

NEWS-1998: This page has been moved to <http://czyborra.com/charsets/iso8859.html>, substantially extended and updated and is now accompanied by additional pages on [ASCII](#), [code pages](#) and [Cyrillic charsets](#).

The ISO 8859 Alphabet Soup

ISO 8859 is a full series of 10 (and soon even [more](#)) standardized multilingual single-byte coded (8bit) graphic character sets for writing in alphabetic languages:

1. [Latin1](#) (West European)
2. [Latin2](#) (East European)
3. [Latin3](#) (South European)
4. [Latin4](#) (North European)
5. [Cyrillic](#)
6. [Arabic](#)
7. [Greek](#)
8. [Hebrew](#)
9. [Latin5](#) (Turkish)
10. [Latin6](#) (Nordic)

The ISO 8859 charsets are not even remotely as complete as the truly great [Unicode](#) but they have been around and usable for quite a while (first [registered Internet charsets](#) for use with [MIME](#)) and have already offered a major improvement over the plain 7bit [US-ASCII](#).

[Unicode \(ISO 10646\)](#) will make this whole chaos of mutually incompatible charsets superfluous because it unifies a superset of all established charsets and is out to cover all the world's languages. But I still haven't seen any software to display all of Unicode on my Unix screen. We're [working](#) on it.

The ISO 8859 charsets were designed in the mid-1980s by the European Computer Manufacturer's Association ([ECMA](#)) and endorsed by the International Standards Organisation ([ISO](#)). The series is currently being revised by the [ISO/IEC JTC1/SC2/WG3](#) working group. The 1998 editions all come with Unicode numbers.

This page exists because the ISO won't provide [free copies](#) of their [published standards](#) (the charset subcommittee [JTC1/SC2](#) has recently called for a free online publication in the future, though, see their [Redmond resolution M08.02: Publication of SC 2 Standards on the web](#)) and the ECMA offers them [on paper only](#).

By clicking at my [TXT]-buttons you can download textual reference tables with Unicode mappings for each of the charsets. You may want to double-check them against more authoritative sources like Keld Simonsen's pioneering [RFC 1345](#), or his [updated and corrected charmaps](#) for i18n@dkuug.dk, mirrored at many Linux's POSIX.2 [/usr/share/i18n/charmap/](#) directory, the [mapping tables on ftp.unicode.org](#), or Kosta Kostis's [transhtm-generated tables](#).

There are [ISO 639 language codes](#) for some 150 of the world's several thousand known languages. The 1998 editions of the ISO-8859 Latin alphabets come with a [table of languages covered](#). A [survey of each language's characters](#) was started by [Harald Alvestrand](#). A more complete but less computerized survey is Akira Nakanishi's colorful book of the "Writing Systems of the World", ISBN [0-8048-1654-9](#). It would be interesting to merge these two into an illustrative [UTF-8](#) text file with [Yudit](#).

The following bitmap GIFs show only the upper [G1](#) portions of the respective charsets. Characters 0 to 127 are always identical with [US-ASCII](#) and the positions 128 to 159 hold some less used control characters: the so-called [C1 set](#) from [ISO 6429](#).

Each image is followed by a link to the textual reference table and the matching [public-domain bitmap font](#) source code in [BDF bitmap distribution format](#) so that you can integrate support for all charsets in your [metamail](#) setup like I did in 1994 in cs.tu-berlin.de/usr/elm/ before our beloved superuser confiscated it because he felt competed or something. Check out the commands [mkfontdir](#) and [xset](#) to install extra fonts on your X terminal. If anybody has converters from BDF to other bitmap formats like those for Windows or MacOS, please send them to me! Most glyphs were extracted from [etl16-unicode.bdf](#) and reassembled using a bunch of perl scripts.

