ECMA

EUROPEAN COMPUTER MANUFACTURERS ASSOCIATION

STANDARD ECMA-94

8-BIT SINGLE-BYTE CODED GRAPHIC CHARACTER SET

March 1985

Free copies of this document are available from ECMA,
European Computer Manufacturers Association
114 Rue du Rhône – 1204 Geneva (Switzerland)

ECMA

EUROPEAN COMPUTER MANUFACTURERS ASSOCIATION

STANDARD ECMA-94

8-BIT SINGLE-BYTE CODED GRAPHIC CHARACTER SET

March 1985

BRIEF HISTORY

The adoption of Standard ECMA-6 (ISO 646) as the agreed international 7-bit code for information interchange has led to the development of many national, international and application-oriented versions of this code which are in wide use today.

These versions have a number of limitations generally inherent to the size of the code:

- they do not provide all graphic characters which may be needed,
- for some characters, specially for accented letters, it is necessary to resort to BACKSPACE sequences, which creates problems when processing data containing such composite characters,
- interchange among different versions is practically limited to the 82 common graphic characters.

With the advent of 8-bit coding it was possible to increase the number of graphic characters. ISO 6937/2, for example, provides a character set covering the requirements of most languages based on the Latin alphabet. This character set, although well suited for text communication, is difficult to use for processing as some graphic characters are represented by one and others by two bit combinations.

Thus the need was recognized for coded graphic character sets, each of which:

- is the same for all users of a given area,
- provides single-byte coding of all graphic characters thus permitting easy processing,
- takes into account character sets used in the industry.

Since 1982 the urgency of the need for an 8-bit single-byte coded character set was recognized in ECMA as well as in ANSI/X3L2 and numerous working papers were exchanged between the two groups. In February 1984 ECMA TC1 submitted to ISO/TC97/SC2 a proposal for such a coded character set. At its meeting of April 1984 SC2 decided to submit to TC97 a proposal for a new item of work for this topic. Technical discussions during and after this meeting led TC1 to adopt the coding scheme proposed by X3L2. Part 1 of Draft International Standard DIS 8859 is based on this joint ANSI/ECMA proposal.

Further code tables for other groups of languages which will be added to this Standard in future revisions are already under development.

Adopted as an ECMA Standard by the General Assembly of Dec. 13-14, 1984.

TABLE OF CONTENTS

1. SCOPE

This ECMA Standard consists of a set of 189 graphic characters identified as Latin Alphabet No. 1, and specifies the coded representation of each of these characters by means of a single 8-bit byte.

2. FIELD OF APPLICATION

This set of graphic characters, the Latin Alphabet No. 1, is intended for use in data and text processing applications and may also be used for information interchange.

This set contains graphic characters used in at least the following countries:

Argentina	Finland	Panama
Australia	France	Paraguay
Austria	Germany	Peru
Belgium	Guatemala	Portugal
Belize	Guyana	El Salvador
Bolivia	Honduras	Spain
Brazil	Iceland	Surinam
Canada	Ireland	Sweden
Chile	Italy	Switzerland
Colombia	Liechtenstein	The Netherlands
Costa Rica	Luxemburg	United Kingdom
Cuba	Mexico	United States
Denmark	New Zealand	Uruguay
Ecuador	Nicaragua	Venezuela
Faroe Islands	Norway	

This set of graphic characters is suitable for use in a version of an 8-bit code according to ECMA-35 or ECMA-43.

3. CONFORMANCE

A set of graphic characters is in conformance with this Standard if it comprises all graphic characters specified herein to the exclusion of any other and if their coded representations are those specified by this Standard.

4. REFERENCES

ECMA-6 : 7-bit Input/Output Coded Character Set

ECMA-35 : Code Extension Techniques

ECMA-43 : 8-bit Code for Information Interchange

ECMA-48 : Additional Control Functions for Character-Imaging Devices

5. DEFINITIONS

For the purpose of this Standard the following definitions apply:

5.1 Bit Combination; Byte

An ordered set of bits that represents a character or is used as a part of the representation of a character.

5.2 Character

A member of a set of elements used for the organization, control or representation of data.

5.3 Coded Character Set; Code

A set of unambiguous rules that establishes a character set and the one-to-one relationship between each character of the set and its coded representation.

5.4 Code Table

A table showing the character allocated to each bit combination in a code.

5.5 Graphic Character

A character, other than a control function, that has a visual representation normally handwritten, printed or displayed, and that has a coded representation consisting of one or more bit combinations.

