

T.51/ISO/IEC 6937

[6 languages](#)

[Article](#) [Talk](#)

[Read](#) [Edit](#) [View history](#) [Tools](#)

From Wikipedia, the free encyclopedia

T.51 / ISO/IEC 6937:2001, *Information technology — Coded graphic character set for text communication — Latin alphabet*, is a multibyte extension of [ASCII](#), or more precisely [ISO/IEC 646-IRV](#).^[1] It was developed in common with [ITU-T](#) (then [CCITT](#)) for telematic services under the name of *T.51*, and first became an ISO standard in 1983. Certain byte codes are used as lead bytes for letters with [diacritics](#). The value of the lead byte often indicates which diacritic that the letter has, and the follow byte then has the ASCII-value for the letter that the diacritic is on.

ISO/IEC 6937's architects were [Hugh McGregor Ross](#), Peter Fenwick, [Bernard Marti](#) and [Loek Zeckendorf](#).

ISO6937/2 defines 327 characters found in modern European languages using the [Latin alphabet](#). Non-Latin European characters, such as [Cyrillic](#) and [Greek](#), are not included in the standard. Also, some diacritics used with the Latin alphabet like the [Romanian comma](#) are not included, using cedilla instead as no distinction between cedilla and comma below was made at the time.

[IANA](#) has registered the charset names *ISO_6937-2-25* and *ISO_6937-2-add* for two (older) versions of this standard (plus control codes). But in practice this character encoding is unused on the Internet.^[*citation needed*]

Single byte characters [\[edit \]](#)

The primary set (first half) originally followed [ISO 646-IRV](#) before the [ISO/IEC 646:1991](#) revision, that is, mostly following [ASCII](#) but with character 0x24 still denoted as an "international currency sign" (¤) instead of the dollar sign (\$). The 1992 edition of ITU T.51 permits existing CCITT services to continue to interpret 0x24 as the international currency sign, but stipulates that new telecommunication applications should use it for

T.51

Latin based coded character sets for telematic services

Status	In force
Year started	1984
Latest version	(09/92) September 1992
Organization	ITU-T
Committee	Study Group VIII
Related standards	T.61 , ETS 300 706 , ISO/IEC 10367 , ISO/IEC 2022 , ISO 5426
Domain	encoding
License	Freely available
Website	https://www.itu.int/rec/T-REC-T.51

T.51

Alias(es)	Code page 20269 ISO-IR-90 (old) ISO-IR-142 (old) ISO-IR-156
Standard	ISO/IEC 6937 ITU T.51
Based on	ITU T.61
Other related encodings	ETS 300 706 · ISO 5426 · NeXT Multinational · PostScript Standard Encoding · ITU T.101

the dollar sign (i.e. following the current ISO 646-IRV), and instead represent the international currency sign using the supplementary set.^[2]

The supplementary set (second half) contains a selection of spacing and non-spacing graphic characters, additional symbols and some locations reserved for future standardisation.

Both of these are [ISO/IEC 2022](#) graphical character sets, with the primary set being a 94-code set and the secondary set being a 96-code set. In contexts where ISO 2022 code extension techniques are not in use, the primary set is designated as the G0 set and invoked over GL (0x20..0x7F), whereas the supplementary set is designated as the G2 set and invoked over GR (0xA0..0xFF) in an 8-bit environment, or by using the control code 0x19 as a single-shift in a 7-bit environment.^[3] This encoding of the Single Shift Two code matches its location in [ISO-IR-106](#).^[4]

The ISO/IEC 2022 [escape sequence](#) to designate the supplementary set of ISO/IEC 6937 as the G2 set is `ESC . R` (hex `1B 2E 52`).^{[2][5][6]} The older ISO 6937/2:1983 supplementary set is registered as a 94-code set, and designated to G2 with `ESC * 1` (hex `1B 2A 6C`).^{[5][7]}

Two byte characters [\[edit\]](#)

Accented letters which are not allocated single codes in the primary or supplementary set are coded using two bytes. The first byte, the "non spacing diacritical mark", is followed by a letter from the base set e.g.:

```
small e with acute accent (é) = [Acute]+e
```

The ITU T.51 standard allocates column 4 of the supplementary set (i.e. 0xC0–CF when used in 8-bit format) to non-spacing diacritic characters.^[2] However, ISO/IEC 6937 defines a fully specified character repertoire, mapping a list of composition sequences to [ISO/IEC 10646](#) character names which match those defined in Unicode. The isolated nonspacing bytes are not included in this repertoire, although spacing variants of the diacritics not otherwise present in ASCII are included, with the ASCII space being the trail byte.^{[5][8]} Hence, only certain combinations of lead byte and follow byte conform to the ISO/IEC standard.