``I'm really terrified to see how difficult it can be for a non-latin1 person to be able to print in his/her own mother tongue!"
-- Akim Demaille, maintainer of [a2ps](#), early 1998

ISO-8859-1 (Latin1)

A0	A1	A2	A3	A4	A5	A6	A7	A8	A9	AA	AB	AC	AD	AE	AF
B0	B1	B2	B3	B4	B5	B6	B7	B8	B9	BA	BB	BC	BD	BE	BF
C0	C1	C2	C3	C4	C5	C6	C7	C8	C9	CA	CB	CC	CD	CE	CF
D0	D1	D2	D3	D4	D5	D6	D7	D8	D9	DA	DB	DC	DD	DE	DF
E0	E1	E2	E3	E4	E5	E6	E7	E8	E9	EA	EB	EC	ED	EE	EF
F0	F1	F2	F3	F4	F5	F6	F7	F8	F9	FA	FB	FC	FD	FE	FF

charset=ISO-8859-1 [TXT] [BDF]

[Latin1](#) covers most **West European** languages, such as French (fr), Spanish (es), Catalan (ca), Basque (eu), Portuguese (pt), Italian (it), Albanian (sq), Rhaeto-Romanic (rm), Dutch (nl), German (de), Danish (da), Swedish (sv), Norwegian (no), Finnish (fi), Faroese (fo), Icelandic (is), Irish (ga), Scottish (gd), and English (en), incidentally also Afrikaans (af) and Swahili (sw), thus in effect also the entire American continent, Australia and much of Africa. The most notable exceptions are Zulu (zu) and other Bantu languages using [Latin Extended-B](#) letters, and of course Arabic in North Africa, and [Guarani](#) (gn) missing GEIUY with ~ tilde. The lack of the ligatures Dutch IJ, French OE and „German“ quotation marks is considered tolerable. The lack of the new C=resembling Euro currency symbol U+20AC has opened the [discussion](#) of a new [Latin0](#).

Latin1 has also been adopted as the first page of ISO 10646 ([Unicode](#)). Latin1 is HTML's [base charset](#) but HTML has now been globalized through [RFC 2070](#). You can browse the [charset smorgasbord](#) or the impressive [IUC10 poster](#) to test your browser or let Andy Flavell tell you more about the [practical problems](#).

DEC-MCS

[ISO-8859-1](#) was derived from the [DEC](#) Multinational Character Set used on the standard DEC VT-220 terminals:

A1	A2	A3	A4	A5	A6	A7	A8	A9	AA	AB	AC	AD	AE	AF
B1	B2	B3	B4	B5	B6	B7	B8	B9	BA	BB	BC	BD	BE	BF
C1	C2	C3	C4	C5	C6	C7	C8	C9	CA	CB	CC	CD	CE	CF
D1	D2	D3	D4	D5	D6	D7	D8	D9	DA	DB	DC	DD	DE	DF
E1	E2	E3	E4	E5	E6	E7	E8	E9	EA	EB	EC	ED	EE	EF
F1	F2	F3	F4	F5	F6	F7	F8	F9	FA	FB	FC	FD	FE	FF

charset=DEC-MCS [TXT] [BDF]

CP1252 (WinLatin1)

You often see Microsoft Windows users (check out my [code page survey](#)) announcing their texts as being in [ISO-8859-1](#) even when in fact they contain funny characters from the CP1252 superset (and they may become more since Microsoft has also added the Euro to their code pages), so here you have a Unix font for them:

B0	B1	B2	B3	B4	B5	B6	B7	B8	B9	BA	BB	BC	BD	BE	BF
C0	C1	C2	C3	C4	C5	C6	C7	C8	C9	CA	CB	CC	CD	CE	CF
D0	D1	D2	D3	D4	D5	D6	D7	D8	D9	DA	DB	DC	DD	DE	DF
E0	E1	E2	E3	E4	E5	E6	E7	E8	E9	EA	EB	EC	ED	EE	EF
F0	F1	F2	F3	F4	F5	F6	F7	F8	F9	FA	FB	FC	FD	FE	FF

charset=Windows-1252 [TXT] [BDF]

ISO-8859-2 (Latin2)

A0	A1	A2	A3	A4	A5	A6	A7	A8	A9	AA	AB	AC	AD	AE	AF
B0	B1	B2	B3	B4	B5	B6	B7	B8	B9	BA	BB	BC	BD	BE	BF
C0	C1	C2	C3	C4	C5	C6	C7	C8	C9	CA	CB	CC	CD	CE	CF
D0	D1	D2	D3	D4	D5	D6	D7	D8	D9	DA	DB	DC	DD	DE	DF
E0	E1	E2	E3	E4	E5	E6	E7	E8	E9	EA	EB	EC	ED	EE	EF
F0	F1	F2	F3	F4	F5	F6	F7	F8	F9	FA	FB	FC	FD	FE	FF

charset=ISO-8859-2 [TXT] [BDF]

