibility decomposition. The canonical mapping is marked with an *identical to* symbol ≡.

00E5 å LATIN SMALL LETTER A WITH RING ABOVE
 = LATIN SMALL LETTER A RING
 • Danish, Norwegian, Swedish
 ≡ 0061 a + 030A ◌

212B Å ANGSTROM SIGN
 → 00C5 Å latin capital letter a with ring above
 ≡ 00C5 Å

Compatibility decompositions are marked with an *almost equal to* symbol ≈. Formatting information may be indicated inside angle brackets.

01F2 Dz LATIN CAPITAL LETTER D WITH SMALL LETTER Z
 ≈ 0044 D + 007A z


FF21 A FULLWIDTH LATIN CAPITAL LETTER A
 ≈ <wide> + 0041 A

Decompositions are not necessarily full decompositions. For example, the decomposition for U+212B ANGSTROM SIGN can be further decomposed using the canonical decomposition for U+00C5 LATIN SMALL LETTER A WITH RING ABOVE. (For more information on decomposition, see *Section 3.6, Decomposition.*)

Information About Languages


An informative note may include a list of the language(s) using that character where this information is considered useful. For case pairs, the annotation is given only for the lower-case form, to avoid needless repetition. An ellipsis “...” indicates that the listed languages cited are merely the principal ones among many.


Reserved Characters



Character codes that are marked “<reserved>” are unassigned and reserved for future encoding. Reserved codes are indicated by a  glyph.

FFFC  <reserved>

Reserved codes may also have cross references to assigned characters located elsewhere.

0373  <reserved>
 → 00A3 £ pound sign

Character codes that are marked “<not a character>” are permanently unassigned; they will never be assigned a character. Reserved codes are indicated by a  glyph.

FFFF  <not a character>
 • the value FFFF  is guaranteed not to be a Unicode character at all

7.2 CJK Unified Ideographs

A character names list is not provided for the *CJK Unified Ideographs* character block since the name of a unified ideograph simply consists of its Unicode value preceded by CJK UNIFIED IDEOGRAPH-. The character charts also take a different form, which show how the Unicode Han ideographic character set maps to several of the source standards (see Table 7-1). This table is also included on the CD-ROM.

Each Unicode character is shown with its Unicode value and a representative glyph. There are six additional fields for each of the primary source character sets. Characters with the compatibility range (U+F900 → U+FAFF) are also found within the main CJK range.

Table 7-1. Han Character Chart Entries

| Key | Prefix | Subset | Code Format | |
|-----|--------|---|-----------------|------------|
| G | 0 | GB 2312-80 | Row/column | |
| | 1 | GB 12345-80 | | |
| | 3 | GB 7589-87 | | |
| | 5 | GB 7590-87 | | |
| | 7 | General Purpose Han Characters for Modern Chinese | | |
| | 8 | GB 8565-89 | | |
| | J | 0 | JIS X 0208-1990 | Row/column |
| | | 1 | JIS X 0212-1990 | |
| K | 0 | KS C 5601-1987 | Row/column | |
| | 1 | KS C 5657-1991 | | |
| B | | Big Five | Hexadecimal | |
| C | 1 | CNS 11643-1986 (1st plane) | Hexadecimal | |
| | 2 | CNS 11643-1986 (2nd plane) | | |
| | E | CNS 11643-1986 (14th plane) | | |
| A | | ANSI Z39.64-1989 | Hexadecimal | |

Their placement is indicated by the small number beneath their Unicode code value in the charts.

- ➔ GB 2312-80 and GB 12345-80 are overlapping standards; most of the characters in the former are also found in the latter. Characters showing a mapping in the charts to GB 2312-80 may also be mapped into GB 12345-80. (For fuller information on how GB 12345-80 maps to Unicode, see the tables available on the CD-ROM.)

A radical/stroke index to CJK ideographs is in *Chapter 8, Han Radical Stroke Index*.

7.3 Hangul Syllables

A character names list is not provided for characters in the *Hangul Syllables Area* (U+AC00 → U+D7A3) since the name of a Hangul syllable can be determined by algorithm (as described in *Section 3.10, Combining Jamo Behavior*).