

Source Standards and Specifications

This section identifies the standards and specifications used as sources for the Unicode Standard. The section also includes selected current standards and specifications relevant to the use of coded character sets. **Last updated April 23, 2020.**

AAT: “About Apple Advanced Typography Fonts.” *TrueType Reference Manual*, Chapter 6: *Font Files*. Apple Computer, ©2014 (last updated 2014).

<https://developer.apple.com/fonts/TrueType-Reference-Manual/RM06/Chap6AATIntro.html>

ANSI X3.4: American National Standards Institute. *Coded character set—7-bit American national standard code for information interchange*. New York: 1986. (ANSI X3.4-1986).

ANSI X3.32: American National Standards Institute. *American national standard graphic representation of the control characters of American national standard code for information interchange*. New York: 1973. (ANSI X3.32-1973).

ANSI Y10.20: American National Standards Institute. *Mathematic signs and symbols for use in physical sciences and technology*. New York: 1988. (ANSI Y10.20-1975 (R1988)).

ANSI Z39.47: American National Standards Institute. *Extended Latin alphabet coded character set for bibliographic use*. New York: 1985. (ANSI Z39.47-1985).

ANSI Z39.64: American National Standards Institute. *East Asian character code for bibliographic use*. New Brunswick, NJ: Transaction, 1991. (ANSI Z39.64-1989).

ARIB STD-B24: Association of Radio Industries and Businesses. *Data Coding and Transmission Specification for Digital Broadcasting*. Tokyo: 2008.

ASMO 449: Arab Organization for Standardization and Metrology. *Data processing 7-bit coded character set for information interchange*. [s.l.]: 1983. (Arab standard specifications, 449-1982). Authorized English translation.

BCP 47: (See RFC 5646 and RFC 4647.)

CCCII: *Zhongwen Zixun Jiaohuanma (Chinese Character Code for Information Interchange)*. Revised edition. Taipei: Xingzhengyuan Wenhua Jianshe Xiaozu (Executive Yuan Committee for Cultural Construction), 1985.

CNS 11643-1986: *Tongyong hanzi biaozhun jiaohuanma (Han character standard interchange code for general use)*. Taipei: Xingzhengyuan (Executive Yuan), 1986.

CNS 11643-1992: *Zhongwen biaozhun jiaohuanma (Chinese standard interchange code)*. Taipei: 1992.

Obsoletes 1986 edition.

DIN 66006: *Informationsverarbeitung—Darstellung von ALGOL-Symbolen auf 5-Spur-Lochstreifen und 80spaltigen Lochkarten. (Information processing—representation of ALGOL symbols on 5-track punched tape and on 80-column punched cards)*. Berlin: Fachnormenausschuß Informationsverarbeitung (FNI) im Deutschen Normenausschuß (DNA), 1965.

Also cited as: *Darstellung von ALGOL/ALCOR-Programmen auf Lochstreifen und Lochkarten.*

EACC: (See ANSI Z39.64.)

ECMA Registry: (See ISO Register.)

ELOT 1373: Hellenic Organization for Standardization (ELOT). *The Greek Byzantine musical notation system.* Athens: 1997.

GB 2312: *Xinxi jiaohuanyong hanzi bianmaj, jibenji (Code of Chinese graphic character set for information interchange, primary set).* Beijing: Jishu Biaozhun Chubanshe (Technical Standards Press), 1981. (GB 2312-80).

GB 13000: *Xinxi jishu—Tongyong duobawei bianma zifuji (UCS)—Diyi bufen: Tixi jiegou yu jiben duowenzhong pingmian (Information technology—Universal multiple-octet coded character set (UCS)—Part 1: Architecture and basic multilingual plane).* Beijing: Jishu Biaozhun Chubanshe (Technical Standards Press), 1993. (GB 13000.1-93). (ISO/IEC 10646.1-1993).

GB 13134: *Xinxi jiaohuanyong yiwen bianma zifuji (Yi coded character set for information interchange),* [prepared by] Sichuansheng Minzushiwu Weiyuanhui. Beijing: Jishu Biaozhun Chubanshe (Technical Standards Press), 1991. (GB 13134-91).