NOTE

In this Standard a single bit combination is used to represent each character.

5.6 Graphic Symbol

A visual representation of a graphic character.

5.7 Position

That part of a code table identified by its column and row co-ordinates.

6. NOTATION, CODE TABLE AND NAMES

6.1 Notation

The bits of the bit combinations of the 8-bit code are identified by b_8, b_7, b_6, b_5, b_4, b_3, b_2 and b_1, where b_8 is the highest-order, or most-significant bit and b_1 is the lowest-order, or least-significant bit.

The bit combinations may be interpreted to represent numbers in the range 0 to 255 in binary notation by attributing the following weights to the individual bits:

Bit	b_8	b_7	b_6	b_5	b_4	b_3	b_2	b_1
Weight	128	64	32	16	8	4	2	1

Using these weights, the bit combinations of the 8-bit code are interpreted to represent numbers in the range 0 to 255.

In this Standard, the bit combinations are identified by notations of the form xx/yy, where xx and yy are numbers in the range 00 to 15. The correspondence between the notations of the form xx/yy and the bit combinations consisting of the bits b_8 to b_1, is as follows:

- xx is the number represented by b_8, b_7, b_6 and b_5 where these bits are given the weights 8, 4, 2 and 1 respectively;
- yy is the number represented by b_4, b_3, b_2 and b_1 where these bits are given the weights 8, 4, 2 and 1 respectively.

6.2 Layout of the Code Table

An 8-bit code table consists of 256 positions arranged in 16 columns and 16 rows. The columns and the rows are numbered 00 to 15.

The code table positions are identified by notations of the form xx/yy, where xx is the column number and yy is the row number.

The positions of the code table are in one-to-one correspondence with the bit combinations of the code. The notation of a code table position, of the form xx/yy, is the same as that of the corresponding bit combination.

6.3 Names and Meanings

This Standard assigns at least one name to each character. In addition, it specifies a graphic symbol for each graphic character. By convention only capital letters, the graphic symbols of small letters and hyphens are used for writing the names of the characters.

The names chosen to denote graphic characters are intended to reflect their customary meaning. However, except for SPACE (SP), NO-BREAK SPACE (NBSP) and SOFT HYPHEN (SHY), this Standard does not define and does not restrict the meanings of graphic characters. Neither does it specify a particular style or font design for imaging graphic characters.

6.3.1 SPACE (SP)

This character may be interpreted as a graphic character, a control character or as both. As a graphic character it has the visual representation consisting of the absence of a graphic symbol.

6.3.2 NO-BREAK SPACE (NBSP)

A graphic character the visual representation of which consists of the absence of a graphic symbol, indicating a point where no line break shall be established by operations determining the layout of text.

6.3.3 SOFT HYPHEN (SHY)

A graphic character that is imaged by a graphic symbol identical with, or similar to, that representing HYPHEN, and which may be inserted or removed by operations determining the layout of text.

7. SPECIFICATION OF THE CODED CHARACTER SET

This Standard specifies 189 characters allocated to the bit combinations of the Code Table. None of these characters are "non-spacing".

The use of control functions, such as BACKSPACE or CARRIAGE RETURN for the coded representation of composite characters is prohibited by this Standard.

7.1 Characters of the Set and their Coded Representation

Bit Combination	Name
02/00	SPACE
02/01	EXCLAMATION MARK
02/02	QUOTATION MARK
02/03	NUMBER SIGN
02/04	DOLLAR SIGN
02/05	PERCENT SIGN
02/06	AMPERSAND
02/07	APOSTROPHE
02/08	LEFT PARENTHESIS
02/09	RIGHT PARENTHESIS
02/10	ASTERISK
02/11	PLUS SIGN
02/12	COMMA
02/13	HYPHEN, MINUS SIGN
02/14	FULL STOP
02/15	SOLIDUS
03/00	DIGIT ZERO
03/01	DIGIT ONE
03/02	DIGIT TWO
03/03	DIGIT THREE
03/04	DIGIT FOUR
03/05	DIGIT FIVE

