ECMA

EUROPEAN COMPUTER MANUFACTURERS ASSOCIATION

STANDARD ECMA-144

8-BIT SINGLE-BYTE
CODED GRAPHIC CHARACTER SETS

LATIN ALPHABET No. 6

December 1990

ECMA

EUROPEAN COMPUTER MANUFACTURERS ASSOCIATION

STANDARD ECMA-144

8-BIT SINGLE-BYTE
CODED GRAPHIC CHARACTER SETS

LATIN ALPHABET No. 6

December 1990

BRIEF HISTORY

The adoption of ECMA-6 (ISO 646) as the agreed international 7-bit code for information interchange had 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 let TC1 to adopt the coding scheme proposed by X3L2. International Standard ISO 8859/1 is based on this joint ANSI/ECMA proposal. ECMA published its corresponding Standard ECMA-94 in March 1985.

After this first publication, the work of ECMA TC1 on further coded graphic character sets has led to the following results:

i)   The second Edition of Standard ECMA-94, dated June 1986, comprising four coded graphic character sets for the Latin script, identified as Latin Alphabet No 1 to No 4. These alphabets have a number of characters in common, in particular those allocated to columns 02 to 07. They have all been adopted by ISO and are the subject of ISO 8859, Parts 1 to 4.

ii)   A series of ECMA Standards for coded graphic character sets comprising those characters of the Latin Alphabets allocated to columns 02 to 07 and characters of another script for multiple-language applications. These Standards ECMA-113,

ECMA-114, ECMA-118, ECMA-121 and ECMA-128 cover the Cyrillic, Arabic, Greek and Hebrew scripts, and a set of characters suitable for use with the Turkish language, respectively. They have also been adopted by ISO and published as Part 5, Part 6, Part 7, Part 8, and Part 9, respectively, of ISO 8859.

iii) The present Standard for a Latin Alphabet No. 6 has been derived from Latin Alphabet No. 1. It specifies the characters needed for all majority and minority languages of the Nordic and Baltic countries, except those needed for the Skolt Lappish dialect.

Adopted as an ECMA Standard by the General Assembly of 13th December 1990.

Table of Contents

1. SCOPE

This Standard specifies a set of 191 graphic characters identified as Latin Alphabet No. 6, and specifies the coded representation of each of these characters by means of a single 8-bit byte. 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.

2. FIELD OF APPLICATION

This set of graphic characters, the Latin Alphabet No. 6, is intended for use in data and text processing applications and may also be used for information interchange.

This set is suited for multiple-language applications involving Danish, English, Estonian, Finnish, German, Greenlandic, Icelandic, Lappish, Latvian, Lithuanian, Norwegian, Faroese, and Swedish.

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-35   Code Extension Techniques

ECMA-43   8-bit Coded Character Set - Structure and Rules

ECMA-48   Control Functions for Coded Character Sets

ECMA-94   8-Bit Single-Byte Coded Graphic Character Sets - Latin Alphabets No 1 to No 4

ECMA-113 8-Bit Single-Byte Coded Graphic Character Sets - Latin/Cyrillic Alphabet

ECMA-114 8-Bit Single-Byte Coded Graphic Character Sets - Latin/Arabic Alphabet

ECMA-118 8-Bit Single-Byte Coded Graphic Character Sets - Latin/Greek Alphabet

ECMA-121 8-Bit Single-Byte Coded Graphic Character Sets - Latin/Hebrew Alphabet.

ECMA-128 8-Bit Single-Byte Coded Graphic Character Sets - Latin Alphabet No. 5

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 1

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 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 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.

NOTE 2

The names of the characters in this Standard are those internationally agreed in ISO/IEC/JTC1/SC2, thus, they may differ from the names for the same characters listed in previous ECMA Standards.

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, for use when a line break is to be prevented in the text as presented.

6.3.3 SOFT HYPHEN (SHY)

A graphic character that is imaged by a graphic symbol identical with, or similar to, that representing HYPHEN, for use when a line break is permitted in the text as presented.

7. SPECIFICATION OF THE CODED CHARACTER SET

This Standard specifies 191 characters allocated to the bit combinations of the Code Table.