GB 18030: *Xinxi jishu zhongwen bianma zifuji. (Information technology—Chinese coded character set).* Beijing: Guojiao zhiliang jishu jianduju, 2005. (GB 18030-2005).

Revision of the 2000 edition.

GBK: *Xinxi jiaohuanyong hanzi bianma kuozhan guifan (Extended Code of Chinese graphic character set for information interchange).* Beijing: Zhongguo dianzi gongye bu [and] Guojiao jishu jianduju, 1995.

The Chinese-specific subset of GB 13000.1-93.

GB/T 12345: *Xinxi jiaohuanyong hanzi bianmaj, fuzhuji (Code of Chinese ideogram set for information interchange supplementary set).* Beijing: Jishu Biaozhun Chubanshe (Technical Standards Press), 1990. (GB/T 12345-90).

GOST 10859-64: USSR. State Committee on Standards, Measures and Measuring Devices of the USSR. *Computational machinery. Alphanumeric Codes for Punchcards and Punchtapes.* Moscow: Standards Publishing, 1964.

HKSCS: *Hong Kong Supplementary Character Set – 2008.* Hong Kong: Office of the Government Chief Information Officer & Official Languages Division, Civil Service Bureau, Government of the Hong Kong Special Administrative Region, 2008.

English: http://www.ogcio.gov.hk/en/business/tech_promotion/ccli/hkscs/

Simplified Chinese: http://www.ogcio.gov.hk/sc/business/tech_promotion/ccli/hkscs/

Traditional Chinese: http://www.ogcio.gov.hk/tc/business/tech_promotion/ccli/hkscs/

Irish Standard 434:1999. *Information technology—8-bit single-byte graphic coded character set for Ogham / Teicneolaíocht eolais—Tacar carachtar grafach Oghaim códaithe go haonbheartach le 8 ngiotán.*

ISCI-88: India. Department of Electronics. *Indian script code for information interchange.* New Delhi: 1988.

ISCI-91: India. Bureau of Indian Standards. *Indian script code for information interchange.* New Delhi: 1991.

ISIRI 3342: Institute of Standards and Industrial Research of Iran. *estaandaard-e tabaadol-e ettelaa'aat-e 8 biti-e faarsi = Farsi 8-bit coded character set for information interchange.* Tehran: 1993 (1372 AP). (ISIRI 3342:1993).

ISIRI 9147: Institute of Standards and Industrial Research of Iran. *fannaavari-e ettelaa'at—chidemaan-e horoof va alaa'em-e faarsi bar safhe-kelid-e raayaane = Information technology—Layout of Persian letters and symbols on computer keyboards.* Tehran: 2007 (1386 AP). (ISIRI 9147:2007).

ISO Publicly Available Standards. Some ISO and ISO/IEC standards are publicly available for free downloads. See the following URL for information about current availability:

<http://standards.iso.org/ittf/PubliclyAvailableStandards/>

ISO Register: International Organization for Standardization. *ISO international register of coded character sets to be used with escape sequences.*

Current register: https://www.itscj.ipsj.or.jp/itscj_english/iso-ir/ISO-IR.pdf

ISO 639: International Organization for Standardization. *Code for the representation of names of languages.* [Geneva]: 1988. (ISO 639:1988).

ISO/IEC 646: International Organization for Standardization. *Information technology—ISO 7-bit coded character set for information interchange.* [Geneva]: 1991. (ISO/IEC 646:1991).

ISO 1073-1: International Organization for Standardization. *Alphanumeric character sets for optical recognition—Part 1: Character set OCR-A—Shapes and dimensions of the printed image.* [Geneva]: 1976. (ISO 1073-1:1976).

ISO/IEC 2022: International Organization for Standardization. *Information processing—ISO 7-bit and 8-bit coded character sets—Code extension techniques.* 3rd ed. [Geneva]: 1986. (ISO 2022:1994).