Bit Combination	Name
03/06	DIGIT SIX
03/07	DIGIT SEVEN
03/08	DIGIT EIGHT
03/09	DIGIT NINE
03/10	COLON
03/11	SEMICOLON
03/12	LESS-THAN SIGN
03/13	EQUALS SIGN
03/14	GREATER-THAN SIGN
03/15	QUESTION MARK
04/00	COMMERCIAL AT
04/01	CAPITAL LETTER A
04/02	CAPITAL LETTER B
04/03	CAPITAL LETTER C
04/04	CAPITAL LETTER D
04/05	CAPITAL LETTER E
04/06	CAPITAL LETTER F
04/07	CAPITAL LETTER G
04/08	CAPITAL LETTER H
04/09	CAPITAL LETTER I
04/10	CAPITAL LETTER J
04/11	CAPITAL LETTER K
04/12	CAPITAL LETTER L
04/13	CAPITAL LETTER M
04/14	CAPITAL LETTER N
04/15	CAPITAL LETTER O
05/00	CAPITAL LETTER P
05/01	CAPITAL LETTER Q
05/02	CAPITAL LETTER R
05/03	CAPITAL LETTER S
05/04	CAPITAL LETTER T
05/05	CAPITAL LETTER U
05/06	CAPITAL LETTER V

Bit Combination	Name
05/07	CAPITAL LETTER W
05/08	CAPITAL LETTER X
05/09	CAPITAL LETTER Y
05/10	CAPITAL LETTER Z
05/11	LEFT SQUARE BRACKET
05/12	REVERSE SOLIDUS
05/13	RIGHT SQUARE BRACKET
05/14	CIRCUMFLEX ACCENT
05/15	LOW LINE, UNDERLINE
06/00	GRAVE ACCENT
06/01	SMALL LETTER a
06/02	SMALL LETTER b
06/03	SMALL LETTER c
06/04	SMALL LETTER d
06/05	SMALL LETTER e
06/06	SMALL LETTER f
06/07	SMALL LETTER g
06/08	SMALL LETTER h
06/09	SMALL LETTER i
06/10	SMALL LETTER j
06/11	SMALL LETTER k
06/12	SMALL LETTER l
06/13	SMALL LETTER m
06/14	SMALL LETTER n
06/15	SMALL LETTER o
07/00	SMALL LETTER p
07/01	SMALL LETTER q
07/02	SMALL LETTER r
07/03	SMALL LETTER s
07/04	SMALL LETTER t
07/05	SMALL LETTER u
07/06	SMALL LETTER v
07/07	SMALL LETTER w
07/08	SMALL LETTER x

Bit Combination	Name
07/09	SMALL LETTER y
07/10	SMALL LETTER z
07/11	LEFT CURLY BRACKET
07/12	VERTICAL LINE
07/13	RIGHT CURLY BRACKET
07/14	TILDE
10/00	NO-BREAK SPACE
10/01	INVERTED EXCLAMATION MARK
10/02	CENT SIGN
10/03	POUND SIGN
10/04	CURRENCY SIGN
10/05	YEN SIGN
10/06	BROKEN BAR
10/07	PARAGRAPH SIGN
10/08	DIAERESIS
10/09	COPYRIGHT SIGN
10/10	FEMININE ORDINAL INDICATOR
10/11	LEFT ANGLE QUOTATION MARK
10/12	NOT SIGN
10/13	SOFT HYPHEN
10/14	REGISTERED TRADE MARK SIGN
10/15	MACRON
11/00	DEGREE SIGN
11/01	PLUS-MINUS SIGN
11/02	SUPERSCRIPT TWO
11/03	SUPERSCRIPT THREE
11/04	ACUTE ACCENT
11/05	SMALL GREEK LETTER MU, MICRO SIGN
11/06	PILCROW SIGN
11/07	MIDDLE DOT
11/08	CEDILLA
11/09	SUPERSCRIPT ONE
11/10	MASCULINE ORDINAL INDICATOR
11/11	RIGHT ANGLE QUOTATION MARK