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

Bit Combination	Name
03/05	DIGIT FIVE
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	LATIN CAPITAL LETTER A
04/02	LATIN CAPITAL LETTER B
04/03	LATIN CAPITAL LETTER C
04/04	LATIN CAPITAL LETTER D
04/05	LATIN CAPITAL LETTER E
04/06	LATIN CAPITAL LETTER F
04/07	LATIN CAPITAL LETTER G
04/08	LATIN CAPITAL LETTER H
04/09	LATIN CAPITAL LETTER I
04/10	LATIN CAPITAL LETTER J
04/11	LATIN CAPITAL LETTER K
04/12	LATIN CAPITAL LETTER L
04/13	LATIN CAPITAL LETTER M
04/14	LATIN CAPITAL LETTER N
04/15	LATIN CAPITAL LETTER O
05/00	LATIN CAPITAL LETTER P
05/01	LATIN CAPITAL LETTER Q
05/02	LATIN CAPITAL LETTER R
05/03	LATIN CAPITAL LETTER S
05/04	LATIN CAPITAL LETTER T
05/05	LATIN CAPITAL LETTER U
05/06	LATIN CAPITAL LETTER V
05/07	LATIN CAPITAL LETTER W
05/08	LATIN CAPITAL LETTER X
05/09	LATIN CAPITAL LETTER Y
05/10	LATIN CAPITAL LETTER Z

Bit Combination	Name
05/11	LEFT SQUARE BRACKET
05/12	REVERSE SOLIDUS
05/13	RIGHT SQUARE BRACKET
05/14	CIRCUMFLEX ACCENT
05/15	LOW LINE
06/00	GRAVE ACCENT
06/01	LATIN SMALL LETTER A
06/02	LATIN SMALL LETTER B
06/03	LATIN SMALL LETTER C
06/04	LATIN SMALL LETTER D
06/05	LATIN SMALL LETTER E
06/06	LATIN SMALL LETTER F
06/07	LATIN SMALL LETTER G
06/08	LATIN SMALL LETTER H
06/09	LATIN SMALL LETTER I
06/10	LATIN SMALL LETTER J
06/11	LATIN SMALL LETTER K
06/12	LATIN SMALL LETTER L
06/13	LATIN SMALL LETTER M
06/14	LATIN SMALL LETTER N
06/15	LATIN SMALL LETTER O
07/00	LATIN SMALL LETTER P
07/01	LATIN SMALL LETTER Q
07/02	LATIN SMALL LETTER R
07/03	LATIN SMALL LETTER S
07/04	LATIN SMALL LETTER T
07/05	LATIN SMALL LETTER U
07/06	LATIN SMALL LETTER V
07/07	LATIN SMALL LETTER W
07/08	LATIN SMALL LETTER X
07/09	LATIN SMALL LETTER Y
07/10	LATIN 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	LATIN CAPITAL LETTER A WITH OGONEK

Bit Combination	Name
10/02	LATIN CAPITAL LETTER E WITH MACRON
10/03	LATIN CAPITAL LETTER G WITH CEDILLA
10/04	LATIN CAPITAL LETTER I WITH MACRON
10/05	LATIN CAPITAL LETTER I WITH TILDE
10/06	LATIN CAPITAL LETTER K WITH CEDILLA
10/07	LATIN CAPITAL LETTER L WITH CEDILLA
10/08	LATIN CAPITAL LETTER N WITH ACUTE
10/09	LATIN CAPITAL LETTER R WITH CEDILLA
10/10	LATIN CAPITAL LETTER S WITH CARON
10/11	LATIN CAPITAL LETTER T WITH STROKE
10/12	LATIN CAPITAL LETTER Z WITH CARON
10/13	SOFT HYPHEN
10/14	LATIN SMALL LETTER KRA (Greenlandic)
10/15	LATIN CAPITAL LETTER ENG (Lappish)
11/00	LATIN SMALL LETTER D WITH STROKE
11/01	LATIN SMALL LETTER A WITH OGONEK
11/02	LATIN SMALL LETTER E WITH MACRON
11/03	LATIN SMALL LETTER G WITH CEDILLA
11/04	LATIN SMALL LETTER I WITH MACRON
11/05	LATIN SMALL LETTER I WITH TILDE
11/06	LATIN SMALL LETTER K WITH CEDILLA
11/07	LATIN SMALL LETTER L WITH CEDILLA
11/08	LATIN SMALL LETTER N WITH ACUTE
11/09	LATIN SMALL LETTER R WITH CEDILLA
11/10	LATIN SMALL LETTER S WITH CARON
11/11	LATIN SMALL LETTER T WITH STROKE
11/12	LATIN SMALL LETTER Z WITH CARON
11/13	SECTION SIGN
11/14	LATIN SMALL LETTER SHARP S (German)
11/15	LATIN SMALL LETTER ENG (Lappish)
12/00	LATIN CAPITAL LETTER A WITH MACRON
12/01	LATIN CAPITAL LETTER A WITH ACUTE
12/02	LATIN CAPITAL LETTER A WITH CIRCUMFLEX
12/03	LATIN CAPITAL LETTER A WITH TILDE
12/04	LATIN CAPITAL LETTER A WITH DIAERESIS
12/05	LATIN CAPITAL LETTER A WITH RING ABOVE
12/06	LATIN CAPITAL LETTER AE
12/07	LATIN CAPITAL LETTER I WITH OGONEK

