Chapter 12

Symbols

The universe of symbols is rich and open-ended. The collection of encoded symbols in the Unicode Standard encompasses the following:

- Currency Symbols
- Letterlike Symbols
- Number Forms
- Mathematical Operators and Arrows
- Technical Symbols
- Geometrical Symbols
- Miscellaneous Symbols and Dingbats
- Enclosed and Square
- Braille Patterns

There are other notational systems not covered by the Unicode Standard. Some symbols mark the transition between pictorial items and text elements; because they do not have a well-defined place in plain text, they are not encoded here.

Combining marks may be used with symbols, particularly the set encoded at U+20D0.. U+20FF (see *Section 7.9, Combining Marks*).

Letterlike and currency symbols, as well as number forms including superscripts and subscripts are typically subject to the same font and style changes as the surrounding text. Some, but not all, of the square and enclosed symbols occur in East Asian contexts and generally follow the prevailing type styles.

Other symbols have an appearance that is independent of type style, or a more limited or altogether different range of type style variation than the regular text surrounding them. Symbols such as mathematical operators can be used with any script, or independent of any script.

In a bidirectional context (see *Section 3.12, Bidirectional Behavior*), symbol characters have no inherent directionality, but resolve according to the Unicode bidirectional algorithm. Where the image of a symbol is not bilaterally symmetric, the mirror image is used when the character is part of the right-to-left text stream (see *Section 4.7, Mirrored—Normative*).

Dingbats and optical character recognition characters are different from all other characters in the standard in that they are encoded based on their precise appearance.

Braille patterns are a special case, because they can be used to write text. They are included as symbols, as the Unicode Standard encodes only their shapes; the association of letters to patterns is left to other standards. When a character stream is intended primarily to convey text information, it should be coded using one of the scripts. Only when it is intended to

convey a particular binding of text to Braille pattern sequence should it be coded using the Braille patterns.

Many symbols encoded in the Unicode Standard are obsolescent, and intended to support legacy implementations, such as terminal emulation or other charcter mode user interfaces. Examples include box drawing components and control pictures.

12.1 Currency Symbols

Currency Symbols: U+20A0–U+20CF

This block contains currency symbols not encoded in other blocks. Where the Unicode Standard follows the layout of an existing standard, such as for the ASCII, Latin 1, and Thai blocks, the currency symbols are encoded in those blocks, rather than here.

Unification. The Unicode Standard does not duplicate encodings where more than one currency is expressed with the same symbol. Many currency symbols are overstruck letters. There are therefore many minor variants, such as the U+0024 DOLLAR SIGN $, with one or two vertical bars, or other graphical variation. The Unicode Standard considers these variants to be typographical and provides a single encoding.

Claims that glyph variants of a certain currency symbol are used consistently to indicate a particular currency could not be substantiated upon further research. See ISO/IEC DIS 10367, Annex B (informative) for an example of multiple renderings for U+00A3 POUND SIGN.

Fonts. Currency symbols are commonly designed to be at digit width. Like letters, they follow the style of the font.

Table 12-1 lists common currency symbols encoded in other blocks.

Table 12-1. Other Currency Symbols

Dollar, milreis, escudo	U+0024	DOLLAR SIGN
Cent	U+00A2	CENT SIGN
Pound	U+00A3	POUND SIGN
General currency	U+00A4	CURRENCY SIGN
Yen or yuan	U+00A5	YEN SIGN
Dutch florin	U+0192	LATIN SMALL LETTER F WITH HOOK
Baht	U+0E3F	THAI CURRENCY SYMBOL BAHT
Riel	U+17DB	KHMER CURRENCY SYMBOL RIEL

For additional forms of currency symbols, see fullwidth forms (U+FFE0..U+FFE6).

Euro Sign. The new single currency for member countries of the European Monetary Union (EMU) is the euro. The euro character is encoded in the Unicode Standard as U+20AC EURO SIGN.