This repertoire is also affixed to the ITU version of the specification as Annex A, although the ITU version does not reference it from the main text. It is described as a "unified superset" of the Latin-script character repertoires.^[2] It corresponds to the repertoire of [ISO/IEC 10367](#) when the ASCII, [Latin-1](#) (or [Latin-5](#)), [Latin-2](#) and [supplementary Latin](#) sets are used.^[5]

This system also differs from the Unicode [combining character](#) system in that the diacritic code precedes the letter (as opposed to following it), making it more similar to [ANSEL](#).

A little anomaly is that *Latin Small Letter G with Cedilla* is coded as if it were with an acute accent, that is, with a 0xC2 lead byte, since due to its descender interfering with a cedilla, the lowercase letter is usually with turned comma above: *Ç ç*.

In total 13 diacritical marks can be followed by the selected characters from the primary set:

Accent	Code	Second character	Result
Grave	0xC1	AEIOUaeiou	ÀÈÌÒÙàèìòù
Acute	0xC2	ACEILNORSUYZacegilmnorsuyz	ÁĆÉÍĹŃÓŔŚÚÝŽácégíĺńórśúýž
Circumflex	0xC3	ACEGHIJOSUWYaceghijosuwy	ÂÊĜĤÎĴÔŜÛŴŶâêĝĥîĵôŝûŵŷ
Tilde	0xC4	AINOUainou	ÃĨÑÕŨãĩñõũ
Macron	0xC5	AEIOUaeiou	ĀĒĪŌŪāēīōū
Breve	0xC6	AGUagu	ĄǪǻąǫ
Dot	0xC7	CEGIZcegiz	ĆĖĠİŻćėġż
Umlaut or diæresis	0xC8	AEIOUYaeiouy	ÄËÏÖŸäëïöÿ
Ring	0xCA	AUau	ÅŮåů
Cedilla	0xCB	CGKLNRSŤcġklnrst	ÇĢĶĻŅŖŞŢçģķļņŗşţ
Double Acute	0xCD	OUou	ÓÚóú
Ogonek	0xCE	AEIUaeiu	ĄĘĮąęį
Caron	0xCF	CDELNRSTZcdeInrstz	ČĎĚĹŇŘŠŤŽčďěĺňřšťž

Code page layout [[edit](#)]

The reference to [combining characters](#) in the U+0300—U+036F range for the codes in the range 0xC1—0xCF below is subject to the caveats mentioned above; they cannot simply be mapped to the codepoints listed. Also, Unicode distinguishes 0xE2 into uppercase [D with stroke](#) and uppercase [Eth](#), which usually look different for the lowercase letters (0xF2 and 0xF3).

The older 1988 edition of ITU T.51 defined two versions of the supplementary set, with the first version lacking the [non-breaking space](#), [soft hyphen](#), not sign (¬) and broken bar (|) present in the second version. The first version was defined as an extension of the [T.61](#) supplementary set, and the second version as an extension of the first version.^[9] The current (1992) edition only includes the second version, deprecates certain characters, and updates the primary set to the current ISO-646-IRV ([ASCII](#)), although existing telematic services are permitted to retain the older behaviour.^[2]

ISO/IEC 6937 or ITU T.51 (Latin)

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0x																
1x																
2x	SP	!	"	#	\$/α ^[a]	%	&	'	()	*	+	,	-	.	/
3x	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?

4x	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5x	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
6x	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7x	p	q	r	s	t	u	v	w	x	y	z	{		}	~	
8x																
9x																
Ax	NBSP	ı	¢	£	\$ ^[b]	¥	# ^[b]	§	¤	‘	“	«	←	↑	→	↓
Bx	°	±	²	³	×	μ	¶	·	÷	’	”	»	¼	½	¾	¿
Cx		ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	ÿ		
Dx	—	¹	®	©	™	♭	¬	¦					⅛	⅜	⅝	⅞
Ex	Ω	Æ	Ð/Ð	ª	℥	^[d]	¡	£	Ł	∅	ƒ	°	þ	ƒ	Đ	ˆ
Fx	κ	æ	đ	ď	ħ	ı	ij	ı	ı	ø	œ	ß	þ	ı	ŋ	SHY

 Differences from T.61

Videotex version [[edit](#)]

Main article: [Videotex character set](#)

The versions of the supplementary set used by the ITU T.101 standard for [Videotex](#) are based on the first supplementary set of the 1988 edition of T.51.