Bit Combination	Name
12/08	LATIN CAPITAL LETTER C WITH CARON
12/09	LATIN CAPITAL LETTER E WITH ACUTE
12/10	LATIN CAPITAL LETTER E WITH OGONEK
12/11	LATIN CAPITAL LETTER E WITH DIAERESIS
12/12	LATIN CAPITAL LETTER E WITH DOT ABOVE
12/13	LATIN CAPITAL LETTER I WITH ACUTE
12/14	LATIN CAPITAL LETTER I WITH CIRCUMFLEX
12/15	LATIN CAPITAL LETTER I WITH DIAERESIS
13/00	LATIN CAPITAL LETTER D WITH STROKE
13/01	LATIN CAPITAL LETTER N WITH CEDILLA
13/02	LATIN CAPITAL LETTER O WITH MACRON
13/03	LATIN CAPITAL LETTER O WITH ACUTE
13/04	LATIN CAPITAL LETTER O WITH CIRCUMFLEX
13/05	LATIN CAPITAL LETTER O WITH TILDE
13/06	LATIN CAPITAL LETTER O WITH DIAERESIS
13/07	LATIN CAPITAL LETTER U WITH TILDE
13/08	LATIN CAPITAL LETTER O WITH STROKE
13/09	LATIN CAPITAL LETTER U WITH OGONEK
13/10	LATIN CAPITAL LETTER U WITH ACUTE
13/11	LATIN CAPITAL LETTER U WITH CIRCUMFLEX
13/12	LATIN CAPITAL LETTER U WITH DIAERESIS
13/13	LATIN CAPITAL LETTER Y WITH ACUTE
13/14	LATIN CAPITAL LETTER THORN (Icelandic)
13/15	LATIN CAPITAL LETTER U WITH MACRON
14/00	LATIN SMALL LETTER A WITH MACRON
14/01	LATIN SMALL LETTER A WITH ACUTE
14/02	LATIN SMALL LETTER A WITH CIRCUMFLEX
14/03	LATIN SMALL LETTER A WITH TILDE
14/04	LATIN SMALL LETTER A WITH DIAERESIS
14/05	LATIN SMALL LETTER A WITH RING ABOVE
14/06	LATIN SMALL LETTER AE
14/07	LATIN SMALL LETTER I WITH OGONEK
14/08	LATIN SMALL LETTER C WITH CARON
14/09	LATIN SMALL LETTER E WITH ACUTE
14/10	LATIN SMALL LETTER E WITH OGONEK
14/11	LATIN SMALL LETTER E WITH DIAERESIS
14/12	LATIN SMALL LETTER E WITH DOT ABOVE
14/13	LATIN SMALL LETTER I WITH ACUTE

Bit Combination	Name
14/14	LATIN SMALL LETTER I WITH CIRCUMFLEX
14/15	LATIN SMALL LETTER I WITH DIAERESIS
15/00	LATIN SMALL LETTER ETH (Icelandic)
15/01	LATIN SMALL LETTER N WITH CEDILLA
15/02	LATIN SMALL LETTER O WITH MACRON
15/03	LATIN SMALL LETTER O WITH ACUTE
15/04	LATIN SMALL LETTER O WITH CIRCUMFLEX
15/05	LATIN SMALL LETTER O WITH TILDE
15/06	LATIN SMALL LETTER O WITH DIAERESIS
15/07	LATIN SMALL LETTER U WITH TILDE
15/08	LATIN SMALL LETTER O WITH STROKE
15/09	LATIN SMALL LETTER U WITH OGONEK
15/10	LATIN SMALL LETTER U WITH ACUTE
15/11	LATIN SMALL LETTER U WITH CIRCUMFLEX
15/12	LATIN SMALL LETTER U WITH DIAERESIS
15/13	LATIN SMALL LETTER Y WITH ACUTE
15/14	LATIN SMALL LETTER THORN (Icelandic)
15/15	LATIN SMALL LETTER U WITH MACRON

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 standards, e.g. ECMA-43 and ECMA-48.

