

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. 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 present Standard for a Latin/Arabic coded graphic character set. In developing this ECMA Standard TC1 closely co-operated with the relevant groups and committees of ASMO, the Arab Organization for Standardization and Metrology, of ATU, the Arab Telecommunication Union, and of different Arabic countries.

- ii) The second Edition of Standard ECMA-94 comprising four coded graphic character sets for the Latin script, identified as Latin Alphabets No 1 to No 4. These alphabets have a number of characters in common, in particular those allocated to columns 02 to 07. Latin Alphabet No 2 has been submitted to ISO and is the subject of ISO 8859/2.
- iii) 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 ECMA Standards cover the Greek and Cyrillic scripts. It is intended to submit them to ISO for further processing as ISO standards.

Adopted as an ECMA Standard by the General Assembly of June 26, 1986.

TABLE OF CONTENTS

	<u>Page</u>
1. SCOPE	1
2. FIELD OF APPLICATION	1
3. CONFORMANCE	1
4. REFERENCES	1
5. DEFINITIONS	1
5.1 Bit Combination; Byte	2
5.2 Character	2
5.3 Coded Character Set; Code	2
5.4 Code Table	2
5.5 Graphic Character	2
5.6 Graphic Symbol	2
5.7 Position	2
6. NOTATION, CODE TABLE AND NAMES	2
6.1 Notation	2
6.2 Layout of the Code Table	3
6.3 Names and Meanings	3
6.3.1 SPACE (SP)	3
6.3.2 NO-BREAK SPACE (NBSP)	3
6.3.3 SOFT HYPHEN (SHY)	4
6.4 Imaging Graphic Symbols	4
6.4.1 Alternative Sets of Digits	4
6.4.2 Mirror Shapes	4
7. SPECIFICATION OF THE CODED CHARACTER SET	4
7.1 Characters of the Set and Their Coded Representation	4
7.2 Code Table	10
7.3 Bit Combination Not To Be Used	10
8. DESIGNATION OF THE CHARACTER	10

1. SCOPE

This ECMA Standard specifies a set of 146 characters identified as the Arabic-Latin Alphabet and the coded representation of each of these characters by means of a single 8-bit byte.

The implementation of this Standard will require the use of specific control functions due to the fact that languages using the Latin script are written from left to right and Arabic is written from right to left. In addition, statements expressed with one of the scripts may have to be included in text using the other script. Also the insertion of numerical values may require a specific treatment. The specification of the control functions to be used with the set of graphic characters of this Standard is outside the scope of this Standard.

2. FIELD OF APPLICATION

This set of graphic characters is intended for use in information interchange as well as in data and text processing applications where both the Arabic and the Latin scripts are used. 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 Coded Character Set - Structure and Rules
- ECMA-48 : Control Functions
- 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- : 8-bit Single-Byte Coded Graphic Character Sets - Latin/Greek Alphabet (in preparation)
- ASMO 449 : 7-bit Coded Arabic Character Set for Information Interchange

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 ECMA Standard a single 8-bit combination is used to represent each graphic 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, 1 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 unique graphic symbol for each graphic character, except for those specified in 6.4. By convention only capital letters, the graphic symbols of small letters and hyphens are used for writing the names of the characters.

Note 2

The graphic symbols of each Arabic alphabetic character may have more than one shape. The appropriate shape to be imaged is determined by contextual analysis, which is outside the scope of this Standard.

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, 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.

6.4 Imaging Graphic Symbols

6.4.1 Alternative Sets of Digits

The characters allocated to positions 03/00 to 03/09 are shown with two alternative graphic symbols. The left hand set of graphic symbols is that of the Latin script, the right hand set is that of the Arabic script.

6.4.2 Mirror Shapes

The graphic symbols shown in the Code Table in positions 02/08, 02/09, 03/12, 03/14, 05/11, 05/13, 07/11 and 07/13 are those to be imaged unless preceded by an appropriate control function. In this case their mirror shape shall be imaged.