[Latin2](#) covers the languages of **Central and Eastern Europe**: Czech (cs), Hungarian (hu), [Polish](#) (pl), Romanian (ro), Croatian (hr), Slovak (sk), Slovenian (sl), Sorbian. For Romanian the S and T had better use commas instead of cedilla as in Turkish: the U+015F LATIN SMALL LETTER S WITH CEDILLA at =BA ought to be read as U+0219 LATIN SMALL LETTER S WITH COMMA BELOW etc.

The German umlauts äöüß are found at exactly the same positions in Latin1, Latin2, Latin3, Latin4, Latin5, Latin6. Thus you can write German+Polish with Latin2 or German+Turkish with Latin5 but there is no 8bit charset to properly mix German+Russian, for instance.

ISO-8859-3 (Latin3)

A0	A1	A2	A3	A4	A5	A6	A7	A8	A9	AA	AB	AC	AD	AE	AF
	H	h	2	3	μ	h	·	1	š	ž	š	ž	-		ž
B0	o	h	2	3	μ	h	·	1	š	ž	š	ž	-		ž
C0	Ā	Ā	Ā	Ā	Ā	Ā	Ā	Ā	Ā	Ā	Ā	Ā	Ā	Ā	Ā
D0	ñ	ñ	ñ	ñ	ñ	ñ	ñ	ñ	ñ	ñ	ñ	ñ	ñ	ñ	ñ
E0	à	à	à	à	à	à	à	à	à	à	à	à	à	à	à
F0	ñ	ñ	ñ	ñ	ñ	ñ	ñ	ñ	ñ	ñ	ñ	ñ	ñ	ñ	ñ

charset=ISO-8859-3 [\[TXT\]](#) [\[BDF\]](#)

Latin3 is popular with authors of [Esperanto](#) (eo) and Maltese (mt), and it covered Turkish before the introduction of [Latin5](#) in 1988.

ISO-8859-4 (Latin4)

A0	A1	A2	A3	A4	A5	A6	A7	A8	A9	AA	AB	AC	AD	AE	AF
	Ā	Ā	Ā	Ā	Ā	Ā	Ā	Ā	Ā	Ā	Ā	Ā	Ā	Ā	Ā
B0	ā	ā	ā	ā	ā	ā	ā	ā	ā	ā	ā	ā	ā	ā	ā
C0	Ā	Ā	Ā	Ā	Ā	Ā	Ā	Ā	Ā	Ā	Ā	Ā	Ā	Ā	Ā
D0	ā	ā	ā	ā	ā	ā	ā	ā	ā	ā	ā	ā	ā	ā	ā
E0	ā	ā	ā	ā	ā	ā	ā	ā	ā	ā	ā	ā	ā	ā	ā
F0	ā	ā	ā	ā	ā	ā	ā	ā	ā	ā	ā	ā	ā	ā	ā

charset=ISO-8859-4 [\[TXT\]](#) [\[BDF\]](#)

Latin4 introduced letters for Estonian (et), the [Baltic](#) languages Latvian (lv, Lettish) and [Lithuanian](#) (lt), Greenlandic (kl) and Lappish. Note that Latvian requires the cedilla on the =BB U+0123 LATIN SMALL LETTER G WITH CEDILLA to jump on top. Latin4 was followed by [Latin6](#).