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 04/..

to designate the G0 and the G1 sets, respectively. According to ECMA-35 the character SPACE does not require designation.

CODE TABLE OF LATIN ALPHABET No. 6

b8					0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
b7					0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
b6					0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
b5					0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
b4	b3	b2	b1		00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
0	0	0	0	00			SP	0	@	P	`	p			NBSP	đ	Ā	Đ	ā	ð
0	0	0	1	01			!	1	A	Q	a	q			Ą	ą	Á	Ŋ	á	ŋ
0	0	1	0	02			"	2	B	R	b	r			Ē	ē	Â	Ō	â	ō
0	0	1	1	03			#	3	C	S	c	s			Ģ	ġ	Ã	Ó	ã	ó
0	1	0	0	04			$	4	D	T	d	t			Ī	ī	Ä	Ô	ä	ô
0	1	0	1	05			%	5	E	U	e	u			Ĩ	ĩ	Å	Õ	å	õ
0	1	1	0	06			&	6	F	V	f	v			Ķ	ķ	Æ	Ö	æ	ö
0	1	1	1	07			'	7	G	W	g	w			Ļ	ļ	Į	Ũ	į	ũ
1	0	0	0	08			(	8	H	X	h	x			Ń	ń	Č	Ø	č	ø
1	0	0	1	09			)	9	I	Y	i	y			Ŗ	ŗ	É	Ų	é	ų
1	0	1	0	10			*	:	J	Z	j	z			Š	š	Ę	Ú	ę	ú
1	0	1	1	11			+	;	K	[	k	{			Ŧ	ŧ	Ë	Û	ë	û
1	1	0	0	12			,	<	L	\	l	|			Ž	ž	Ė	Ü	ė	ü
1	1	0	1	13			-	=	M	]	m	}			SHY	§	Í	Ý	í	ý
1	1	1	0	14			.	>	N	^	n	~			ĸ	ß	Î	Þ	î	þ
1	1	1	1	15			/	?	O	_	o				Ŋ	ŋ	Ï	Ū	ï	ū

APPENDIX A

(This Appendix is not part of the Standard)

The coded character set specified by this ECMA Standard comprises the characters needed for the Southern, Lule, Northern and Enare dialects of the Lappish language in modern orthography. The Skolt Lappish dialect used in parts of Finland requires certain additional characters that are not specified in this Standard. Those additional characters, as well as certain characters used in other historic Lappish documents, have been registered according to ISO 2375 as Registration No. 158. For text in the Skolt Lappish dialect, as well as texts using older Lappish orthography, it is recommended to use the character set specified in this ECMA Standard the G0 and G1 set together with the character set of this Registration No. 158 as the G2 or G3 set according to level 2 of Standard ECMA-43.

The character set according to Registration No. 158 is designated with the following escape sequence:

G2: ESC 02/14 ../..
G3: ESC 02/15 ../..

For convenience, the code table of Registration No. 158 is reproduced below.

b₇		0	0	0	0	1	1	1	1
b₆		0	0	1	1	0	0	1	1
b₅		0	1	0	1	0	1	0	1
b₄ b₃ b₂ b₁		0	1	2	3	4	5	6	7
0 0 0 0	0			′	'	Ă		ă	
0 0 0 1	1					À		à	
0 0 1 0	2					Ā		ā	
0 0 1 1	3					Ȧ		ȧ	
0 1 0 0	4					Ǣ		ǣ	
0 1 0 1	5					Ĕ		ĕ	
0 1 1 0	6					È		è	
0 1 1 1	7					G̣		g̣	
1 0 0 0	8					Ğ		ğ	
1 0 0 1	9					Ǩ		ǩ	
1 0 1 0	10					Ŏ		ŏ	
1 0 1 1	11					Ò		ò	
1 1 0 0	12					Q		q	
1 1 0 1	13					Q̄		q̄	
1 1 1 0	14					Ʒ		ʒ	
1 1 1 1	15					Ǯ		ǯ	