12.2 Letterlike Symbols

Letterlike Symbols: U+2100–U+214F

Letterlike symbols are symbols derived in some way from ordinary letters of an alphabetic script. This block includes symbols based on Latin, Greek, and Hebrew letters. Many of these symbols are encoded for compatibility. In general, the usage of distinct codes for letterlike symbols that are merely font variants or alternative representations of other character sequences is strongly discouraged. When using letters as symbols in equations and formulae, as well as in other contexts, use normal alphabetic forms in the appropriate styles. For example, to represent degrees Celsius "°C", use a sequence of U+00B0 DEGREE SIGN + U+0043 LATIN CAPITAL LETTER C, rather than U+2103 DEGREE CELSIUS. For searching, treat these two sequences as identical.

Despite its name, U+2118 SCRIPT CAPITAL P is neither script nor capital—it is uniquely the Weierstrass elliptic function derived from a calligraphic lowercase p.

U+2116 NUMERO SIGN is provided both for Cyrillic use, where it looks like №, and for compatibility with Asian standards, where it looks like №. The French practice is to use "N" followed by the degree sign: N°.

In the context of East Asian typography, letterlike symbols are rendered as "wide" characters occupying a full cell. They remain upright in vertical layout, contrary to the rotated rendering of their regular letter equivalents.

Where the letterlike symbols have alphabetic equivalents, they collate in alphabetic sequence; otherwise, they should be treated as neutral symbols. The letterlike symbols may have different directional properties than normal letters; for example, the four transfinite cardinal symbols (U+2135..U+2138) are used in ordinary mathematical text and do not share the strong right-to-left directionality of the Hebrew letters from which they are derived.

Styles. The letterlike symbols include some of the few instances in which the Unicode Standard encodes stylistic variants of letters as distinct characters. For example, there are instances of black letter, double-struck, and script styles for certain Latin letters used as mathematical symbols. The choice of these stylistic variants for encoding reflects their common use as distinct symbols. It is recognized that a particular style can be applied to any Latin letter with a resulting semantic distinction in mathematical or logical text; applications that require such systematic stylistic semantics should achieve them by using styles directly, rather than by seeking to extend the character-by-character encoding of such variants in the Unicode Standard.

The black letter style is often referred to as *Fraktur* or *Gothic* in various sources. Technically, Fraktur and Gothic typefaces are distinct designs from black letter, but no encoding distinctions are implied in the various symbol sources. The Unicode Standard simply uses black letter forms as the archetypes.

A similar consideration applies to the double-struck style. This style is not literally double-struck, but is instead an open outline design that gives the visual appearance of being struck twice with a horizontal shift. For encoding purposes, this style can be considered equivalent to letterlike symbols rendered in outlined or shadowed typefaces to carry conventional semantic distinctions.

Standards. The Unicode Standard encodes letterlike symbols from many different national standards and corporate collections.

12.3 Number Forms

Number Forms: U+2150–U+218F

Number form characters are encoded solely for compatibility with existing standards. The same considerations with respect to compatibility apply as noted in the discussion of letterlike symbols.

Fractions. The vulgar fraction characters encoded in this block can be equivalently represented using U+2044 FRACTION SLASH.

Roman Numerals. The Roman numerals can be composed of sequences of the appropriate Latin letters. Upper- and lowercase variants of the Roman numerals through 12, plus L, C, D, and M, have been encoded for compatibility with East Asian standards.

U+2180 ROMAN NUMERAL ONE THOUSAND C D and U+216F ROMAN NUMERAL ONE THOUSAND can be considered to be glyphic variants of the same Roman numeral, but are distinguished because they are not generally interchangeable, and because U+2180 cannot be considered to be a compatibility equivalent to the Latin letter M. U+2181 ROMAN NUMERAL FIVE THOUSAND and U+2182 ROMAN NUMERAL TEN THOUSAND are distinct characters used in Roman numerals; they do not have compatibility decompositions in the Unicode Standard. U+2183 ROMAN NUMERAL REVERSED ONE HUNDRED is a form used in combinations with C and/or I to form large numbers—some of which vary with single character number forms such as D, M, U+2181, or others.