ISO-8859-5 (Cyrillic)

A0	A1	A2	A3	A4	A5	A6	A7	A8	A9	AA	AB	AC	AD	AE	AF
	Е	Ѣ	Ѥ	Ѧ	Ѩ	Ѭ	Ѯ	Ѱ	Ѳ	Ѵ	Ѷ	Ѹ	-	Ѻ	Ѽ
B0	А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О
C0	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю
D0	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о
E0	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю
F0	ѐ	ѣ	ѥ	ѧ	ѩ	ѫ	ѭ	ѯ	ѱ	ѳ	ѵ	ѷ	ѹ	ѻ	ѽ

charset=ISO-8859-5 [\[TXT\]](#) [\[BDF\]](#)

With these **Cyrillic** letters you can type Bulgarian (bg), Byelorussian (be), Macedonian (mk), Russian (ru), Serbian (sr) and pre-1990 (no [ghe with upturn](#)) Ukrainian (uk). The ordering is based on the (incompatibly) revised GOST 19768 of 1987 with the Russian letters except for ѐ sorted by Russian alphabet (ABVGDE).

Note that [several other Cyrillic charsets](#) are used on the net. Have a look at my neighboring [Cyrillic charsets page](#).

ISO-8859-6 (Arabic)

[illegible]

This is the **Arabic** alphabet, unfortunately the basic alphabet for the Arabic (ar) language only and not containing the four extra letters for Persian (fa) nor the eight extra letters for Pakistani Urdu (ur). This fixed font is not well-suited for text display. Each Arabic letter occurs in up to four (2²) presentation forms: initial, medial, final or separate. To make Arabic text legible you'll need a display engine that analyses the context and combines the appropriate glyphs on top of a handler for the reverse writing direction shared with [Hebrew](#). The rendering algorithm is described in the [Unicode book](#) and I have implemented it in my [arabjoin](#) perl script.

A0	A1	A2	A3			A6	A7	A8	A9		A8	AC	AD	-	AF
B0	B1	B2	B3	B4	B5	B6	B7	B8	B9	BA	BB	BC	BD	BE	BF
C0	C1	C2	C3	C4	C5	C6	C7	C8	C9	CA	CB	CC	CD	CE	CF
D0	D1	D2	D3	D4	D5	D6	D7	D8	D9	DA	DB	DC	DD	DE	DF
E0	E1	E2	E3	E4	E5	E6	E7	E8	E9	EA	EB	EC	ED	EE	EF
F0	F1	F2	F3	F4	F5	F6	F7	F8	F9	FA	FB	FC	FD	FE	FF

This is (modern monotonic) **Greek** (el) to me. ISO-8859-7 was formerly known as [ELOT-928](#) or ECMA-118:1986.

A0		A2	A3	A4	A5	A6	A7	A8	A9	AA	AB	AC	AD	AE	AF
B0	B1	B2	B3	B4	B5	B6	B7	B8	B9	BA	BB	BC	BD	BE	BF
	±	²	³	⁴	⁵	μ	¶	·	₁	÷	»	¼	½	¾	
															DF
															=
E0	E1	E2	E3	E4	E5	E6	E7	E8	E9	EA	EB	EC	ED	EE	EF
	λ	υ	λ	γ	η	ι	π	υ	ι	γ	λ	κ	υ	υ	λ
F0	F1	F2	F3	F4	F5	F6	F7	F8	F9	FA	FB	FC	FD	FE	FF
	υ	υ	γ	γ	γ	υ	π	γ	γ	ω	π				

And this is the [Hebrew](#) script used by **Hebrew** (iw) and Yiddish (ji). Like [Arabic](#) it is [written leftwards](#), so get your dusty old bidirectional typewriters out of the closet! We are promised to see a Bidirectional Algorithm Reference Implementation published as [Unicode Technical Report #9](#) in the near future.

A0	A1	A2	A3	A4	A5	A6	A7	A8	A9	AA	AB	AC	AD	AE	AF
B0	B1	B2	B3	B4	B5	B6	B7	B8	B9	BA	BB	BC	BD	BE	BF
C0	C1	C2	C3	C4	C5	C6	C7	C8	C9	CA	CB	CC	CD	CE	CF
D0	D1	D2	D3	D4	D5	D6	D7	D8	D9	DA	DB	DC	DD	DE	DF
E0	E1	E2	E3	E4	E5	E6	E7	E8	E9	EA	EB	EC	ED	EE	EF
F0	F1	F2	F3	F4	F5	F6	F7	F8	F9	FA	FB	FC	FD	FE	FF

Latin5 replaces the rarely needed Icelandic letters ðýþ in **Latin1** with the **Turkish** ones.