Edition 4 (ISO/IEC 2022:1994) has title: *Information technology—Character code structure and extension techniques.*

ISO 2033: International Organization for Standardization. *Information processing—Coding of machine-readable characters (MICR and OCR).* 2nd ed. [Geneva]: 1983. (ISO 2033:1983).

ISO 2047: International Organization for Standardization. *Information processing—Graphical representations for the control characters of the 7-bit coded character set.* [Geneva]: 1975. (ISO 2047:1975).

ISO/IEC 2375: International Organization for Standardization. *Information technology—Procedure for registration of escape sequences and coded character sets.* [Geneva]: 2003. (ISO/IEC 2375:2003).

ISO 3166: International Organization for Standardization. *Codes for the representation of names of countries and their subdivisions.* [Geneva]. Part 1: *Country Codes* (ISO 3166-1:1997). Part 2: *Country subdivision code* (ISO 3166-2:1998). Part 3: *Code for formerly used names of countries* (ISO 3166-3:1999).

ISO/IEC 4873: International Organization for Standardization. *Information technology—ISO 8-bit code for information interchange—Structure and rules for implementation.* [Geneva]: 1991. (ISO/IEC 4873:1991).

ISO 5426: International Organization for Standardization. *Extension of the Latin alphabet coded character set for bibliographic information interchange.* 2nd ed. [Geneva]: 1983. (ISO 5426:1983).

ISO 5426-2: International Organization for Standardization. *Information and documentation—Extension of the Latin alphabet coded character set for bibliographic information interchange—Part 2: Latin characters used in minor European languages and obsolete typography.* [Geneva]: 1996. (ISO 5426-2:1986).

ISO 5427: International Organization for Standardization. *Extension of the Cyrillic alphabet coded character set for bibliographic information interchange.* [Geneva]: 1984. (ISO 5427:1984).

ISO 5428: International Organization for Standardization. *Greek alphabet coded character set for bibliographic information interchange.* [Geneva]: 1984. (ISO 5428-1984).

ISO/IEC 6429: International Organization for Standardization. *Information technology—Control functions for coded character sets.* 3rd ed. [Geneva]: 1992. (ISO/IEC 6429:1992).

ISO 6438: International Organization for Standardization. *Documentation—African coded character set for bibliographic information interchange.* [Geneva]: 1983. (ISO 6438:1983).

ISO 6861:1996. International Organization for Standardization. *Information and documentation—Glagolitic alphabet coded character set for bibliographic information interchange.* [Geneva]: 1996. (ISO 6861:1996).

ISO 6862: International Organization for Standardization. *Information and documentation—Mathematics character set for bibliographic information interchange.* [Geneva]: 1996. (ISO 6862:1996).

ISO/IEC 6937: International Organization for Standardization. *Information processing—Coded character sets for text communication.* [Geneva]: 1984.

Edition 3 (ISO/IEC 6937:2001) has the following title: *Information technology—Coded graphic character set for text communication—Latin alphabet.*

ISO/IEC 8859: International Organization for Standardization. *Information processing—8-bit single-byte coded graphic character sets*. [Geneva]: 1987–.

These parts of ISO/IEC 8859 predate the Unicode Standard, Version 1.0, and were used as resources: Part 1, *Latin alphabet No. 1*; Part 2, *Latin alphabet No. 2*; Part 3, *Latin alphabet No. 3*; Part 4, *Latin alphabet No. 4*; Part 5, *Latin/Cyrillic alphabet*; Part 6, *Latin/Arabic alphabet*; Part 7, *Latin/Greek alphabet*; Part 8, *Latin/Hebrew alphabet*; and Part 9, *Latin alphabet No. 5*.

The other parts of ISO/IEC 8859 are Part 10, *Latin alphabet No. 6*; Part 11, *Latin/Thai alphabet*; Part 13, *Latin alphabet No. 7*; Part 14, *Latin alphabet No. 8 (Celtic)*; Part 15, *Latin alphabet No. 9*; and Part 16, *Latin alphabet No. 10*. There is no Part 12.