For other number forms, see the Hangzhou numerals (U+3021..U+3029, U+3038..U+303A) and the fractions in the Latin-1 block (U+00BC..U+00BE).

Superscripts and Subscripts: U+2070–U+209F

Superscripts and subscripts have been included in the Unicode Standard only to provide compatibility with existing character sets. In general, the Unicode character encoding does not attempt to describe the positioning of a character above or below the baseline in typographical layout. The superscript digits one, two, and three are coded in the Latin-1 Supplement block to provide code point compatibility with ISO 8859-1.

Standards. The characters in this block are from sets registered with ECMA under ISO 2374 for use with ISO 2022.

12.4 Mathematical Operators

Mathematical Operators: U+2200–U+22FF

The Mathematical Operators block includes character encodings for operators, relations, geometric symbols, and a few other symbols with special usages confined largely to mathematical contexts.

In addition to the characters in this block, mathematical operators are also found in the Basic Latin (ASCII) and Latin-1 Supplement blocks. A few of the symbols from the Miscellaneous Technical block and characters from General Punctuation are also used in mathematical notation. Latin letters in special font styles that are used as mathematical operators, such as U+210B $\mathscr{H}$ SCRIPT CAPITAL H , as well as the Hebrew letter *alef* used as the operator first transfinite cardinal encoded by U+2135 ℵ ALEF SYMBOL, are encoded in the block for letterlike symbols.

Standards. Many national standards' mathematical operators are covered by the characters encoded in this block. These standards include such special collections as ANSI Y10.20, ISO 6862, ISO 8879, and portions of the collection of the American Mathematical Society, as well as the original repertoire of TEX.

Encoding Principles. Mathematical operators often have more than one meaning. Therefore the encoding of this block is intentionally rather shape-based, with numerous instances in which several semantic values can be attributed to the same Unicode value. For example, U+2218 ∘ RING OPERATOR may be the equivalent of *white small circle* or *composite function* or *apl jot*. The Unicode Standard does not attempt to distinguish all possible semantic values that may be applied to mathematical operators or relation symbols.

On the other hand, mathematical operators, and especially relation symbols, may appear in various standards, handbooks, and fonts with a large number of purely graphical variants. Where variants were recognizable as such from the sources, they were not encoded separately.

Unifications. Mathematical operators such as *implies* ⟹ and *if and only if* ⟷ have been unified with the corresponding arrows (U+21D2 RIGHTWARDS DOUBLE ARROW and U+2194 LEFT RIGHT ARROW, respectively) in the Arrows block.

The operator U+2208 ELEMENT OF is occasionally rendered with a taller shape than shown in the code charts. Mathematical handbooks and standards consulted treat these characters as variants of the same glyph. U+220A SMALL ELEMENT OF is a distinctively small version of the *element of* that originates in mathematical pi fonts.

The operators U+226B MUCH GREATER-THAN and U+226A MUCH LESS-THAN are sometimes rendered in a nested shape. Because no semantic distinction applies, the Unicode Standard provides a single encoding for each operator.

A large class of unifications applies to variants of relation symbols involving equality, similarity, and/or negation. Variants involving one- or two-barred *equal signs*, one- or two-*tilde similarity signs*, and vertical or slanted *negation slashes* and *negation slashes* of different lengths are not separately encoded. Thus, for example, U+2288 NEITHER A SUBSET OF NOR EQUAL TO, is the archetype for at least six different glyph variants noted in various collections.

In two instances, essentially stylistic variants are separately encoded: U+2265 GREATER-THAN OR EQUAL TO is distinguished from U+2267 GREATER-THAN OVER EQUAL TO; the same distinction applies to U+2264 LESS-THAN OR EQUAL TO and U+2266 LESS-THAN OVER

EQUAL TO. This exception to the general rule regarding variation results from requirements for character mapping to some Asian standards that distinguish the two forms.