04.07.2005 15:53:13

A0	A1	A2	A3	A4	A5	A6	A7	A8	A9	AA	AB	AC	AD	AE	AF
B0	B1	B2	B3	B4	B5	B6	B7	B8	B9	BA	BB	BC	BD	BE	BF
C0	C1	C2	C3	C4	C5	C6	C7	C8	C9	CA	CB	CC	CD	CE	CF
D0	D1	D2	D3	D4	D5	D6	D7	D8	D9	DA	DB	DC	DD	DE	DF
E0	E1	E2	E3	E4	E5	E6	E7	E8	E9	EA	EB	EC	ED	EE	EF
F0	F1	F2	F3	F4	F5	F6	F7	F8	F9	FA	FB	FC	FD	FE	FF

charset=ISO-8859-10 [TXT] [BDF]

Introduced in 1992, Latin6 rearranged the [Latin4](#) characters, dropped some symbols and the Latvian ¸, added the last missing Inuit (Greenlandic Eskimo) and non-Skolt Sami (Lappish) letters and reintroduced the Icelandic ðýþ to cover the entire **Nordic** area. Skolt Sami still needs a few more accents. Note that RFC 1345 and [GNU recode](#) contain errors and use a preliminary and different latin6.

Future Perspective

From information to be found on [Michael Everson's website](#) and the official [WG 3 website](#) I gathered that in the near future we shall get to see new parts to ISO-8859 which may look like these:

ISO-8859-11 (Thai)

A1	A2	A3	A4	A5	A6	A7	A8	A9	AA	AB	AC	AD	AE	AF
B1	B2	B3	B4	B5	B6	B7	B8	B9	BA	BB	BC	BD	BE	BF
C1	C2	C3	C4	C5	C6	C7	C8	C9	CA	CB	CC	CD	CE	CF
D1	D2	D3	D4	D5	D6	D7	D8	D9	DA	DB	DC	DD	DE	DF
E1	E2	E3	E4	E5	E6	E7	E8	E9	EA	EB	EC	ED	EE	EF
F1	F2	F3	F4	F5	F6	F7	F8	F9	FA	FB	FC	FD	FE	FF

charset=ISO-8859-11 [TXT] [BDF]

The Thai TIS620 is likely to be published as ISO-8859-11 Latin/Thai (th). It contains some combining vowel and tone marks that have to be written above or below the consonants.

ISO-8859-12

There is currently no draft numbered ISO-8859-12. This number might be reserved for ISCII Indian.

It is unlikely that there will ever be a Vietnamese part. Vietnamese (vi) seems to be the language using the most accentuated letters of all languages using the Latin script. Some letters carry a combination of two different accents. They are so many that they simply don't fit into the model of ISO-8859. You can use [VISCII](#) instead.

ISO-8859-13 (Latin7)

A0	A1	A2	A3	A4	A5	A6	A7	A8	A9	AA	AB	AC	AD	AE	AF
B0	B1	B2	B3	B4	B5	B6	B7	B8	B9	BA	BB	BC	BD	BE	BF
C0	C1	C2	C3	C4	C5	C6	C7	C8	C9	CA	CB	CC	CD	CE	CF
D0	D1	D2	D3	D4	D5	D6	D7	D8	D9	DA	DB	DC	DD	DE	DF
E0	E1	E2	E3	E4	E5	E6	E7	E8	E9	EA	EB	EC	ED	EE	EF
F0	F1	F2	F3	F4	F5	F6	F7	F8	F9	FA	FB	FC	FD	FE	FF

charset=ISO-8859-13 [TXT] [BDF]

Latin7 is going to cover the **Baltic Rim** and re-establish the Latvian (lv) support lost in Latin6 and may introduce the local quotation marks. It resembles [WinBaltic](#).