The default G2 set for Data Syntax 2 adds a “ at 0xC0, for combination with codes from a [Greek](#) primary set.^[10]

The supplementary set for Data Syntax 3 adds non-spacing marks for a "vector overbar" and [solidus](#) and several [semigraphic characters](#).^[11]

ETS 300 706 version [[edit](#)]

The ETS 300 706 standard for [World System Teletext](#) bases its [G2 set](#) on ISO 6937.^[12] It is a superset of the supplementary set of T.61, and a superset of the first supplementary set of the 1988 edition of T.51, but collides with the current edition of T.51 in certain positions. Diacritic codes in the ETS version are specified as being "for association with" characters from the [G0 set in use](#),^[12] such as [US-ASCII](#) or [BS_viewdata](#). This version is shown in the chart below.

World System Teletext, Latin G2 Set (ETS 300 706:1997)^[12]

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
Ax	SP	ı	¢	£	\$	¥	#	§	¤	‘	“	«	←	↑	→	↓
Bx	°	±	²	³	×	μ	¶	·	÷	’	”	»	¼	½	¾	¿

Cx	◌̇	◌̈	◌̉	◌̊	◌̋	◌̌	◌̍	◌̎	◌̏	◌̐	◌̑	◌̒	◌̓	◌̔	◌̕	
Dx	—	1	®	©	™	♪	€	‰	α				¼	⅓	⅕	⅞
Ex	Ω	Æ	Ð/ð	ª	℥		¡	£	Ł	Ø	Œ	◌̊	Ɔ	Ʀ	Ɔ	Ƨ
Fx	κ	æ	đ	ď	ħ	ı	ij	ı	ı	ø	œ	ß	þ	ţ	η	■

 Differences from T.51

See also [[edit](#)]

- [ITU T.50](#)
- [ITU T.61](#), a closely related character encoding for [Teletex](#) use

Footnotes [[edit](#)]

- ^ Continued use for ◌̐ permitted for existing CCITT services only.^[2]
- ^ a b Permitted for existing CCITT services only, otherwise the ASCII representation should be used.^[2]
- ^ Noted in the ITU version of the standard as having existing use for [underlined](#) text, in combination with any other character including accented characters. Although the 1988 ITU edition includes this code,^[9] the 1992 ITU edition discourages sending this code in favour of [ANSI escape sequences](#), although it does mention that it should be correctly interpreted when received by applicable systems.^[2] Previous editions of the ISO/IEC version of the standard also allowed combining this code with any character in the defined repertoire,^[7] whereas more recent revisions do not include this code.^[5]
- ^ An early draft placed j in this position.

References [[edit](#)]

- ^ "T.51 : Latin based coded character sets for telematic services" . [www.itu.int](#). Archived from the original on 2019-10-08. Retrieved 2019-11-14.
- ^ a b c d e f g h CCITT (1992-09-18). *Latin based coded character sets for telematic services* (1992 ed.). Recommendation T.51.
- ^ ITU-T (1995-08-11). *Recommendation T.51 (1992) Amendment 1* .
- ^ ITU (1985-08-01). *Teletex Primary Set of Control Functions* (PDF). ITSCJ/IPSJ. ISO-IR-106.
- ^ a b c d e ISO/IEC JTC 1/SC 2/WG 3 (1998-04-15). *WD 6937, Coded graphic character set for text communication - Latin alphabet* (PDF). JTC1/SC2/N454.
- ^ ISO/IEC JTC 1/SC 2/WG 3 (1991-12-15). *Supplementary Set of ISO/IEC 6937:1992* (PDF). ITSCJ/IPSJ. ISO-IR-156. (The left-hand side is [US-ASCII](#) .)
- ^ a b ISO/TC97/SC2/WG4 (1985-01-10). *Supplementary Set of Latin Alphabetic and non-Alphabetic Graphic Characters* (PDF). ITSCJ/IPSJ. ISO-IR-90.
- ^ Petersen, J. K. (2002-05-29). *The Telecommunications Illustrated Dictionary* . CRC Press. p. 888. ISBN 978-1-4200-4067-8.
- ^ a b CCITT (1988). *Coded character sets for telematic services* (1988 ed.). Recommendation T.51.
- ^ CCITT (1988-11-01). *Supplementary Set of Graphic Characters for Videotex* (PDF). ITSCJ/IPSJ. ISO-IR-70.
- ^ CCITT (1986-11-30). *Supplementary Set of Graphic Characters for CCITT Recommendation T.101, Data Syntax III* (PDF). ITSCJ/IPSJ. ISO-IR-128.