Greek-Derived Symbols. Several mathematical operators derived from Greek characters have been given separate encodings to match usage in existing standards. These operators may occasionally occur in context with Greek-letter variables. They include U+2206 Δ INCREMENT , U+220F ∏ N-ARY PRODUCT, and U+2211 ∑ N-ARY SUMMATION.

Other duplicated Greek characters are those for U+00B5 µ MICRO SIGN in the Latin-1 Supplement block, U+2126 Ω OHM SIGN in Letterlike Symbols, and several characters among the APL functional symbols in the Miscellaneous Technical block. All other Greek characters with special mathematical semantics are found in the Greek block because duplicates were not required for compatibility.

N-ary Operators. N-ary operators are distinguished from binary operators by their larger size and the fact that in mathematical layout, they take limit expressions.

Miscellaneous Symbols. U+2212 − MINUS SIGN is a mathematical operator, to be distinguished from the ASCII-derived U+002D - HYPHEN-MINUS, which may look the same as a minus sign, or may be shorter in length. (For a complete list of dashes in the Unicode Standard, see *Table 6-2*.) U+22EE..U+22F1 are a set of ellipses used in matrix notation.

Mathematical Property. A list of characters with the mathematical property is provided in *Section 4.9, Mathematical Property*.

Arrows: U+2190–U+21FF

Arrows are used for a variety of purposes: to imply directional relation, to show logical derivation or implication, or to represent the cursor control keys.

The Unicode Standard attempts to provide fairly complete encodings for generic arrow shapes, especially where there are established usages with well-defined semantics. It does not attempt to encode every possible stylistic variant of arrows separately, especially where their use is mainly decorative. For most arrow variants, the Unicode Standard provides encodings in the two horizontal directions, often in the four cardinal directions. For the single and double arrows, the Unicode Standard provides encodings in eight directions.

Standards. The Unicode Standard encodes arrows from many different national standards and corporate collections.

Unifications. Arrows expressing mathematical relations have been encoded in the arrows block. An example is U+21D2 ⇒ RIGHTWARDS DOUBLE ARROW, which may be the equivalent of *implies*.

Long and short arrow forms encoded in glyph standards or typesetting systems such as TEX are not represented by separate Unicode values.

Encoding Principles. Because the arrows have such a wide variety of applications, there may be several semantic values for the same Unicode character value. For example, U+21B5 ↵ DOWNWARDS ARROW WITH CORNER LEFTWARDS may be the equivalent of *carriage return*; U+2191 ↑ UPWARDS ARROW may be the equivalent of *increases* or *exponent*.

12.5 Technical Symbols

Control Pictures: U+2400–U+243F

The need to show the presence of the C0 control codes and the SPACE unequivocally when data are displayed has led to conventional representations for these nongraphic characters.

By definition, control codes themselves are manifested only by their action. However, it is sometimes necessary to show the position of a control code within a data stream. Conventional illustrations for the ASCII C0 control codes have been developed.

By definition, the SPACE is a blank graphic. Conventions have also been established for the visible representation of the space.

Standards. The CNS 11643 standard encodes characters for pictures of control codes. Standard representations for control characters have been defined—for example, in ANSI X3.32 and ISO 2047—but for the control code graphics U+2400..U+241F only the semantic is encoded in the Unicode Standard. This choice allows a particular application to use the graphic representation it prefers.

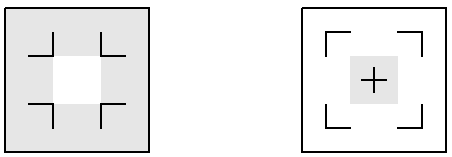
Pictures for ASCII Space. Two specific glyphs are provided that may be used to represent the ASCII space character (U+2420 and U+2422).