Bit Combination	Name
11/12	VULGAR FRACTION ONE QUARTER
11/13	VULGAR FRACTION ONE HALF
11/14	VULGAR FRACTION THREE QUARTERS
11/15	INVERTED QUESTION MARK
12/00	CAPITAL LETTER A WITH GRAVE ACCENT
12/01	CAPITAL LETTER A WITH ACUTE ACCENT
12/02	CAPITAL LETTER A WITH CIRCUMFLEX ACCENT
12/03	CAPITAL LETTER A WITH TILDE
12/04	CAPITAL LETTER A WITH DIAERESIS
12/05	CAPITAL LETTER A WITH RING ABOVE
12/06	CAPITAL DIPHTHONG A WITH E
12/07	CAPITAL LETTER C WITH CEDILLA
12/08	CAPITAL LETTER E WITH GRAVE ACCENT
12/09	CAPITAL LETTER E WITH ACUTE ACCENT
12/10	CAPITAL LETTER E WITH CIRCUMFLEX ACCENT
12/11	CAPITAL LETTER E WITH DIAERESIS
12/12	CAPITAL LETTER I WITH GRAVE ACCENT
12/13	CAPITAL LETTER I WITH ACUTE ACCENT
12/14	CAPITAL LETTER I WITH CIRCUMFLEX ACCENT
12/15	CAPITAL LETTER I WITH DIAERESIS
13/00	CAPITAL ICELANDIC LETTER ETH
13/01	CAPITAL LETTER N WITH TILDE
13/02	CAPITAL LETTER O WITH GRAVE ACCENT
13/03	CAPITAL LETTER O WITH ACUTE ACCENT
13/04	CAPITAL LETTER O WITH CIRCUMFLEX ACCENT
13/05	CAPITAL LETTER O WITH TILDE
13/06	CAPITAL LETTER O WITH DIAERESIS
13/07	*(This position shall not be used)*
13/08	CAPITAL LETTER O WITH OBLIQUE STROKE
13/09	CAPITAL LETTER U WITH GRAVE ACCENT
13/10	CAPITAL LETTER U WITH ACUTE ACCENT
13/11	CAPITAL LETTER U WITH CIRCUMFLEX ACCENT
13/12	CAPITAL LETTER U WITH DIAERESIS

Bit Combination	Name
13/13	CAPITAL LETTER Y WITH ACUTE ACCENT
13/14	CAPITAL ICELANDIC LETTER THORN
13/15	SMALL GERMAN LETTER SHARP s
14/00	SMALL LETTER a WITH GRAVE ACCENT
14/01	SMALL LETTER a WITH ACUTE ACCENT
14/02	SMALL LETTER a WITH CIRCUMFLEX ACCENT
14/03	SMALL LETTER a WITH TILDE
14/04	SMALL LETTER a WITH DIAERESIS
14/05	SMALL LETTER a WITH RING ABOVE
14/06	SMALL DIPHTHONG a WITH e
14/07	SMALL LETTER c WITH CEDILLA
14/08	SMALL LETTER e WITH GRAVE ACCENT
14/09	SMALL LETTER e WITH ACUTE ACCENT
14/10	SMALL LETTER e WITH CIRCUMFLEX ACCENT
14/11	SMALL LETTER e WITH DIAERESIS
14/12	SMALL LETTER i WITH GRAVE ACCENT
14/13	SMALL LETTER i WITH ACUTE ACCENT
14/14	SMALL LETTER i WITH CIRCUMFLEX ACCENT
14/15	SMALL LETTER i WITH DIAERESIS
15/00	SMALL ICELANDIC LETTER ETH
15/01	SMALL LETTER n WITH TILDE
15/02	SMALL LETTER o WITH GRAVE ACCENT
15/03	SMALL LETTER o WITH ACUTE ACCENT
15/04	SMALL LETTER o WITH CIRCUMFLEX ACCENT
15/05	SMALL LETTER o WITH TILDE
15/06	SMALL LETTER o WITH DIAERESIS
15/07	*(This position shall not be used)*
15/08	SMALL LETTER o WITH OBLIQUE STROKE
15/09	SMALL LETTER u WITH GRAVE ACCENT
15/10	SMALL LETTER u WITH ACUTE ACCENT
15/11	SMALL LETTER u WITH CIRCUMFLEX ACCENT
15/12	SMALL LETTER u WITH DIAERESIS
15/13	SMALL LETTER y WITH ACUTE ACCENT

Bit Combination	Name
15/14	SMALL ICELANDIC LETTER THORN
15/15	SMALL LETTER y WITH DIAERESIS

7.2 Code Table

The Code Table shows the characters listed at the position in the code table corresponding to the specified bit combination.

The shaded positions correspond to bit combinations that do not represent graphic characters. Their use is outside the scope of this Standard, it is specified in other ECMA Standards, e.g. ECMA-6 or ECMA-48.

7.3 Bit Combinations Not To Be Used

Bit combinations 13/07 and 15/07 are reserved for future standardization and shall not be used. They are cross-hatched in the Code Table.