ISO 8879: International Organization for Standardization. *Information processing—Text and office systems—Standard generalized markup language (SGML)*. [Geneva]: 1986. (ISO 8879:1986).

ISO 8957: International Organization for Standardization. *Information and documentation—Hebrew alphabet coded character sets for bibliographic information interchange*. [Geneva]: 1996. (ISO 8957:1996).

ISO 9036: International Organization for Standardization. *Information processing—Arabic 7-bit coded character set for information interchange*. [Geneva]: 1987. (ISO 9036:1987).

ISO/IEC 9573-13: International Organization for Standardization. *Information technology—SGML support facilities—Techniques for using SGML—Part 13: Public entity sets for mathematics and science*. [Geneva]: 1991. (ISO/IEC TR 9573-13:1991).

ISO/IEC 9995-7: International Organization for Standardization. *Information technology—Keyboard layouts for text and office systems—Part 7: Symbols used to represent functions*. [Geneva]: 1994. (ISO/IEC 9995-7:1994).

ISO/IEC 10367: International Organization for Standardization. *Information technology—Standardized coded graphic character sets for use in 8-bit codes*. [Geneva]: 1991. (ISO/IEC 10367:1991).

ISO 10585: International Organization for Standardization. *Information and documentation—Armenian alphabet coded character set for bibliographic information interchange*. [Geneva]: 1996. (ISO 10585:1996).

ISO 10586: International Organization for Standardization. *Information and documentation—Georgian alphabet coded character set for bibliographic information interchange*. [Geneva]: 1996. (ISO 10586:1996).

ISO/IEC 10646: International Organization for Standardization. *Information Technology—Universal Multiple-Octet Coded Character Set (UCS)*. [Geneva]: 2017. (ISO/IEC 10646:2017).

ISO 10754: International Organization for Standardization. *Information and documentation—Extension of the Cyrillic alphabet coded character set for non-Slavic languages for bibliographic information interchange*. [Geneva]: 1996. (ISO 10754:1996).

ISO/TR 11548-1: International Organization for Standardization. *Communication aids for blind persons—Identifiers, names and assignation to coded character sets for 8-dot Braille characters—Part 1: General guidelines for Braille identifiers and shift marks.* [Geneva]: 2001. (ISO/TR 11548:2001).

ISO/TR 11548-2: International Organization for Standardization. *Communication aids for blind persons—Identifiers, names and assignation to coded character sets for 8-dot Braille characters—Part 2: Latin alphabet based character sets.* [Geneva]: 2001. (ISO/TR 11548-2:2001).

ISO 11822: International Organization for Standardization. *Information and documentation—Extension of the Arabic coded character set for bibliographic information interchange.* [Geneva]: 1996. (ISO 11822:1996).

ISO/IEC 14496-22: International Organization for Standardization. *Information technology—Coding of audio-visual objects—Part 22: Open Font Format.* [Geneva]: 2015. (ISO/IEC 14496-22:2015). See also OpenType.

ISO/IEC 14651: International Organization for Standardization. *Information technology—International string ordering and comparison—Method for comparing character strings and description of the common template tailorable ordering.* [Geneva]: 2001. (ISO/IEC 14651:2001).

ISO 15285: International Organization for Standardization. *Information Technology—An Operational Model for Characters and Glyphs.* [Geneva]: 1998. (ISO/IEC TR 15285:1998).

ISO 15919: International Organization for Standardization. *Information and documentation—Transliteration of Devanagari and related Indic scripts into Latin characters.* [Geneva]: 2001. (ISO 15919:2001).

ISO 15924: International Organization for Standardization. *Information and Documentation—Codes for the representation of names of scripts = Information et documentation—Codes pour la représentation des noms d'écritures.* Bilingual edition = Édition bilingue. [Geneva: 2004]. (ISO 15924:2004).

ISO/IEC TR 19769: International Organization for Standardization. *Information technology—Programming languages, their environments and system software interfaces—Extensions for the programming language C to support new character data types.* [Geneva]: 2004. (ISO/IEC TR 19769:2004).