Code Points for Pictures for Control Codes. The remaining code points in this block are not associated with specific glyphs, but rather are available to encode *any* desired pictorial representation of the given control code. The assumption is that the particular pictures used to represent control codes are often specific to different systems, and are not often the subject of text interchange between systems.

Miscellaneous Technical: U+2300–U+23FF

This block encodes technical symbols, including keytop labels such as U+232B ERASE TO THE LEFT. Excluded from consideration were symbols that are not normally used in one-dimensional text but are intended for two-dimensional diagrammatic use, such as symbols for electronic circuits. An unusually large expansion space is provided because it is anticipated that a large number of technical symbols could eventually be considered for addition to the Unicode Standard.

Crops and Quine Corners. Crops and quine corners are most properly used in two-dimensional layout but may be referred to in plain text. The usage of crops and quine corners is as indicated in the following diagram:



Use of crops *Use of quine corners*

APL Functional Symbols. APL (A Programming Language) makes extensive use of functional symbols constructed by composition with other, more primitive functional symbols. It made extensive use of backspace and overstrike mechanisms in early computer implementations. In principle, functional composition is productive in APL; in practice,

however, a relatively small number of composed functional symbols have become standard operators in APL. This relatively small set is encoded in entirety in this block. All other APL extensions can be encoded by composition of other Unicode characters. For example, the APL symbol *a underbar* can be represented by U+0061 LATIN SMALL LETTER A + U+0332 COMBINING LOW LINE.

Optical Character Recognition: U+2440–U+245F

This block includes those symbolic characters of the OCR-A character set that do not correspond to ASCII characters, and magnetic ink character recognition (MICR) symbols used in check processing.

Standards. Both sets of symbols are specified in ISO 2033.

12.6 Geometrical Symbols

Box Drawing: U+2500–U+257F

The characters in the Box Drawing block are encoded solely to facilitate the support of legacy implementations, such as terminal emulation.

Standards. GB 2312, KS C 5601, and industry standards were used to develop this block.

Block Elements: U+2580–U+259F

The Block Elements block represents a graphic compatibility zone in the Unicode Standard. A number of existing national and vendor standards, including IBM PC Code Page 437, contain a number of characters intended to enable a simple kind of display cell graphics by filling some fraction of each cell, or by filling each display cell by some degree of shading. The Unicode Standard does not encourage this kind of character-based graphics model but includes a minimal set of such characters for backward compatibility with the existing standards.

Half-block fill characters are included for each half of a display cell, plus a graduated series of vertical and horizontal fractional fills based on one-eighth parts. Also included is a series of shades based on one-quarter shadings. The fractional fills do not form a logically complete set but are intended only for backward compatibility.

Geometric Shapes: U+25A0–U+25FF

The Geometric Shapes are a collection of characters intended to encode prototypes for various commonly used geometrical shapes—mostly squares, triangles, and circles. The collection is somewhat arbitrary in scope; it is a compendium of shapes from various character and glyph standards. The typical distinctions more systematically encoded include black versus white, large versus small, basic shape (square versus triangle versus circle), orientation, and top versus bottom or left versus right part.

The hatched and cross-hatched squares at U+25A4..U+25A9 are derived from the Korean national standard (KS C 5601), in which they were probably intended as representations of fill patterns. Because the semantics of those characters is insufficiently defined in that standard, the Unicode character encoding simply carries the glyphs themselves as geometric shapes to provide a mapping for the Korean standard.

U+25CA ◊ LOZENGE is a typographical symbol seen in PostScript and in the Macintosh character set. It should be distinguished from both the generic U+25C7 WHITE DIAMOND and the U+2662 WHITE DIAMOND SUIT, as well as from another character sometimes called a lozenge, U+2311 SQUARE LOZENGE.

The squares and triangles at U+25E7..U+25EE are derived from the Linotype font collection. U+25EF LARGE CIRCLE is included for compatibility with the JIS X 0208-1990 Japanese standard.

Standards. The Geometric Shapes are derived from a large range of national and vendor character standards.