8. DESIGNATION OF THE CHARACTER SET

The graphic characters of this Standard constitute a single coded character set. However, when this character set is implemented together with other coding standards such as ECMA-35 or ECMA-43, the Code Table of this Standard shall be considered to consist of the following components:

- The character SPACE represented by bit combination 02/00.
- A 94-character G0 graphic character set represented by bit combinations 02/01 to 07/14.
- A 96-character G1 graphic character set represented by bit combinations 10/00 to 15/15.

When required by other coding standards, e.g. ECMA-35 or ECMA-43 the following pair of escape sequences shall be used:

ESC 02/08 04/02
ESC 02/13 F*

to designate the G0 and the G1 sets, respectively. According to ISO 2022 the character SPACE does not require designation.

** The G1 will be registered according to ISO 2375 and the Final character F will be replaced by that allocated by the Registration Authority.*

CODE TABLE

b_4	b_3	b_2	b_1		00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
				b_7	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
				b_6	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
				b_5	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
0	0	0	0	00			SP	0	@	P	`	p			NBSP	°	À	Ð	à	ð
0	0	0	1	01			!	1	A	Q	a	q			¡	±	Á	Ñ	á	ñ
0	0	1	0	02			"	2	B	R	b	r			¢	²	Â	Ò	â	ò
0	0	1	1	03			#	3	C	S	c	s			£	³	Ã	Ó	ã	ó
0	1	0	0	04			$	4	D	T	d	t			¤	´	Ä	Ô	ä	ô
0	1	0	1	05			%	5	E	U	e	u			¥	µ	Å	Õ	å	õ
0	1	1	0	06			&	6	F	V	f	v			¦	¶	Æ	Ö	æ	ö
0	1	1	1	07			'	7	G	W	g	w			§	·	Ç	▒	ç	▒
1	0	0	0	08			(	8	H	X	h	x			¨	¸	È	Ø	è	ø
1	0	0	1	09			)	9	I	Y	i	y			©	¹	É	Ù	é	ù
1	0	1	0	10			*	:	J	Z	j	z			ª	º	Ê	Ú	ê	ú
1	0	1	1	11			+	;	K	[	k	{			«	»	Ë	Û	ë	û
1	1	0	0	12			,	<	L	\	l	|			¬	¼	Ì	Ü	ì	ü
1	1	0	1	13			-	=	M	]	m	}			SHY	½	Í	Ý	í	ý
1	1	1	0	14			.	>	N	^	n	~			®	¾	Î	Þ	î	þ
1	1	1	1	15			/	?	O	_	o				¯	¿	Ï	ß	ï	ÿ

APPENDIX I

RELATIONSHIP OF THIS ECMA STANDARD WITH ISO 6937

The character set defined in this ECMA Standard consists of SPACE, 184 characters that form a sub-set of the repertoire of graphic characters of ISO 6937, and the following four characters that are not defined in the 1983 edition of ISO 6937/2:

¦	BROKEN BAR
¬	NOT SIGN
NBSP	NO-BREAK SPACE
SHY	SOFT HYPHEN

It is ECMA's intention to promote the inclusion of these four characters in ISO 6937, in order that its character repertoire be a super-set of that of this 8-bit single-byte coded graphic character set.

APPENDIX II

7-BIT CODED NATIONAL VERSIONS

Whilst it is expected that the 8-bit coded graphic character set of this ECMA Standard will be adopted as the unique internationally agreed 8-bit single-byte coded character set, it might be necessary during a hopefully short transition period to maintain compatibility with existing 7-bit coded national versions. Where this will be the case it is recommended to apply the following rules when establishing an 8-bit single-byte code table for specific national use.

II.1 The character POUND SIGN can be allocated to bit combination 02/03 and the character NUMBER SIGN to bit combination 10/03.

II.2 The character CURRENCY SIGN can be allocated to bit combination 02/04 and the character DOLLAR SIGN to bit combination 10/04.

II.3 Any other character allocated to a bit combination in the range 10/01 to 15/15 can be allocated to one of the bit combinations 04/00, 05/11, 05/12, 05/13, 05/14, 06/00, 07/11, 07/12, 07/13 or 07/14. The character thus replaced shall be allocated to the bit combination to which the character that has replaced it is allocated in the Code Table of this Standard.

II.4 Any such code table shall comprise all 189 characters.

Although a code table derived from that of this Standard will not be in conformance with this Standard, a simple algorithm will allow conversion from, and to, the standard Code Table.