External links [[edit](#)]

- ITU Recommendation T.51
- ISO pages: [ISO 6937-1:1983](#) , [ISO 6937-2:1983](#) , [ISO 6937-2:1983/Add 1:1989](#) , [ISO/IEC 6937:1994](#) , [ISO/IEC 6937:2001](#)
- WD 6937, Coded graphic character set for text communication - Latin alphabet (Revision of ISO/IEC 6937:1994) [Archived](#) 2015-01-04 at the [Wayback Machine](#) (ISO/IEC 6937:1994 draft)
- [ISO-IR-156](#) (ISO-IR registration of right-hand part)

V · T · E	Character encodings	[hide]
Early telecommunication	Telegraph code (Needle · Morse (Non-Latin · Wabun/Kana · Chinese · Cyrillic) · Baudot and Murray) · Fielddata · ASCII (ISO/IEC 646) · BCDIC · Teletex and Videotex/Teletext (T.51/ISO/IEC 6937 · ITU T.61 · ITU T.101 · World System Teletext (background · sets)) · Transcode	
ISO/IEC 8859	Approved parts (-1 (Western Europe) · -2 (Central Europe) · -3 (Maltese/Esperanto) · -4 (North Europe) · -5 (Cyrillic) · -6 (Arabic) · -7 (Greek) · -8 (Hebrew) · -9 (Turkish) · -10 (Nordic) · -11 (Thai) · -13 (Baltic) · -14 (Celtic) · -15 (New Western Europe) · -16 (Romanian)) · Abandoned parts (-12 (Devanagari)) · Proposed but not approved (KOI-8 Cyrillic · Sámi) · Adaptations (Welsh · Estonian · Ukrainian Cyrillic)	
Bibliographic use	MARC-8 (ANSEL · CCCII/EACC) · ISO 5426 · 5426-2 · 5427 · 5428 · 6438 · 6862	
National standards	ArmSCII · Big5 · BraSCII · BSCII · CNS 11643 · DIN 66003 · ELOT 927 · GOST 10859 · GB 2312 · GB 12345 · GB 12052 · GB 18030 · HKSCS · ISCII · JIS X 0201 · JIS X 0208 · JIS X 0212 · JIS X 0213 · KOI-7 · KPS 9566 · KS X 1001 · KS X 1002 · LST 1564 · LST 1590-4 · PASCII · Shift JIS · SI 960 · TIS-620 · TSCII · VISCI · VSCII · YUSCII	
ISO/IEC 2022	ISO/IEC 8859 · ISO/IEC 10367 · Extended Unix Code (EUC)	
Code pages	Armenian · Arabic · Barents Cyrillic · Celtic · Central European · Croatian · Cyrillic · Devanagari · Font X (Kermit) · Gaelic · Georgian · Greek · Gujarati · Gurmukhi · Hebrew · Iceland · Inuit · Keyboard · Latin (Kermit) · Maltese/Esperanto · Ogham · Roman · Romanian · Sámi · Turkish · Turkic Cyrillic · Ukrainian · VT100	
Mac OS ("scripts")		
DOS	437 · 737 · 850 · 858 · 861 · 862 · 863 · 864 · 865 · 866 · 867 · 868 · 869 · 899 · 904 · 932 · 936 · 942 · 949 · 950 · 951 · 1040 · 1043 · 1046 · 1098 · 1115 · 1116 · 1117 · 1118 · 1127 · ABICOMP · CS Indic · CSX Indic · CSX+ Indic · CWI-2 · Iran System · Kamenický · Mazovia · MIK	
IBM AIX	895 · 896 · 912 · 915 · 921 · 922 · 1006 · 1008 · 1009 · 1010 · 1012 · 1013 · 1014 · 1015 · 1016 · 1017 · 1018 · 1019 · 1046 · 1133	
Windows	CER-GS · 932 · 936 (GBK) · 950 · Extended Latin-8 · 1250 · 1251 · 1252 · 1253 · 1254 · 1255 · 1256 · 1257 · 1258 · 1270 · Cyrillic + French · Cyrillic + German · Polytonic Greek	
EBCDIC	Japanese language in EBCDIC · DKOI	
DEC terminals (VTx)	Multinational (MCS) · National Replacement (NRCS) (French Canadian · Swiss · Spanish · United Kingdom · Dutch · Finnish · French · Norwegian and Danish · Swedish · Norwegian and Danish (alternative)) ·	