12.7 Miscellaneous Symbols and Dingbats

Miscellaneous Symbols: U+2600–U+26FF

The Miscellaneous Symbols block consists of a very heterogeneous collection of symbols that do not fit in any other Unicode character block and that tend to be rather pictographic in nature. These symbols are typically used for text decorations, but they may also be treated as normal text characters in applications such as typesetting chess books, card game manuals, and horoscopes.

Characters in the Miscellaneous Symbols block may be rendered in more than one way, unlike characters in the Dingbats block, in which characters correspond to an explicit glyph. For example, both U+2641 EARTH and U+2645 URANUS have common alternative glyphs. EARTH can be rendered as ♁ or ⊕, and URANUS can be rendered as ♅ or ⛢.

The order of the Miscellaneous Symbols is completely arbitrary, but an attempt has been made to keep like symbols together and to group subsets of them into meaningful orders. Some of these subsets include weather and astronomical symbols, pointing hands, religious and ideological symbols, the I Ching trigrams, planet and zodiacal symbols, chess pieces, card suits, and musical dingbats. (For other moon phases, see the circle-based shapes in the Geometric Shapes block.)

Corporate logos and collections of pictures of animals, vehicles, foods, and so on are not included because they tend either to be very specific in usage (logos, political party symbols) or nonconventional in appearance and semantic interpretation (pictures of cows or of cats; fizzing champagne bottles), and hence are inappropriate for encoding as characters. The Unicode Standard recommends that such items be incorporated in text via higher protocols that allow intermixing of graphic images with text, rather than by indefinite extension of the number of Miscellaneous Symbols encoded as characters. However, a large unassigned space has been set aside in this block with the expectation that other conventional sets of such symbols may be found appropriate for character encoding in the future.

Standards. The Miscellaneous Symbols are derived from a large range of national and vendor character standards.

Dingbats: U+2700–U+27BF

The Dingbats are a well-established set of symbols comprising the industry standard "Zapf Dingbat" font—currently available in most laser printers. Other series of dingbats also exist but are not encoded in the Unicode Standard because they are not widely implemented in existing hardware and software as character-encoded fonts. Dingbats that are part of other standards have been encoded in the Geometrical Shapes, Enclosed Alphanumerics, and Miscellaneous Symbols blocks. The order of the remaining dingbats follows the PostScript encoding.

The treatment of the Dingbats in the Unicode Standard differs from that of all other characters. These symbols are encoded as specific glyph shapes, rather than as glyphic archetypes for abstract characters that can be represented in different faces and styles. Thus, it would be incorrect to arbitrarily replace U+279D → TRIANGLE-HEADED RIGHTWARDS ARROW with any other right arrow dingbat or with any of the generic arrows from the Arrows block (U+2190..U+21FF). In other words, because the Zapf Dingbat font refers to glyphs from a specific typeface, their semantic value *is* their shape.

Unifications. A number of the Dingbats represent shapes that overlap with regular Unicode symbol characters. Instead of coding both a Zapf Dingbat glyph shape and a separate character whose glyphic representation is normally indistinguishable from that shape, the Unicode Standard unifies the two. The characters in question include card suits, BLACK STAR, BLACK TELEPHONE, and BLACK RIGHT-POINTING INDEX (see "Miscellaneous Symbols"); BLACK CIRCLE and BLACK SQUARE (see "Geometric Shapes"); white encircled numbers 1 to 10 (see "Enclosed Alphanumerics"); and several generic arrows (see "Arrows"). These four entries appear elsewhere in this section.

The positions of these unified characters are left unassigned in the Dingbats block and are cross-referenced to the assigned positions in the other blocks. Applications may use alternative glyphs for representing those characters (as for any normal Unicode characters), including, of course, the exact shapes required for rendering them in the Zapf Dingbat font on an imaging device.