ITU-T Recommendation T.101: International Interworking for Videotex Services (November, 1994).

<http://www.itu.int/rec/T-REC-T.101/en>

JIS C 6226: Japanese Industrial Standards Committee. *Jouhou koukan you kanji fugou kei (Code of the Japanese graphic character set for information interchange).* Tokyo: Japanese Standards Association, 1978. (JIS C 6226:1978).

Revised as JIS C 6226:1983, then re-designated as JIS X 0208:1983.

JIS X 0208: Japanese Industrial Standards Committee. *7 bitto oyobi 8 bitto no 2 baito jouhou koukan you fugouka kanji shuugou (7-bit and 8-bit double byte coded kanji sets for information interchange)*. Tokyo: Japanese Standards Association, 1997. (JIS X 0208:1997).

Revision of the 1990 edition, which was the original source for the Unicode Standard. Originally designated as JIS X 0208:1983.

JIS X 0212: Japanese Industrial Standards Committee. *Jouhou koukan you kanji fugou—hojo kanji (Code of the supplementary Japanese graphic character set for information interchange)*. Tokyo: Japanese Standards Association, 1990. (JIS X 0212:1990).

JIS X 0213: Japanese Industrial Standards Committee. *7 bitto oyobi 8 bitto no 2 baito jouhou koukan you fugouka kakuchou kanji shuugou (7-bit and 8-bit double byte coded extended kanji sets for information interchange)*. Tokyo: Japanese Standards Association, 2004. (JIS X 0213:2004).

Revision of the 2000 edition.

JIS X 0221: Japanese Industrial Standards Committee. *Information Technology—Universal Multiple-Octet Coded Character Set (UCS)—Part 1: Architecture and Basic Multilingual Plane*. Tokyo: Japanese Standards Association, 2001. (JIS X 0221-1:2001).

Identical to ISO/IEC 10646-1:2000.

JIS X 4051: Japanese Industrial Standards Committee. *Nihongo Bunsho no Kumihan Houhou. (Formatting rules for Japanese documents)*. Tokyo: Japanese Standards Association, 2004. (JIS X 4051:2004).

Revision of the 1995 edition, used in Unicode Standard Annex #14, *Line Breaking Properties*.

JIS X 4052: Japanese Industrial Standards Committee. *Nihongo Bunsho no Kumihan Shitei Koukan Keishiki. (Exchange format for Japanese documents with composition markup)*. Tokyo: Japanese Standards Association, 2000. (JIS X 4052:2000).

KPS 9566: Committee for Standardization of the Democratic People's Republic of Korea. *(Code of the Korean graphic character set for information interchange)*. Pyongyang: 1997. (KPS 9566:1997).

KPS 10721: Committee for Standardization of the Democratic People's Republic of Korea. *(Code of the supplementary Korean hanja set for information interchange)*. Pyongyang: 2000. (KPS 10721:2000).

KS C 5601: Korea Industrial Standards Association. *Chongbo kyohwanyong puho (Han'gul mit Hancha)*. Seoul: 1989. (KS C 5601:1987).

KS X 1001: Korean Agency for Technology and Standards. *Chongbo kyohwanyong puho (Han'gul mit Hancha). (Code for information interchange (Hangeul and hanja))*. Seoul: 2004. (KS X 1001:2004).

Last confirmed 2014. Originally designated as KS C 5601:1992.

KS X 1002: Korean Agency for Technology and Standards. *Chongbo kyohwanyong puho hwakchang se'u. (Extension code for information interchange.)* Seoul: 1991. (KS X 1002:1991).

Last confirmed 2011. Originally designated as KS C 5657:1991.

KS X 1026-1: Korean Agency for Technology and Standards. *Information Technology – Universal Multiple Octet Coded Character Set – Hangul, Part 1, Hangul processing guide for information interchange.* Seoul: 2008. (KS X 1026-1:2008).

Last confirmed 2013.

MIME: (See RFCs 2045-2049, 4648-4649.)