7. SPECIFICATION OF THE CODED CHARACTER SET

This Standard specifies 146 graphic 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

Bit Combination	Name
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
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

Bit Combination	Name
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
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
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

Bit Combination	Name
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
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	(This position shall not be used)
10/02	(This position shall not be used)
10/03	(This position shall not be used)
10/04	CURRENCY SIGN
10/05	(This position shall not be used)
10/06	(This position shall not be used)
10/07	(This position shall not be used)
10/08	(This position shall not be used)
10/09	(This position shall not be used)
10/10	(This position shall not be used)
10/11	(This position shall not be used)
10/12	ARABIC COMMA
10/13	SOFT HYPHEN
10/14	(This position shall not be used)
10/15	(This position shall not be used)
11/00	(This position shall not be used)
11/01	(This position shall not be used)
11/02	(This position shall not be used)
11/03	(This position shall not be used)
11/04	(This position shall not be used)

Bit Combination	Name
11/05	(This position shall not be used)
11/06	(This position shall not be used)
11/07	(This position shall not be used)
11/08	(This position shall not be used)
11/09	(This position shall not be used)
11/10	(This position shall not be used)
11/11	ARABIC SEMICOLON
11/12	(This position shall not be used)
11/13	(This position shall not be used)
11/14	(This position shall not be used)
11/15	ARABIC QUESTION MARK
12/00	(This position shall not be used)
12/01	HAMZA
12/02	MADDA ON ALEF
12/03	HAMZA ON ALEF
12/04	HAMZA ON WAW
12/05	HAMZA UNDER ALEF
12/06	HAMZA ON YEH
12/07	ALEF
12/08	BEH
12/09	TEH MARBUTA
12/10	TEH
12/11	THEH
12/12	JEEM
12/13	HAH
12/14	KHAH
12/15	DAL
13/00	THAL
13/01	RA
13/02	ZAIN
13/03	SEEN
13/04	SHEEN
13/05	SAD
13/06	DAD
13/07	TAH
13/08	ZAH

Bit Combination	Name
13/09	AIN
13/10	GHAIN
13/11	(This position shall not be used)
13/12	(This position shall not be used)
13/13	(This position shall not be used)
13/14	(This position shall not be used)
13/15	(This position shall not be used)
14/00	TATWEEL
14/01	FEH
14/02	QAF
14/03	KAF
14/04	LAM
14/05	MEEM
14/06	NOON
14/07	HEH
14/08	WAW
14/09	ALEF MAKSURA
14/10	YEH
14/11	FATHATAN
14/12	DAMMATAN
14/13	KASRATAN
14/14	FATHA
14/15	DAMMA
15/00	KASRA
15/01	SHADDA
15/02	SUKUN
15/03	(This position shall not be used)
15/04	(This position shall not be used)
15/05	(This position shall not be used)
15/06	(This position shall not be used)
15/07	(This position shall not be used)
15/08	(This position shall not be used)
15/09	(This position shall not be used)
15/10	(This position shall not be used)

Bit Combination	Name
15/11	(This position shall not be used)
15/12	(This position shall not be used)
15/13	(This position shall not be used)
15/14	(This position shall not be used)
15/15	(This position shall not be used)

7.2 Code Table

The Code Table on the next page 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. They are reserved for the representation of control functions.

7.3 Bit Combinations Not To Be Used

A number of bit combinations are reserved for future standardization and shall not be used. They are 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 ECMA-35 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.

b.	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	
b.	0	0	0	0	1	1	1	0	0	0	0	1	1	1	1	
b.	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	
b.	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	
	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15

b.	b.	b.	b.																			
0	0	0	0	00			SP	0	·	Q	P	`	p			NBSP			ذ	—	ا	
0	0	0	1	01			!	1	؛	A	Q	a	q						ء	ر	ف	س
0	0	1	0	02			"	2	؛	B	R	b	r						آ	ز	ق	ه
0	0	1	1	03			#	3	؛	C	S	c	s						أ	س	ك	
0	1	0	0	04			\$	4	؛	D	T	d	t			⌘			ؤ	ش	ل	
0	1	0	1	05			%	5	؛	E	U	e	u						إ	ص	م	
0	1	1	0	06			&	6	؛	F	V	f	v						ع	ض	ن	
0	1	1	1	07			'	7	؛	G	W	g	w						ا	ط	ه	
1	0	0	0	08			(8	؛	H	X	h	x						ب	ظ	و	
1	0	0	1	09)	9	؛	I	Y	i	y						ة	ع	ى	
1	0	1	0	10			*	:		J	Z	j	z						ت	غ	ي	
1	0	1	1	11			+	;		K	[k	{						؛	ث	=	
1	1	0	0	12			/	<		L	\	l							،	ج	"	
1	1	0	1	13			-	=		M]	m	}						SHY	ح	=	
1	1	1	0	14			.	>		N	^	n	~							خ	'	
1	1	1	1	15			/	?		O	_	o							؟	د	,	