ISO-8859-14 (Latin8)

A0	A1	A2	A3	A4	A5	A6	A7	A8	A9	AA	AB	AC	AD	AE	AF
B0	B1	B2	B3	B4	B5	B6	B7	B8	B9	BA	BB	BC	BD	BE	BF
C0	C1	C2	C3	C4	C5	C6	C7	C8	C9	CA	CB	CC	CD	CE	CF
D0	D1	D2	D3	D4	D5	D6	D7	D8	D9	DA	DB	DC	DD	DE	DF
E0	E1	E2	E3	E4	E5	E6	E7	E8	E9	EA	EB	EC	ED	EE	EF
F0	F1	F2	F3	F4	F5	F6	F7	F8	F9	FA	FB	FC	FD	FE	FF

charset=ISO-8859-14 [TXT] [BDF]

Latin8 adds the last Gaelic and Welsh (cy) letters to Latin1 to cover all Celtic languages.

ISO-8859-15 (Latin9)

A0	A1	A2	A3	A4	A5	A6	A7	A8	A9	AA	AB	AC	AD	AE	AF
B0	B1	B2	B3	B4	B5	B6	B7	B8	B9	BA	BB	BC	BD	BE	BF
C0	C1	C2	C3	C4	C5	C6	C7	C8	C9	CA	CB	CC	CD	CE	CF
D0	D1	D2	D3	D4	D5	D6	D7	D8	D9	DA	DB	DC	DD	DE	DF
E0	E1	E2	E3	E4	E5	E6	E7	E8	E9	EA	EB	EC	ED	EE	EF
F0	F1	F2	F3	F4	F5	F6	F7	F8	F9	FA	FB	FC	FD	FE	FF

charset=ISO-8859-15 [TXT] [BDF]

The new [Latin9](#) nicknamed **Latin0** aims to update [Latin1](#) by replacing the less needed symbols 'ı', '¼', '¾' with forgotten French and Finnish letters and placing the U+20AC Euro sign in the cell =A4 of the former international currency sign ₭.

On 1998-06-28 I suggested to heed the lesson learned and base Latin9 on [ISO-8859-9](#) instead of Latin1 because there is a much greater use for Turkish than for Icelandic but apparently that proposal did not sway the [WG3 standardizers](#).

From: misha.wolf@reuters.com

Date: 22 Jun 1998

To: unicode@unicode.org

Subject: Re: Outlook & the Euro

> ISO 8859-15 will probably be implemented by a number of vendors, but it will take some time until a large percentage of the users start using those versions. Until then, it might be wise *not* to make 8859-15 the default when sending mail.

We have just the place for ISO 8859-15 here in London. It is called the Science Museum and is full of charming historical relics, like Babagge's difference engine, used by Ada Lovelace (I think that was her family name).

What a relief that we now have Unicode and won't have to implement this amusing piece of history.

But with [good Unicode support](#), adding yet another charset is a [piece of cake](#). And the Euro will be needed on systems limited to 8bit. ISO-8859-15 fonts and keysyms have already been included in X11 [R6.4 fix #02](#).

Blurb

I started this page as <http://www.cs.tu-berlin.de/~czyborra/charsets/> on February 27, 1995, in reaction to a request for ISO-8859 code charts on [comp.std.internat](#). Until then, there had only been lousy scans of the ISO charts floating around on the net besides textual tables. I could easily throw this together since I had already gathered all the necessary X11 fonts from [MULE's](#), [Barry Bouwsma's](#) and [Kosta Kostis'](#) collections. Since then has the charsets page had more than 585850 accesses, got copied, included in books, CD-ROMs, and even [translated into French](#). Because of network turbulences at cs.tu-berlin.de that shook the referer database I can only offer you an old list of [who referred](#) to the charsets page.

Thanks go to Sven-Ove Westberg, Alexandre Khalil, Andreas Prilop, [Jacob Andersen](#), Stavros Macrakis, Doug Newell, Chrystopher Nehaniv, Alan Watson, Aaron Irvine, Jonathan Rosenne, Christine Kluka, Clint Adams, Arnold Krivoruk, Van Le, [Jörg Knappen](#), Thomas Henlich, Chris Maden, Paul Keinänen, [Christian Weisgerber](#), Kent Karlsson, [Markus Kuhn](#), [Pino Zollo](#), Imants Metra, [Jukka Korpela](#), and Paul Hill who provided valuable hints for corrections to this page. You are welcome to mail your criticism to roman@czyborra.com.

[Roman Czyborra](#)

\$Date: 1998/12/01 12:39:22 \$