OpenType: *OpenType™ Specification*, version 1.7. (Microsoft, 2015) See also ISO/IEC 14496-22:2015.

<http://www.microsoft.com/typography/otspec/>

The RFCs listed below are available through the RFC Editor website, which can be found at <https://www.rfc-editor.org/>. This page provides for retrieval by RFC number.

RFC 2045: *Multipurpose Internet Mail Extensions (MIME). Part One: Format of Internet message bodies*, by N. Freed and N. Borenstein. November 1996. (Status: DRAFT STANDARD).

Updated by RFC 2184, RFC 2231.

RFC 2046: *Multipurpose Internet Mail Extensions (MIME). Part Two: Media types*, by N. Freed and N. Borenstein. November 1996. (Status: DRAFT STANDARD).

Updated by RFC 2646, RFC 3798.

RFC 2047: *MIME (Multipurpose Internet Mail Extensions). Part Three: Message header extensions for non-ASCII text*, by K. Moore. November 1996. (Status: DRAFT STANDARD).

Updated by RFC 2184, RFC 2231.

RFC 2048: Obsoleted by RFC 4288, RFC 4289.

RFC 2049: *Multipurpose Internet Mail Extensions (MIME). Part Five: Conformance criteria and examples*, by N. Freed and N. Borenstein. November 1996. (Status: DRAFT STANDARD).

RFC 2152: *UTF-7: A mail-safe transformation format of Unicode*, by D. Goldsmith and M. Davis. May 1997. (Status: INFORMATIONAL).

RFC 3066: Obsoleted by RFC 4646, RFC 4647.

RFC 3629: *UTF-8: A transformation format of ISO 10646*, by F. Yergeau. November 2003. (Also STD0063). (Status: STANDARD).

RFC 4288: *Media type specifications and registration procedures*, by N. Freed and J. Klensin. December 2005. (Also BCP0013). (Status: BEST CURRENT PRACTICE).

RFC 4289: *Multipurpose Internet Mail Extensions (MIME). Part Four: Registration procedures*, by N. Freed and J. Klensin. December 2005. (Also BCP0013). (Status: BEST CURRENT PRACTICE).

RFC 4646: Obsoleted by RFC 5646.

RFC 4647: *Matching of language tags*, edited by A. Phillips and M. Davis. 2006. (Also BCP0047). (Status: BEST CURRENT PRACTICE).

RFC 5646: *Tags for Identifying Languages*, edited by A. Phillips and M. Davis. 2009. (Status: BEST CURRENT PRACTICE).

SI 1311.1: Standards Institution of Israel. *Information technology: ISO 8-bit coded character set with Hebrew points*. [Tel Aviv: 1996]. (SI 1311.1 (1996)).

SI 1311.2: Standards Institution of Israel. *Information technology: ISO 8-bit coded character set with Hebrew accents*. [Tel Aviv: 1996]. (SI 1311.2 (1996)).

SLS 1134: Sri Lanka Standards Institution. *Sinhala character code for information interchange*. Third revision. Colombo: 2011. (SLS 1134: 2011).

TGH 2013: *Tongyong Guifan Hanzibiao (List of Generally Used Standardized Chinese Characters)*. Beijing: Renmin Chubanshe (People's Press), 2013.

TIS 620-2529: Thai Industrial Standards Institute, Ministry of Industry. *Thai Industrial Standard for Thai character code for computer*. Bangkok: 1986. (TIS 620-2529–1986).

TIS 620-2533: Thai Industrial Standards Institute. *Standard for Thai character codes for computers*. Bangkok: 1990. (TIS 620-2533–1990). ISBN 974-606-153-4.

In Thai. Online version: <http://www.nectec.or.th/it-standards/std620/std620.html>

Extensible Markup Language (XML) 1.0. 5th ed. (W3C Recommendation 26 November 2008). Editors: Tim Bray, Jean Paoli, C. M. Sperberg-McQueen, Eve Maler, [and] François Yergeau.

<http://www.w3.org/TR/xml/>