[8-bit Greek](#) · [8-bit Turkish](#) · [SI 960](#) · [Hebrew](#) · [Special Graphics](#) · [Technical \(TCS\)](#)

Platform specific [1052](#) · [1053](#) · [1054](#) · [1055](#) · [1058](#) · [Acorn RISC OS](#) · [Amstrad CPC](#) · [Apple II](#) · [ATASCII](#) · [Atari ST](#) · [BICS](#) · [Casio calculators](#) · [CDC](#) · [Compucolor 8001](#) · [Compucolor II](#) · [CP/M+](#) · [DEC RADIX 50](#) · [DEC MCS/NRCS](#) · [DG International](#) · [Galaksija](#) · [GEM](#) · [GSM 03.38](#) · [HP Roman](#) · [HP FOCAL](#) · [HP RPL](#) · [SQUOZE](#) · [LICS](#) · [LMBCS](#) · [MSX](#) · [NEC APC](#) · [NeXT](#) · [PETSCII](#) · [PostScript Standard](#) · [PostScript Latin 1](#) · [SAM Coupé](#) · [Sega SC-3000](#) · [Sharp calculators](#) · [Sharp MZ](#) · [Sinclair QL](#) · [Teletext](#) · [TI calculators](#) · [TRS-80](#) · [Ventura International](#) · [WISCII](#) · [XCCS](#) · [ZX80](#) · [ZX81](#) · [ZX Spectrum](#)

Other [ABICOMP](#) · [ASMO 449](#) · [Digital encoding of APL symbols \(ISO-IR-68\)](#) · [ARIB STD-B24](#) · [Fieldata](#) · [HZ](#) · [IEC-P27-1](#) · [INIS \(7-bit · 8-bit\)](#) · [ISO-IR-169](#) · [ISO 2033](#) · [KOI \(KOI8-R · KOI8-RU · KOI8-U\)](#) · [Mojikyō](#) · [SEASCII](#) · [Stanford/ITS](#) · [Symbol](#) · [TRON](#) · [Unified Hangul Code](#)

Unicode, ISO/IEC 10646 [UTF-1](#) · [UTF-7](#) · [UTF-8](#) · [UTF-16](#) · [UTF-32](#) · [UTF-EBCDIC](#) · [GB 18030](#) · [DIN 91379](#) · [BOCU-1](#) · [CESU-8](#) · [SCSU](#) · [TACE16](#) · [Comparison of Unicode encodings](#)

TeX typesetting [Cork](#) · [LY1](#) · [OML](#) · [OMS](#) · [OT1](#)

Control character [Morse prosigns](#) · [C0 and C1 control codes \(ISO/IEC 6429 · JIS X 0211\)](#) · [Unicode control, format and separator characters](#) · [Whitespace characters](#)

Related topics [CCSID](#) · [Character encodings in HTML](#) · [Charset detection](#) · [Han unification](#) · [Hardware code page](#) · [MICR code](#) · [Mojibake](#) · [Variable-length encoding](#)

 [Character sets](#)

[V](#) · [T](#) · [E](#)

[International Organization for Standardization \(ISO\) standards](#)

[\[show\]](#)

Categories: [ITU-T recommendations](#) | [ITU-T T Series Recommendations](#) | [Character encoding](#) | [Character sets](#) | [Computer-related introductions in 1983](#)

This page was last edited on 27 February 2026, at 23:41 (UTC).

Text is available under the [Creative Commons Attribution-ShareAlike 4.0 License](#); additional terms may apply. By using this site, you agree to the [Terms of Use](#) and [Privacy Policy](#). Wikipedia® is a registered trademark of the [Wikimedia Foundation, Inc.](#), a non-profit organization.

[Privacy policy](#) [About Wikipedia](#) [Disclaimers](#) [Contact Wikipedia](#) [Legal & safety contacts](#) [Code of Conduct](#) [Developers](#) [Statistics](#)

[Cookie statement](#) [Mobile view](#)