To illustrate this distinction, an application encoding an encircled digit one with U+2460 ① CIRCLED DIGIT ONE may render that encircled digit in any appropriate typeface—serif or sans serif, roman or italic, and with the circle rendered in different thicknesses. On the other hand, an application encoding an encircled digit one with the Dingbat U+2780 ➀ SANS SERIF CIRCLED DIGIT ONE requires an explicit sans serif glyph from the Zapf Dingbat font for rendering.

12.8 Enclosed and Square

Enclosed Alphanumerics: U+2460–U+24FF

The enclosed numbers and Latin letters of this block come from several sources, chiefly East Asian standards, and are provided for compatibility with them.

Standards. Enclosed letters and numbers occur in the Korean national standard, KS C 5601, and in the Chinese national standard, GB 2312, as well as in various East Asian industry standards.

The Zapf Dingbat character set in widespread industry use contains four sets of encircled numbers (including encircled zero). The black-on-white set that has numbers with serifs is encoded here (U+2460..U+2468, and U+24EA). The other three sets are encoded in the range U+2776..U+2793 in the Dingbats block.

Decompositions. The parenthesized letters or numbers may be decomposed to a sequence of opening parenthesis, letter or digit(s), closing parenthesis. The numbers with a period may be decomposed to digit(s), followed by a period. The encircled letters and single-digit numbers may be decomposed to a letter or digit followed by U+20DD COMBINING ENCLOSING CIRCLE. Decompositions for the encircled numbers 10 through 20 are not supported in Unicode plain text. (For more information, see *Chapter 2, General Structure*, and *Chapter 3, Conformance*.)

Enclosed CJK Letters and Months: U+3200–U+32FF

Standards. This block provides mapping for all the enclosed Hangul elements from Korean standard KS C 5601 as well as parenthesized ideographic characters from JIS X 0208-1990 standard, CNS 11643, and several corporate registries.

CJK Compatibility: U+3300–U+33FF

CJK squared Katakana words are Katakana-spelled words that fill a single display cell (em-square) when intermixed with CJK ideographs. Likewise, squared Latin abbreviation symbols are designed to fill a single character position when mixed with CJK ideographs.

These characters are provided solely for compatibility with existing character encoding standards. Modern software can supply an infinite repertoire of Kana-spelled words or squared abbreviations on the fly.

Standards. CJK Compatibility characters are derived from the KS C 5601 and CNS 11643 national standards, and from various company registries.

Japanese Era Names. The Japanese era names refer to the dates given in *Table 12-2*.

Table 12-2. Japanese Era Names

U+337B	SQUARE ERA NAME HEISEI	1989-01-07 to present day
U+337C	SQUARE ERA NAME SYOUWA	1926-12-24 to 1989-01-06
U+337D	SQUARE ERA NAME TAISYOU	1912-07-29 to 1926-12-23
U+337E	SQUARE ERA NAME MEIZI	1867 to 1912-07-28

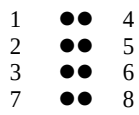
12.9 Braille

Braille: U+2800–U+28FF

Braille is a writing system used by blind people worldwide. It uses a system of six or eight raised dots, arranged in two vertical rows of three or four dots respectively. Eight-dot systems build on six-dot systems by adding two extra dots above or below the core matrix. Six-dot Braille allows 64 possible combinations, and eight-dot Braille allows 256 possible patterns of dot combinations. There is no fixed correspondence between a dot pattern and a character or symbol of any given script. Dot pattern assignments are dependent on context and user community. A single pattern can represent an abbreviation or a frequently occurring short word. For a number of contexts and user communities, the series of international standards starting with ISO 11548-1 provide standardized correspondence tables as well as invocation sequences to indicate a context switch.

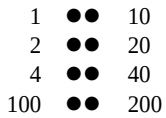
The Unicode Standard encodes a single complete set of 256 eight-dot patterns. This set includes the 64 dot patterns needed for six-dot Braille.

The character names for Braille patterns are based on the assignments of the dots of the Braille pattern to digits 1 to 8 as follows:



The designation of dots 1 to 6 corresponds to that of 6-dot Braille. The additional dots 7 and 8 are added beneath. The character name for a Braille pattern consists of BRAILLE PATTERN DOTS-1234567, where only those digits corresponding to dots in the pattern are included. The name for the empty pattern is BRAILLE PATTERN BLANK.

The 256 Braille patterns are arranged in the same sequence as in ISO 11548-1, which is based on an octal number generated from the pattern arrangement. Octal numbers are associated with each dot of a Braille pattern in the following way:

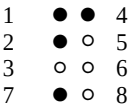


The octal number is obtained by adding the values corresponding to the dots present in the pattern. Octal numbers smaller than 100 are expanded to three digits by inserting leading zeroes. For example, the dots of BRAILLE PATTERN DOTS-1247 are assigned to the octal values of 1_8, 2_8, 10_8, and 100_8. The octal number representing the sum of these values is 113_8.

The assignment of meanings to Braille patterns is outside the scope of this standard.

Example. According to ISO 11548-2, the character LATIN CAPITAL LETTER F can be represented in eight-dot Braille by the combination of the dots 1, 2, 4, and 7 (BRAILLE PATTERN

DOTS-1247). A full circle corresponds to a tangible (set) dot, and empty circles serve as position indicators for dots not set within the dot matrix:



Usage Model. The eight-dot Braille patterns in the Unicode Standard are intended to be used with either style of eight-dot Braille system, whether the additional two dots are considered to be in the top row or in the bottom row. These two systems are never intermixed in the same context, so their distinction is a matter of convention. The intent of encoding the 256 Braille patterns in the Unicode Standard is to allow input and output devices to be implemented that can interchange Braille data without having to go through a context-dependent conversion from semantic values to patterns, or vice versa. In this manner, final form documents can be exchanged and faithfully rendered. On the other hand, processing of textual data that require semantic support is intended to take place using the regular character assignments in the Unicode Standard.

Imaging. When output on a Braille device, dots shown as black are intended to be rendered as tangible. Dots shown in the standard as open circles are blank (not rendered as tangible). The Unicode Standard does not specify any physical dimension of Braille characters.

In the absence of a higher-level protocol, Braille patterns are output from left to right. When used to render final form (tangible) documents, Braille patterns are normally not intermixed with any other Unicode characters except control codes.

This PDF file is an excerpt from *The Unicode Standard, Version 3.0*, issued by the Unicode Consortium and published by Addison-Wesley. The material has been modified slightly for this online edition, however the PDF files have not been modified to reflect the corrections found on the Updates and Errata page (see http://www.unicode.org/unicode/uni2errata/UnicodeErrata.html). More recent versions of the Unicode standard exist (see http://www.unicode.org/unicode/standard/versions/).

Dai Kan-Wa Jiten used as the source of reference Kanji codes was written by Tetsuji Morohashi and published by Taishukan Shoten.

ISBN 0-201-61633-5

This book is set in Minion, designed by Rob Slimbach at Adobe Systems, Inc. It was typeset using FrameMaker 5.5 running under Windows NT. ASMUS, Inc. created custom software for chart layout. The Han radical-stroke index was typeset by Apple Computer, Inc. The following companies and organizations supplied fonts:

Apple Computer, Inc.
Atelier Fluxus Virus
Beijing Zhong Yi (Zheng Code) Electronics Company
DecoType, Inc.
IBM Corporation
Monotype Typography, Inc.
Microsoft Corporation
Peking University Founder Group Corporation
Production First Software

Additional fonts were supplied by individuals as listed in the *Acknowledgments*.

The publisher offers discounts on this book when ordered in quantity for special sales. For more information please contact:

Corporate, Government, and Special Sales
Addison Wesley Longman, Inc.
One Jacob Way
Reading, Massachusetts 01867

Visit A-W on the Web: http://www.awl.com/cseng/

First printing, January 2000.